

Prévision de la turbidité par apprentissage statistique: application au captage AEP d'Yport (Normandie)

Michael Savary

▶ To cite this version:

Michael Savary. Prévision de la turbidité par apprentissage statistique : application au captage AEP d'Yport (Normandie). Sciences de la Terre. Normandie Université, 2018. Français. NNT : 2018NORMR091 . tel-01992886

HAL Id: tel-01992886 https://theses.hal.science/tel-01992886

Submitted on 24 Jan 2019 $\,$

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

Pour obtenir le diplôme de doctorat Spécialité (Sciences de la Vie et de la Santé) Préparée au sein de « Université de Rouen Normandie »

Prévision de la turbidité par apprentissage statistique : application au captage AEP d'Yport (Normandie)

Présentée et soutenue par Michaël SAVARY

Thèse soutenue publiquement le (date de soutenance) devant le jury composé de						
Monsieur David LABAT	Professeur / UMR 5563 CNRS : GET (Université de Toulouse)	Rapporteur				
Monsieur Séverin PISTRE	Professeur / UMR 5569 CNRS : HSM (Université de Montpellier)	Rapporteur				
Madame Nicole BARAN	Ingénieur de recherche / BRGM Orléans	Examinatrice				
Madame Danièle VALDES-LAO	Maitre de Conférences / UMR 7619 METIS	Examinatrice				
Monsieur Nicolas MASSEI	Professeur / UMR 6143 CNRS : M2C (Université de Rouen)	Directeur de thèse				
Madame Anne JOHANNET	Professeur / Laboratoire LGEI (Ecole des mines d'Alès)	Directrice de thèse				
Monsieur Jean-Paul DUPONT	Professeur / UMR 6143 CNRS : M2C (Université de Rouen)	Encadrant				

Thèse dirigée par Monsieur Nicolas MASSEI, Professeur, Laboratoire M2C (Université de Rouen) et co-dirigée par Madame Anne JOHANNET, Professeure, Laboratoire LGEI (IMT Mines d'Alès)









AVANT-PROPOS

Cette thèse intitulée "Prévision de la turbidité par apprentissage statistique : application au captage AEP d'Yport" a été financée par une allocation de la Région Haute-Normandie ainsi que par l'Agence de l'Eau Seine Normandie (AESN).

Ce travail de thèse a été codirigé par Monsieur Nicolas Massei au sein de l'UMR CNRS 6143 Morphodynamique Continentale et Côtière (M2C) à l'Université de Rouen et par Madame Anne Johannet du Laboratoire de Génie de l'Environnement Industriel (LGEI) de l'IMT Mines Alès.

REMERCIEMENTS

Durant mes travaux de thèse, de nombreuses personnes sont intervenues aussi bien du point de vue du travail que du point de vue social et je tiens à tous les remercier.

Je commencerai bien évidemment par remercier mes directeurs de thèse Anne Johannet, Nicolas Massei et mon encadrant Jean-Paul Dupont. Individuellement, je remercie Anne en tant que directrice pour toute l'aide qu'elle m'a apporté et la bienveillance dont elle a fait preuve à mon égard. Je tiens à la remercier également en tant qu'amie pour les nombreuses sorties spéléo et repas passés en sa compagnie et j'en profite, par la même occasion, pour remercier toute sa famille pour les agréables moments de détente que nous avons passés ensemble. J'aimerai pouvoir remercier Nicolas Massei de la même manière mais malheureusement la distance géographique nous ayant séparé, notre amitié n'a pu se développer à un tel niveau. Je ne manquerai pas de relever cependant les agréables moments passés ensemble lors de colloques et de réunions. Je le remercie bien évidemment aussi pour son investissement dans ma thèse qui m'a permis de mener mon aventure à bien. Je tiens à remercier également Catherine Gonzales et Sylvie Spinelli pour leur aide inestimable sur l'axe produits phytosanitaires de ma thèse. J'ai également une pensée pour Vivianne Pereira-Machado qui fut l'une de mes stagiaires et qui désormais continue son travail au Brésil.

Il va de soi que je remercie l'agence de l'Eau Seine Normandie, au travers de Jerôme Ratiarson ainsi que la CODAH au travers d'Emmanuel Auchard, sans qui cette thèse n'aurait jamais pu financièrement et techniquement être possible. Je tiens également à remercier les employés de la CODAH (...) sans qui je n'aurai pas pu obtenir la majeure partie des données utilisées dans ma thèse.

Mais cette thèse ne fut pas seulement l'objet de travaux mais fut également une formidable opportunité pour me lier d'amitié avec de nombreuses personnes. Ainsi, je tiens à remercier quatre des doctorants ayant débuté leur thèse en même temps que moi. Je citerai ainsi Marie Nicole, Flavie Druine, Arnaud Venin et Aziz Saley. Les nombreux fous rires et moments de détente passés à leurs côtés m'ont permis de tenir le cap durant les moments les plus difficiles. Je tiens à remercier tout particulièrement Marie Nicole pour m'avoir hébergé durant mes nombreux déplacements sur Rouen mais surtout, pour son amitié qui m'est fort précieuse.

Je n'oublie bien évidemment pas toutes les personnes faisant partie du laboratoire M2C et les remercie de tous les moments que j'ai passés avec eux. Ayant travaillé à l'école de Mines durant plus de la moitié de ma thèse, je tiens à remercier également les personnes du LGEI et tout particulièrement Aurélie Hinsberger, Aliénor Chauvin (infiltrée du C2MA), Valentin Ledoux (infiltrée du C2MA), Nicolas Akil, Emilie Marchand, Perrine Branchet et Adrien Coutouis.

Je tiens naturellement à remercier ma famille qui m'a supporté et soutenu sans faiblir et sans jamais rien attendre en retour.

Bien évidement je remercie Emeline Bourcerie/Savary pour tout son amour qu'elle m'a témoigné depuis notre rencontre et sans qui je n'aurai pu mener ma thèse à bien.

Et enfin, je vous remercie vous, lecteurs, grâce à qui ma thèse, perdurera pour les années futures. Maintenant, il ne vous reste plus qu'environ 240 pages à lire!

SOMMAIRE

Table des figures	9
Liste des tableaux	15
Introduction	
Chapitre 1 : Importance de la mesure de la turbidité pour la qualité de l'eau souterraine	19
1.1 Introduction	21
1.2 Qualité de l'eau	
1.3 Questions scientifiques	
1.4 Conclusion du chapitre 1	
Chapitre 2 : Présentation du Site d'Yport et des données associées	37
2.1 Introduction	
2.2 Contexte regional	
2.3 Territoire d'étude : le bassin d'alimentation du captage d'Yport	
2.4 Base de données	50
2.5 Conclusion du chapitre 2	58
Chapitre 3 : Outils et Méthodes	59
3.1 Introduction	61
3.2 Analyses correlatoires	61
3.3 Analyses en ondelettes et analyses multirésolution	64
3.4 Modélisation hydrodynamique par réseaux de neurones	69
3.5 Les réseaux de neurones formels	
3.6 Analyse des produits phytosanitaires	
3.7 Conclusion du chapitre 4	82
Chapitre 4 : Analyse des données	83
4.1 Introduction	85
4.2 Comparaisons simples	85
4.3 Analyses corrélatoires	
4.4 Conclusion du chapitre 4	100
Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance	103
5.1 Introduction	105
5.2 Résumé	105
5.3 Compléments sur l'article	155
5.4 Conclusion chapitre 5	156
Chapitre 6 : Couplage de l'analyse multirésolution et des réseaux de neurones	159
6.1 Introduction	161
6.2 Analyse du résidu de prédiction	161

	6.3 Encadrement de la mise en place du modèle	165
	6.4 Résultats	168
	6.5 Conclusion du chapitre 6	174
С	Chapitre 7 : Vers un indicateur de la présence de produits phytosanitaires	177
	7.1 Introduction	178
	7.2 Echantillonnage	178
	7.3 Résultats des analyses	181
	7.4 Conclusion du chapitre 7	201
Conclusi	ion	203
Bibli	ographie	207
Anne	exe 1	217
Anne	exe 2	221
Anne	exe 3	223
Anne	exe 4	225
Anne	exe 5	227
Anne	exe 6	229
Anne	exe7	231
Anne	exe 8	243
Anne	exe 9	245
Anne	exe 10	
Anne	exe 11	257
Anne	exe 12	
Anne	exe 13	275

TABLE DES FIGURES

Figure 1: Courbes théoriques de variation du pourcentage de substances transportées sur les MES pou différentes valeurs de K_d (Voltz, M., & Louchart, X. 2001)
Figure 2 : Schéma fonctionnel du karst d'après Mangin. A. (1998)
Figure 3 : Précipitations et températures en Normandie relevées durant l'année 2015. Source : Métée France
Figure 4 : Carte de répartition des cumuls de précipitations quotidiennes (normales calculées sur la période de 1971 à 2000). Source : Météo France
Figure 5 : Coupe schématique des formations superficielles de la Seine Maritime (Chédeville, S. & <i>al.</i> 2014a) 4
Figure 6 : Schéma structural du Nord-Ouest du Bassin de Paris. (Hauchard, E. & Laignel, B. 2008) 4:
Figure 7 : Représentation des bassins d'alimentation "continu" et "discontinu" du captage d'Yport 44
Figure 8 : Carte de recensement des bétoires sur le bassin d'alimentation "continu" d'Yport4
Figure 9 : Carte d'occupation du sol sur le territoire du bassin d'alimentation du captage d'Yport (Pereira Machado, V. 2017)
Figure 10 : Carte piézométrique moyenne au voisinage du captage d'Yport (Khammari B. A. 2013)
Figure 11 : Turbidité dans le forage et turbidité à l'entrée de la station de traitement durant la période allan du 11 janvier 2015 au 25 janvier 2015
Figure 12 : Classes des mesures de turbidités pour les données allant de 2009 à 2015
Figure 13 : Evolution de la conductivité en fonction de la turbidité à l'entrée de l'usine et dans le forage 5
Figure 14 : Niveau piézométrique mesuré au niveau du captage d'Yport
Figure 15 : Niveau piézométrique du point ممرح1X0004/S1 de la commune d'Auberville-La-Renault 5
Figure 16 : Carte de répartition des pluviomètres par rapport au bassin d'alimentation du captage d'Yport
Figure 17 : Précipitations mesurées par la station de Froberville de juillet 2009 à juillet 2015
Figure 18 : Représentation du nombre de mesures pour une classe d'intensité de pluviométrie5
Figure 19 : Mesures de concentration en AMPA effectuées au point 00568X0061/P5
Figure 20 : Mesures de concentration en Atrazine effectuées au point من 68X0045/P
Figure 21 : Fonction d'autocorrélation de la turbidité mesurée à l'entrée de la station d'Yport du 1 ^{er} juille 2009 au 1 ^{er} juillet 2015.
Figure 22 : Fonction de corrélation croisée (zoom des valeurs pour les décalages allants de o et 2000) entre la pluviométrie relevée à Goderville et la turbidité mesurée à l'entrée de la station durant la période allant du 1 juillet 2009 au 1 ^{er} juillet 2015
Figure 23 : Signal composé dans sa première partie d'une sinusoïde de période 30 et d'amplitude 1 et dans sa deuxième partie d'une sinusoïde de période 15 et d'amplitude 1
Figure 24 : Décomposition en ondelettes continue (ondelette de Morlet) du signal présenté en Figure 23 La couleur rouge représente les composantes de faible puissance et la couleur jaune/blanc représente les signaux de forte puissance
Figure 25 : Schéma opérationnel de la décomposition en ondelettes discrète.
Figure 26 : Décomposition en ondelettes discrète du signal artificiel de la Figure 2368
Figure 27 : Représentation de la réponse impulsionnelle d'un système linéaire
Figure 28 : Représentation figurée d'un neurone formel

Figure 29 : Représentation schématique du perceptron multicouche.	71
Figure 30 : Schéma d'un réseau de neurones non bouclé	72
Figure 31 : Représentation schématique d'un réseau bouclé dirigé à deux couches	72
Figure 32 : Représentation schématique d'un réseau bouclé non dirigé d'ordre 1 à deux couches	73
Figure 33 : Evolution du biais (trait pointillé) et de la variance (trait continu) en fonction du nombre paramètres libres du modèle	e de
Figure 34 : Illustration du principe de validation croisée.	
Figure 35 : Évolution du score de validation croisée en fonction du nombre de neurones de la couche cach	, hée. 76
Figure 36: Représentation des critères du P_D et P_P en fonction de la prédiction (ligne en tirets) et l'observée (ligne en pointillée).	, t de 78
Figure 37 : Courbe d'accumulation des molécules en fonction du temps de déploiement c échantillonneur passif (Vrana, B. & al., 2006)	d'un 80
Figure 38 : Schémas éclaté d'un POCIS (Source : AFFINISEP)	81
Figure 39 : Equipement analytique du LGEI. Chromatographe liquide (à gauche) et spectromètre (à dro	ite). 82
Figure 40 : Précipitations cumulées pour chaque station de mesure sur la période allant de 2009 à 2015.	. 85
Figure 41 : Comparaison entre les chroniques de précipitations mesurées à Gonneville et au Havre durar période allant de juillet 2009 à janvier 2015.	nt la 86
Figure 42 : Précipitations cumulées pour chaque station de mesures pour l'évènement allant du 01/12/2	2011
au 31/12/2011	87
Figure 43 : Comparaison entre la turbidité à l'entrée de la station d'Yport (sonde 1) et les précipitati	ions
horaires moyennes	88
Figure 44 : Comparaison entre la turbidité au forage d'Yport et les précipitations moyennes cumulées	sur
10 jours.	89
Figure 45 : Comparaison entre la turbidite au forage d'i port et les precipitations moyennes comulees s	89
Figure 46 : Évolution de la hauteur d'eau dans le forage d'Yport et de la turbidité à l'entrée de l'usine (so 1) durant la période allant du 3 juillet 2013 au 5 février 2015.	nde
Figure 47 : Évolution de la hauteur d'eau dans le forage d'Yport et de la turbidité à l'entrée de l'usine (so	onde
1) durant la période allant du 11/10/2013 au 16/10/2013	90
Figure 48 : Évolution de la hauteur d'eau dans la nappe et de la turbidité mesurée à l'entrée de l'usine d'Yr (sonde 1)	port 91
Figure 49 : Évolutions de la conductivité dans le forage en fonction de la turbidité à l'entrée de l'usine (so 1).	nde 92
Figure 50 : Comparaison entre la concentration en atrazine et la turbidité à l'entrée de la station (sonde	e 1).
Figure 51 : Effet du ré-échantillonnage de la turbidité sur le modèle de l'échantillonnage des prod phytosanitaires	luits 94
Figure 52 : Maxima des corrélations croisées pluie-maximum de turbidité, en fonction du pas de ter d'échantillonnage	mps 97
Figure 53 : Représentation des 22 évènements sélectionnés	98
Figure 54: Illustration de la fonction de coût et des minimums locaux	.155
Figure 55: Illustration de l'impact du nombre de modélisation sur l'écart type des résultats de modélisat	ion.
	156

Figure 56 : Résultat de la décomposition en ondelettes continue de la turbidité observée pour l'événement 11 (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Les échelles de couleurs ne correspondent pas à des valeurs absolues mais à des valeurs relatives, on ne peut donc pas comparer les puissances respectives entre les deux graphiques, mais seulement les puissances au sein d'un même graphique.

Figure 67 : Résultats des analyses des POCIS. Les produits : de couleur bleue correspondent aux herbicides interdits d'utilisation en France, de couleur verte les herbicides autorisés à l'utilisation en France, de couleur rouge les fongicides interdits d'utilisation en France, de couleur jaune les fongicides autorisés à l'utilisation en France et en noir les insecticides autorisés à l'utilisation en France. Les noms soulignés correspondent aux produits de dégradation du produit les précédant dans la liste du graphique (lecture de gauche à droite et de haut en bas). La concentration en tébuconazole n'est pas affichée à sa vraie valeur afin de ne pas "écraser" toutes les autres valeurs.

Figure 68 : Classement des molécules suivant leur fréquence de dépassement du seuil de détection (analyse des POCIS). Figure 69 : Classement des produits suivant leur fréquence de dépassement du seuil de quantification (analyse des POCIS).

(unuryse des		
Figure 70	o : Nombre d'analyses supérieures aux seuils de détection et de quantification su	r la totalité des
déploiement	S	189

Figure 71 : Visualisation des précipitations mesurées au moment de l'échantillonnage et la concentratio trouvée via l'analyse des échantillons d'eau19	n
Figure 72: Coefficient de corrélation entre les précipitations et les concentrations en produit phytosanitaires trouvée via l'analyse des échantillons d'eau19	:S
Figure 73 : Visualisation de la turbidité mesurée au moment de l'échantillonnage et la concentration trouvé	e
via l'analyse des échantillons d'eau19	2
Figure 74: Coefficient de corrélation entre la turbidité et les concentrations en produits phytosanitaire trouvée via l'analyse des échantillons d'eau	:s 3
Figure 75 : Conductivité mesurée au moment de l'échantillonnage et concentration trouvée via l'analyse de échantillons d'eau	:S 4
Figure 76 : Coefficient de corrélation entre la conductivité et les concentrations en produits phytosanitaire trouvée via l'analyse des échantillons d'eau. 19	s 5
Figure 77 : Visualisation des précipitations cumulées mesurées durant les déploiements et le concentrations trouvées via l'analyse des POCIS.	:s 7
Figure 78 : Coefficient de corrélation entre les précipitations et les concentrations en produit phytosanitaires via l'analyse des POCIS	:s 7
Figure 79 : Visualisation de la turbidité mesurée durant les déploiements et les concentrations trouvées vi l'analyse des POCIS.	а 8
Figure 80 : Coefficient de corrélation entre la turbidité et les concentrations en produits phytosanitaire trouvée via l'analyse des POCIS	:s 9
Figure 81 : Visualisation de la conductivité moyenne mesurée durant les déploiements et les concentration trouvées via l'analyse des POCIS	15 0
Figure 82 : Coefficient de corrélation entre la conductivité et les concentrations en produits phytosanitaire trouvées via l'analyse des POCIS	25 0
Figure 83: Illustration de l'évènement 123	1
Figure 84: Illustration de l'évènement 223	1
Figure 85: Illustration de l'évènement 3 23	2
Figure 86: Illustration de l'évènement 4 23	2
Figure 87: Illustration de l'évènement 523	3
Figure 88: Illustration de l'évènement 623	3
Figure 89: Illustration de l'évènement 7 23	4
23 Figure 90: Illustration de l'évènement 8	4
Figure 91: Illustration de l'évènement 923	5
Figure 92: Illustration de l'évènement 1023	5
Figure 93: Illustration de l'évènement 11	6
Figure 94: Illustration de l'évènement 12	6
Figure 95: Illustration de l'évènement 13	7
Figure 96: Illustration de l'évènement 14	7
۔ 23 Figure 97: Illustration de l'évènement 15	8
Figure 98: Illustration de l'évènement 16	8
۔ Figure 99: Illustration de l'évènement 17	9
Figure 100: Illustration de l'évènement 18	9 9
۔ 24 Figure 101: Illustration de l'évènement 19	0

Figure 102: Illustration de l'évènement 20 2	240
Figure 103: Illustration de l'évènement 21 2	241
Figure 104: Illustration de l'évènement 22 2	241
Figure 105 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénemen	de 1t 2
Figure 106 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénemen 2	de 1t 3 245
Figure 107 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénemen 2	de 1t 6 246
Figure 108 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement	de : 10 246
Figure 109 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement	de t 11 247
Figure 110 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement	de : 16 247
Figure 111 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement	de : 22 248
Figure 112 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue pa modèle dirigé à deux branches pour la prédiction à 12h de l'événement 2. (a) comparaison entre turbidité préd et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et N 	r le lite M3. 249
Figure 113 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue pa modèle dirigé à deux branches pour la prédiction à 12h de l'événement 3. (a) comparaison entre turbidité préd et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et N 	r le lite M3. 250
Figure 114 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue pa modèle dirigé à deux branches pour la prédiction à 12h de l'événement 6. (a) comparaison entre turbidité préd et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et N	r le lite V3. 251
Figure 115 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue pa modèle dirigé à deux branches pour la prédiction à 12h de l'événement 10. (a) comparaison entre turbid prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 M3.	r le lité 2 et 252
Figure 116 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue pa modèle dirigé à deux branches pour la prédiction à 12h de l'événement 11. (a) comparaison entre turbid prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 M3.	r le lité ₂ et 253
Figure 117 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue pa modèle dirigé à deux branches pour la prédiction à 12h de l'événement 16. (a) comparaison entre turbid	r le lité

rédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et
13
Figure 118 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le
nodèle dirigé à deux branches pour la prédiction à 12h de l'événement 22. (a) comparaison entre turbidité
rédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et
13

LISTE DES TABLEAUX

Tableau 1 : Tableaux récapitulatif des travaux effectués sur la prévision de la turbidité par des modèles
statistiques. Pour les modèles, les abréviations; MP, NN, NF, MLR, SRC, SVM, GEP, PSO correspondent
respectivement a; Modèle probabiliste, Neural Network, Neurone Flous, Multilinear Regression, Sediment
Rating Curve, Support Vector Machine, Genetic Expression Programming, Particul Swarm Optimisation. Pour
les pas de temps, les abréviations M, J et H représentent respectivement des pas de temps Mensuel, Journalier
et Horaire
Tableau 2 : Tableaux récapitulatif des valeurs maximales, minimales et moyennes de turbidité, de
précipitation et de hauteur d'eau dans le forage, observées de 2010 à 2014
Tableau 3 : Normes AFNOR régissant l'échantillonnage dans le cadre du suivi de la qualité des masses d'eau
(Ibrahim, I., 2013)
Tableau 4 : Récapitulatif des maxima des corrélations croisées interpluviomètres, au pas de temps horaire,
sur la période allant de juillet 2009 à janvier 2015
Tableau 5 : Récapitulatif des temps de réponses sur la période allant de juillet 2009 à janvier 2015. A noter
que les chroniques de pluie de Gonneville et du Havre ont une majeure partie commune
Tableau 6 : Synthèses des caractéristiques des 22 évènements sélectionnés
Tableau 7 : Synthèse des analyses corrélatoires au pas de temps horaire pour chaque évènement
Tableau 8 : Tableau récapitulatif des architectures des modèles non-dirigés168
Tableau 9 : Résultats des critères de Nash, de pourcentage de restitution (P_p) et de délais du pic (P_d) pour
les différents modèles à deux branches non-dirigé avec un horizon de prévision de 12h170
Tableau 10 : Tableau récapitulatif des paramètres des modèles dirigés
Tableau 11 : Résultats des critères de Nash, de pourcentage de restitution (P_p) et de retard du pic (P_d) pour
les différents modèles à deux branches dirigés avec un horizon de prévision de 12h
Tableau 12 : Résultats des critères de Nash pour les différents modèles à deux branches avec un horizon de
prévision de 12h
Tableau 13: Résultats des critères de pourcentage de pic (P_P) pour les différents modèles à deux branches
avec un horizon de prévision de 12h
Tableau 14: Résultats des critères de pourcentage de pic (<i>P</i> _d) pour les différents modèles à deux branches
avec un horizon de prévision de 12h

INTRODUCTION

L'eau est une des ressources nécessaires à la vie sur Terre. Dans le monde, près de 25% de l'alimentation en eau potable des populations se fait par l'intermédiaire d'exploitations sur l'eau provenant du karst souterraine. Malgré sa plus faible sensibilité aux pollutions en comparaison avec l'eau de surface, l'eau souterraine n'est pas épargnée par les contaminations, que les gestionnaires de forage tentent de diminuer au maximum.

La région Haute-Normandie, de par ses caractéristiques climatiques, géologiques et anthropiques présente une hypersensibilité de l'eau souterraine face aux contaminations. Avec 70% de la superficie du sol de la région employée pour l'agriculture, elle se place dans les trois régions avec la part de territoires agricoles la plus importante (devant le Les Hauts-de-France et les Pays de la Loire). Parmi ces terrains dédiés à l'agriculture, on recense 55% de terres arables (fortement disponibles à l'érosion) et 45% de parcelles enherbées. Seuls 17% des terrains sont boisés, et sont ainsi moins affectés par le ruissellement (AGRESTE 2015). Une podologie particulière causée par le couplage de l'exploitation agricole et de la présence d'argiles dans le sol entraîne la formation d'une couche imperméable (croûte de battance) qui augmente fortement le processus de ruissellement. Cette imperméabilisation est couplée avec la présence de "bétoires" qui véhiculent l'eau de ruissellement directement dans le système hydrogéologique souterrain. Cette association entre imperméabilité et présence de bétoires est responsable de la forte proportion de forages contaminés par de l'eau trouble. L'association entre cette vulnérabilité de l'eau souterraine et les activités anthropiques induit une possibilité accrue de pollution des eaux destinées à l'alimentation en eau potable par des contaminants tels que les produits phytosanitaires, les bactéries, les hydrocarbures,....

Il apparait inévitable que cette sensibilité de l'eau souterraine induise des conséquences socioenvironnementales importantes. Dans ce contexte, 50% des infrastructures de production d'eau potable en Haute-Normandie présentent une pollution par des matières en suspension, induisant l'arrêt temporaire (quelques jours par an) de l'alimentation en eau potable de 1 000 à 2 000 habitants.

La présence de matières en suspension, qui peut être mise en relation avec la présence de produits phytosanitaires, implique la présence de ces derniers dans l'eau de consommation de façon chronique pour 50 000 habitants, ou ponctuellement pour près de 100 000 habitants (ARS 2013). Depuis la fin des années 90, de nombreuses stratégies et outils ont été développés afin de limiter les risques de contamination des populations par le biais de l'eau potable.

Dans ce contexte, en tant que paramètre intervenant dans les mesures de qualité de l'eau, la turbidité est un des éléments clés des suivis qualitatifs de la ressource en eau et permet rapidement de quantifier les matières en suspension dans l'eau. Ainsi, la valeur de 1 NTU est utilisée comme limite pour les eaux destinées à l'alimentation en eau potable. Cependant, les processus de traitement de la turbidité ne permettent pas de cibler et d'éliminer les produits phytosanitaires libres dans l'eau. Actuellement, le lien entre turbidité et produits phytosanitaires n'est pas clairement établi. Ce lien se trouve fortement affecté par les pratiques agricoles ainsi que par la non-linéarité des systèmes hydrogéologiques et nécessite une étude propre à chacun de ces systèmes. Ainsi, dans le cas du bassin versant d'Yport sur lequel nous avons focalisé nos travaux, cette thèse se propose, d'une part, d'effectuer la prévision de la turbidité au captage d'Yport grâce à une méthode qui permet de s'affranchir de la prise en compte fine des processus, les réseaux de neurones artificiels, et d'identifier la relation éventuelle entre les produits phytosanitaires et la turbidité et avec d'autres variables telles que les précipitations et la conductivité électrique. Ainsi une de ces variables pourrait être utilisée comme PROXY de la présence de produits phytosanitaires afin de permettre une meilleure gestion qualitative de la ressource avec un coût modéré pour les analyses d'eau.

Le premier chapitre a pour intérêt de faire un état des lieux de la compréhension et de la modélisation de la turbidité mais également de faire un historique de l'évolution de la législation face à la qualité de l'eau potable vis-à-vis des produits phytosanitaires. Ce chapitre nous permet de montrer l'intérêt de l'utilisation des réseaux de neurones pour une telle prédiction mais aussi de montrer l'aspect novateur de notre travail.

Le second chapitre a pour objectif de faire une description des différentes méthodes que nous avons utilisées pour nos travaux. Ainsi le lecteur aura à sa disposition toutes les clés lui permettant une bonne compréhension des différents chapitres.

Le troisième chapitre s'attache à décrire notre site d'étude et mettre en avant les caractéristiques de ce dernier. Le lecteur pourra ainsi juger du fonctionnement atypique d'un tel système et entrevoir les enjeux sociétaux et scientifiques que présente notre travail.

Le vif du sujet débute dans le chapitre quatre avec la présentation des analyses préliminaires des données à notre disposition. Nous y dressons un panorama des caractéristiques hydrogéologiques de notre bassin, indispensables, à la suite de notre travail.

Par la suite, grâce à ce travail préliminaire, nous nous employons dans le chapitre 5 à de réaliser des modèles prédictifs par réseaux de neurones. La question à laquelle nous tenterons d'apporter une réponse est la suivante : avec quelle précision est-il possible de prévoir les arrivées d'eaux troubles dans un système souterrain à l'aide des réseaux de neurones ? Deux architectures de modèles seront mises en concurrence afin d'obtenir les meilleurs résultats possibles pour une prévision de la turbidité 12h et 24h en avance. Ce chapitre sera présenté sous la forme d'un article soumis à la revue "*Journal of Hydrology*".

Un travail scientifique sur la recherche d'amélioration des précédents modèles sera exposé dans le chapitre 6. Nous proposerons alors d'associer l'analyse multirésolution, inspirée de la théorie des ondelettes, avec les réseaux de neurones. Le choix de la décomposition de la turbidité y est explicité et la méthode de décomposition justifiée.

Enfin le chapitre 7 présente la totalité des travaux portant sur les produits phytosanitaires ayant été effectués durant les trois ans de thèse. Le contenu de ce chapitre portera sur; (i) la mise en avant d'une nouvelle méthode d'échantillonnage (POCIS), (ii) l'identification d'un PROXI des produits phytosanitaires.

CHAPITRE 1 : IMPORTANCE DE LA MESURE DE LA TURBIDITE POUR LA QUALITE DE L'EAU SOUTERRAINE

1.1 INTRODUCTION

Ce chapitre, à partir d'une description "historique", se propose de décrire les besoins et les réglementations encadrant l'alimentation en eau potable ainsi que leurs relations avec les qualités respectives de l'eau souterraine et de l'eau de surface. Après cette présentation, nous nous intéressons aux méthodes de mesure de la turbidité, en tant que telles, et en tant que marqueur d'autres pollutions : physico-chimiques et bactériologiques. Ce double rôle supposé de la turbidité nous permet de positionner les questions scientifiques de ce travail de thèse. Afin d'y répondre, nous présentons ensuite les études de modélisation de la turbidité au moyen de modèles physiques, probabilistes ou par apprentissage statistique. Dans la troisième partie, nous nous intéressons au comportement des produits phytosanitaires, et à leurs interactions avec la turbidité. Un focus sur les écoulements karstiques est alors proposé afin d'envisager les comportements spécifiques à ces aquifères.

1.2 QUALITE DE L'EAU

1.2.1 EVOLUTION HISTORIQUE DE LA NOTION DE QUALITE DE L'EAU

Au XVII^e siècle, le jugement de la qualité de l'eau s'effectuait sur des critères visuels (transparence, couleur,...) et organoleptiques (odeur, goût,...). De nos jours, cette approche rapide et peu coûteuse est toujours utilisée par tout un chacun, mais ne permet pas de déterminer avec précision la qualité de l'eau. En effet, certains micro-organismes ayant une taille de l'ordre du micromètre (o,5 à 3 µm par exemple pour Escherichia Coli), il n'est pas possible d'appréhender leur présence à l'œil nu. De même, une eau troublée par seulement des particules minérales pourra être consommable sans risque sanitaire. Egalement, si l'on considère les critères organoleptiques, une eau sans odeur pourrait très bien contenir de l'arsenic, qui est inodore, mais considéré comme cancérigène (Smith, A. H. & *al.* 2003), alors qu'une eau contenant du chlore (comme l'eau du robinet) peut être odorante, mais propre à la consommation.

Face à la complexité de la définition d'une eau potable, que nous avons rapidement illustrée précédemment, et suite au développement de l'approvisionnement public en eau potable, il est apparu nécessaire de déterminer les critères permettant d'attester de la qualité d'une eau. C'est en 1902 avec la création du conseil supérieur d'hygiène publique que s'est enclenchée la réflexion pour préciser la notion de qualité de l'eau. C'est bien avant cela (1853) qu'apparait la première société de distribution de l'eau potable sous le nom de compagnie Générale des Eaux, puis en 1880 la Lyonnaise des eaux et enfin la SAUR en 1933 (Haghe, J. P., 1998). En 1854, John Snow, montre déjà que la transmission du choléra à Londres se fait par l'intermédiaire de l'eau contaminée (Joy, R. J., 2004) ; en 1880, Pasteur, Koch et Erbech montrent que 90% des maladies proviennent de la consommation de l'eau (Roche, P. A. 2002). Vers la fin du XIXe siècle, les techniques d'analyses chimiques et bactériologiques permettent de mettre en avant des seuils de contamination de l'eau, au-delà desquels la santé humaine peut être impactée. Durant cette même période, les travaux de l'institut Pasteur permettront d'éliminer les micro-organismes de l'eau, jusqu'à présent ne subissant qu'une décantation, en améliorant le traitement avec la mise en place des filtres à sable associés à de la coagulation/décantation (Husson, G. P., 1996). C'est depuis les années 1950 que des normes quantitatives ont pu être développées, en particulier grâce aux progrès de l'épidémiologie. Il est ensuite apparu que la contamination anthropique de l'eau (par les industries, l'agriculture,...) entrainait un besoin grandissant de traitement de l'eau pour atteindre les normes de potabilité. Les traitements chimiques font alors leur apparition au début du XXe siècle et les techniques de désinfection par l'ozone et le chlore sont mises en place. En France, c'est d'ailleurs après la première querre mondiale que la désinfection par le chlore (appelée verdunisation) se généralise. Le développement des techniques de traitement de l'eau permettant une plus grande consommation d'eau potable, la loi sur l'eau du 16 décembre 1964 instaurera une gestion de l'eau répartie suivant les bassins hydrographiques. Cette loi fait également apparaître, pour la première fois, le principe fondamental de pollueur-payeur. La délimitation des bassins hydrographiques avec l'observation des lignes de partage des eaux définit ainsi les 6 grands bassins

actuels : Artois-Picardie, Rhin-Meuse, Seine-Normandie, Loire-Bretagne, Adour-Garonne et Rhôneméditerranée-Corse. La gestion de ces bassins est administrée par des agences financières de bassin que l'on connaît actuellement sous le nom d'Agences de l'eau. En 1980 la quasi-totalité des Français pourront bénéficier de l'eau potable dite "courante". Le décret nº89-3 du 3 janvier 1989 fixe donc les normes de qualité de l'eau destinée à la consommation humaine. Ce décret met également en place une norme sur la qualité de l'eau potable ainsi que l'instauration d'une autorisation de prélèvement et de normes d'hygiènes pour les installations de distribution de l'eau potable. En 1992, la loi sur l'eau du 3 janvier 1992 (Article L-1321-2 du code de la santé public) instaure, pour les prélèvements de plus de 100 m³/j, les périmètres de protection des captages (PPC) permettant de limiter l'impact anthropique sur la ressource en eau. Ces PPC sont au nombre de trois ; le Périmètre de Protection Immédiat (PPI), dans lequel toutes les activités humaines sont interdites (sauf entretien des installations de captage); le Périmètre de Protection Rapproché (PPR), plus étendu, dans leguel les activités polluantes sont proscrites ; et enfin le Périmètre de Protection Eloigné (PPE), qui est facultatif et qui peut renforcer la protection d'un bassin vulnérable (par exemple karstique) sur toute l'étendue de son bassin d'alimentation (Durand, F. & Petit, V. 1997; Petit, V. & al. 1998; Muet, P. & Vier, E. 2006). Cette loi du 3 janvier 1992 fait également apparaitre les Schémas Directeurs d'Aménagement et de Gestion des Eaux (SDAGE) propres à chaque bassin. Par la suite, l'arrêté du 11 janvier 2007 permet d'établir, dans un premier temps, les limites de qualité et de références des eaux destinées à la consommation humaine (Annexe 1 et 2). Cet arrêté définit également les limites de qualité des eaux brutes de toutes origines utilisées pour la production d'eau destinée à la consommation humaine (Annexe 3). Enfin l'arrêté du 11 janvier permet la mise en place d'une grille définissant les limites de qualité des eaux douces superficielles utilisées pour la production d'eau destinée à la consommation humaine (Annexe 4). De nos jours, c'est l'Europe, avec la Directive Cadre sur l'Eau (DCE) de 2000 qui impose les normes de qualité relatives à l'eau souterraine et de surface. C'est alors la loi du 21 avril 2004 qui permettra d'acter dans le droit français les dispositions mises en place par la DCE.

1.2.2 LES MOYENS RAPIDES DE JUGER DE LA QUALITE DE L'EAU

Suites aux évolutions concernant la détermination de la qualité de l'eau depuis le XVIIe siècle les gestionnaires des sites de production d'eau potable ont donc à leur disposition des normes établies afin d'éviter la contamination des populations qu'ils approvisionnent. Les nombreux paramètres, bien que nécessaires à la détermination de la qualité de l'eau, ne peuvent cependant pas tous être analysés en temps réel. La qualité de l'eau est évaluée régulièrement : une fréquence est ainsi définie par l'arrêté du 21 janvier 2010 suivant le débit de production et le nombre d'habitants desservis par l'usine). Afin de pouvoir identifier de possibles contaminations susceptibles de nuire à la santé humaine, des dispositifs d'analyses en temps réel sont mis en place dans les usines de production d'eau potable. On notera par exemple la présence de poisson tel que l'Apteronotus albifrons (usine d'eau potable vichyssoise) dans un bac alimenté par l'eau sortant de la station. Ces poissons, lorsqu'ils deviennent malades, ou meurent, permettent d'effectuer un test de toxicité, permettent d'arrêter la production de l'eau dans le cas d'une légère variation physico-chimique de l'eau produite par la station. Egalement, la turbidité qui représente la non-transparence de l'eau, est prise comme indicateur de la présence de particules en suspension, et peut être un révélateur de la présence de micro-organismes si l'origine de la turbidité et de la présence de micro-organismes est commune. La simplicité de sa mesure a conduit à la définition d'un seuil d'une valeur de 1 NTU (nephelometric turbidity unit) en sortie de station. La mesure de turbidité est donc un outil décisionnel pour la mise en place de traitements tels que les filtres à sable ou bien la quantité de produit utilisé pour la floculation/décantation.

1.2.3 LA TURBIDITE

La turbidité est un phénomène physique, lié à la présence de matières en suspension. Elle est généralement appréhendée par son effet sur l'opacité d'une solution (Bin Omar, A. F. & al. 2009). La turbidité est en effet définie comme étant la capacité d'une solution à se laisser traverser par un rayon lumineux (Glysson, G. D., &

Gray, J. R. 2002). La turbidité s'exprime en NTU (*Nephelometric Turbidity Unit*) ou en NFU (*Nephelometric Formazine Unit*). On peut noter que la relation entre NFU et NTU n'est pas linéaire :

- Pour des turbidités inférieures à 20 NTU, 1NTU = 1 NFU
- Pour des turbidités supérieures à 20 NTU, 1 NFU = 0,6 NFU

La mesure de la turbidité peut être effectuée par différentes méthodes telles que :

- la transmissométrie et néphélométrie
- la diffraction laser
- la rétrodiffusion acoustique

Chacune de ces méthodes possède ses avantages et ses inconvénients qui sont présentés ci-après.

LA TRANSMISSOMETRIE ET LA NEPHELOMETRIE

Les turbidimètres peuvent être classés dans deux catégories ; les transmissomètres et les néphélomètres. Les transmissomètres mesurent la lumière transmise après passage au travers d'une solution. Les néphélomètres s'intéressent à la lumière rétrodiffusée par les particules en suspension.

Les méthodes de mesure relatives à la transmission du rayon lumineux sont généralement utilisées dans le cadre de l'étude de la turbidité de l'air. Cependant, quelques appareils peuvent être utilisés dans les milieux aquatiques (Holdaway, G. P. & *al.* 1999). La mesure par transmissométrie utilise généralement un rayonnement lumineux de longueur d'onde 665 nm (Holdaway, G. P. & *al.* 1999). D'un point de vue de la concentration, la qualité des mesures pour cette méthode est assurée pour des concentrations comprises entre 0,1 et 500 mg/L. Cependant, cette mesure n'est pas univoque : par exemple, il a été montré que des matières en suspension de dimension de l'ordre de 3 µm à une concentration de 0,7 mg/L atténuaient autant la lumière que des matières en suspension de dimension de l'ordre de 8 µm pour une concentration de 9,8 mg/L (Bunt, J. A. & *al.* 1999). La forme des particules joue également un rôle dans la capacité d'arrêter le rayon lumineux. Ainsi, à dimension équivalente, une matière en suspension de forme plate a moins de chance d'atténuer le rayon lumineux qu'une particule sphérique.

Les méthodes relatives à l'utilisation de la rétrodiffusion sont plus nombreuses et plus « faciles » à mettre en œuvre dans le karst. Parmi celles-ci, le capteur le plus communément utilisé est l'OBS (Optical Backscatter Sensor) (Downing, J. P. & al. 1981). Ce dernier mesure les rayonnements lumineux rétrodiffusés compris entre 700 nm et 900 nm. Il fonctionne de la manière suivante : dans un premier temps, un rayon lumineux est émis depuis le capteur ; lorsque le rayon lumineux interagit avec une particule en suspension, une partie de ce rayonnement est difractée/rétrodiffusée. Les rayonnements rétrodiffusés selon un angle allant de 90° à 180° par rapport à l'angle du rayon d'origine sont captés (Gray, J. R. & al. 2009). C'est cette mesure qui est convertie en termes de turbidité. Ce type de sonde est relativement facile à utiliser car de taille petite, elle peut donc être facilement intégrée dans un conduit karstique. De plus, cette dernière est généralement équipée d'un balai mécanique servant au nettoyage régulier de l'instrument. Cependant, ce type d'instrument présente une limitation concernant la concentration en matières en suspension présentes dans le milieu (Gardner, W. D. & al. 1985). En effet, il a été montré que pour des concentrations en matières en suspension inférieures à 2 g/L, l'OBS avait une réponse globalement linéaire, mais au-delà de cette concentration, la relation n'est plus linéaire car les rayonnements rétrodiffusés interagissent avec d'autres particules avant d'atteindre le capteur. Ce capteur a des performances différentes en fonction de la taille et de la forme des particules. Suivant la taille des particules, l'équation de rétrodiffusion n'est pas la même, on applique deux théories différentes qui sont : la théorie de Mie pour les particules de grandes tailles est la théorie de Rayleight pour les particules de petites tailles (Lawler, D. M. 2005). En effet, plus la taille des particules et grande, plus le rayonnement rétrodiffusé est faible (Bunt, J. A. & al. 1999). On note alors l'importance de la calibration de l'OBS en fonction d'une classe de taille de particules donnée. Si l'on calibre l'OBS avec de petites particules, on risque de ne pas prendre en compte les particules de grosse taille responsables de la turbidité. Ainsi, les mesures effectuées dans un milieu où la taille des particules varie sont moins fiables (Creed, E. L. & al. 2001). La forme des particules joue également un rôle dans la mesure (Bunt, J. A. & al. 1999), bien que moins important que tous les autres facteurs décrits (Hatcher, A. & al. 2000).

En effet, des particules de forme sphérique rétrodiffusent préférentiellement la lumière selon un angle de 120° alors que des particules plates ou de forme carrée rétrodiffusent préférentiellement la lumière selon un angle de 90° à 100°. La nature des particules (biologique ou minérale) influence aussi la mesure, par exemple, dans le cas de matières en suspension organiques, des phénomènes de floculation/défloculation peuvent se produire et faire varier la taille des particules en suspension. Les propriétés optiques des matières en suspension peuvent également entrainer une réponse différente en fonction de la longueur d'onde utilisée pour la mesure (Bowers, D. G., & Binding, C. E. 2006 ; Hatcher, A. & al. 2000). Dans le cas d'une non stationnarité relative au type de matière en suspension, une re-calibration du capteur est donc nécessaire.

En synthèse, le transmissomètre est un instrument facile à utiliser et à installer mais sa réponse dépend du type de matières en suspension. Il ne peut alors pas donner des mesures fiables dans un environnement où la taille et la concentration en matières en suspension varient. Les méthodes relatives à l'utilisation de la rétrodiffusion, en particulier l'OBS, sont plus faciles à mettre en place en milieux naturel et délivrent une mesure directement utilisable dans le cadre de l'étude de la qualité de l'eau, mais souffrent de devoir être re-calibrées en fonction de l'évolution des propriétés physico-chimiques des matières en suspension (Gartner, J. W. 2004).

DIFFRACTION LASER

La méthode de diffraction laser LISST (Laser In-Situ Scattering Transmissometry) se base sur la diffraction laser. Elle délivre la distribution des tailles des matières en suspension mais pas la turbidité, l'information est donc complémentaire à celle délivrée par un turbidimètre. Il existe trois catégories de LISST ; Le LISST-100B qui mesure les particules entre 1,25 et 250µm, le LISST-100C qui mesure les particules entre 2,5 et 500 µm et le LISST-FLOC qui mesure les particules entre 7,5 et 1500 µm. Le LISST utilise un rayonnement lumineux avec une longueur d'onde de 670 nm. Il utilise en tout 32 anneaux de détection pour mesurer la lumière renvoyée par une solution, ce qui permet la classification des matières en suspension selon 32 classes de taille de particules (Gray, J.R., & Gartner, J.W. 2009).

Pour cet instrument, la calibration joue un rôle très important. La calibration doit être réalisée en utilisant comme bornes les tailles maximales et minimales des particules en suspension afin d'utiliser au mieux les capacités d'analyses de l'instrument (Jago, C. F. & *al.* 2000). De plus, le LISST fonctionne de manière optimale pour des particules en suspension de forme ronde et lisse car elles renvoient la lumière selon des angles constants, quelles que soient leurs orientations (Bunt, J. A & *al.* 1999). Le LISST n'étant pas équipé d'un système d'auto-nettoyage de son capteur, son entretien régulier en fait une méthode difficilement utilisable en milieu naturel. Pour finir, le LISST est un appareil qui a été développé afin de mesurer la distribution des tailles des particules en suspension au niveau des zones côtières. Par conséquent, ce dernier est relativement encombrant et coûteux. Il est donc beaucoup plus difficile de s'en servir lorsque l'on cherche à étudier les particules en suspension présentes dans un conduit karstique.

En synthèse, le LISST est un instrument performant, capable de mesurer de nombreuses informations (distribution des tailles et concentration des particules en suspension). Cependant, sa mise en place au sein d'un système karstique est limitée par sa taille et par la nécessité d'entretiens réguliers. De plus, la calibration doit s'adapter au type de particules, limitant ainsi sa fiabilité lors de modifications du milieu (Bishop, J. K. 1986).

RETRODIFFUSION ACOUSTIQUE

La méthode de rétrodiffusion acoustique se base sur le fonctionnement du sonar. Parmi ces méthodes, l'ADCP (*Acoustic Doppler Current Profiler*) est l'un des capteurs les plus utilisés. Le capteur mesure une onde sonore rétrodiffusée, le déphasage entre l'onde émise et l'onde réfléchie donne la distance, tandis que la puissance reçue indique la turbidité. Il est possible de réaliser des profils de particules en suspension en fonction de la profondeur (Holdaway, G. P. & *al.* 1999). Dans cette méthode, la fréquence de l'onde peut être adaptée à la taille des matières en suspension étudiées. Contrairement aux autres méthodes, en réalisant des profils de profondeur, cette méthode permet de minimiser les risques de mesurer une évolution localisée de la concentration en matières en suspension. L'ADCP permet également de réaliser des profils de vitesse d'écoulement. Cette méthode n'utilisant pas de rayonnement lumineux pour la mesure, son encrassement par

des algues est limité et diminue les besoins d'entretien. L'un des points faibles de l'ADCP vient de la nécessité d'une mesure de la turbidité, ou de la concentration en matière en suspension, en parallèle, afin de convertir la mesure acoustique réalisée. De plus, cette méthode est sensible aux variations des propriétés physicochimiques de l'eau (température, pression, salinité). Le contrôle de ces paramètres se trouve alors indispensable pour prévenir toute dérive de la mesure. Comme pour l'OBS, il a été montré que pour des concentrations en matières supérieures à 10g/L, la relation entre onde réfléchie et concentration en matières en suspension n'est plus linéaire car les rayonnements rétrodiffusés interagissent avec d'autres particules avant d'atteindre le capteur (Glysson, G. D., & Gray, J. R. 2002). Comme pour les autres méthodes de mesure, la calibration (comprenant également le choix de la fréquence des ondes acoustiques) joue un rôle important. Si l'on prend l'exemple d'un ADCP réglé avec une fréquence de 1 200 kHz, seules les particules de tailles inférieures ou égales à 400 µm pourront être détectées. Ainsi, plus la fréquence est ajustée par rapport à la taille des particules en suspension présentes dans le milieu, plus la sensibilité de la mesure sera élevée.

En synthèse, L'ADCP permet de mesurer des profils de concentration en matières en suspension et des vitesses d'écoulement. Cependant, sa dépendance à un autre système de mesure des matières en suspension, le rend sensible aux erreurs de mesures pouvant être faites par ces derniers. De plus, le choix de la fréquence des ondes acoustiques lors de la calibration peut affecter grandement la qualité des mesures.

ORIGINE ET NATURE DE LA TURBIDITE

Grâce à l'étude de la turbidité, couplée à la mesure de différents paramètres chimiques de l'eau, des études ont montré qu'il était possible de différencier l'origine des matières en suspension. En se servant des chroniques de turbidité associées aux chroniques de conductivité et de débit, Massei, N. & *al.* (2003) ont utilisé une méthode servant dans l'interprétation des tests de traçage pour différencier les apports autochtones (remise en suspension) et les apports allochtones (transfert en provenance de la surface). Shevenell L. & *al.* (2002) ont associé la mesure de la turbidité avec les mesures de pH, de carbone organique total, de température et d'oxygène dissout. Atteia, O., & Kozel, R. (1997) ont associé la mesure des tailles de particules avec le pH, pour identifier les particules issues d'une remise en suspension, de celles issues directement de la surface. Il a également été montré que la turbidité pouvait être utilisée comme traceur des écoulements de l'eau souterraine en appliquant des méthodes statistiques telles que des corrélations. (Massei, N. & *al.* 2003; Massei, N., & *al.* 2006). Amraoui, F., & *al.* (2003) ont montré la possibilité d'une remise en suspension des dépôts contenus dans le karst, celle–ci se faisant à la faveur d'une augmentation rapide du courant comme mécanisme initiateur de la remise en suspension (Kneller, B., & Buckee, C., 2000).

Ainsi, les matières en suspension peuvent être de deux types ; minérales ou organiques. La matière minérale peut être caractérisée dans le cas d'un aquifère possédant différentes structures géologiques afin d'identifier les origines hydrogéologiques des arrivées d'eaux (Herman, E. K. & *al.* 2007 ; Dupont J-P., & *al.* 2001). Pour les matières organiques, il a été montré que celles-ci pouvaient être utilisées comme un indicateur de contamination bactériologique. Pour cela, Dussart-Baptista, L., & *al.* (2003) ont analysé la concentration en bactéries avant l'entrée dans le système karstique puis la turbidité et la concentration en bactéries au niveau d'un forage. Ils ont démontré que l'arrivée d'eau turbide était associée à l'arrivée de bactéries, même si la concentration en bactéries était moins forte au niveau du forage (due à des phénomènes de dilution). Ils ont également montré que la concentration en bactéries pouvait augmenter même si la turbidité n'augmentait pas. Mahler, B. J. & *al.* (2000) ont effectué le même type d'analyse et ont montré que durant une période sèche, 50% des bactéries étaient associées à de la matière en suspension alors qu'après un évènement pluvieux, 100% des bactéries (de type E. Coli) avaient plus de chances d'être associées à des particules de petites tailles (entre o,9 µm et 2 µm).

En synthèse, il apparaît donc que la turbidité, à elle seule, ne peut permettre d'identifier son origine ; elle doit impérativement être couplée à d'autres mesures physico-chimiques. Enfin la relation entre la turbidité et la concentration en micro-organismes apparaît comme fortement non-linéaire et influencée par diverses variables comme le débit.

1.3 QUESTIONS SCIENTIFIQUES

Comme présenté précédemment, II apparait que la qualité de l'eau, physico-chimique et bactériologique, est soumise à des variations rapides d'origines complexes. Ces variations ne peuvent pas faire l'objet d'une analyse en continu à un coût raisonnable. En revanche, la turbidité est une mesure intervenant dans les normes de qualité de l'eau potable, qui peut être mesurée facilement et en continu sans engager de fortes dépenses. La turbidité est également un indicateur de contamination de l'eau souterraine par l'eau de surface au cours des épisodes pluvieux. Cependant la turbidité apparaît comme le résultat de nombreux phénomènes non linéaires difficilement mesurables. La turbidité peut être entrainée par l'eau de surface, remobilisée en souterrain, ou mal estimée par sa mesure qui dépend de manière non univoque des caractéristiques des matières en suspension. Egalement, la capacité de la turbidité à être le révélateur d'autres pollutions, bactériologiques comme physicochimique apparaît comme non linéaire et fortement sensible à l'état du bassin considéré (le débit de la rivière par exemple).

Ainsi, au vu de la présentation précédente, et en regard avec l'importance de la turbidité dans les normes de potabilité de l'eau, deux questions scientifiques se dégagent : la première concerne la capacité à modéliser et à prévoir la turbidité au sein d'un aquifère, afin d'anticiper sur les traitements possibles, la seconde s'intéresse à la possibilité d'utiliser la turbidité comme indicateur de pollutions bactériologiques ou physico-chimiques.

1.3.1 MODELISATION DE LA TURBIDITE

Compte tenu de l'enjeu, plusieurs études se sont intéressées à la prévision de la turbidité, mais compte tenu des difficultés précédemment soulevées, assez peu d'études ont été publiées. Compte tenu du manque d'informations sur les propriétés et caractéristiques physiques du sol et du sous-sol, les modèles physiques seuls ne peuvent généralement pas être utilisés ; néanmoins, lorsqu'ils le sont, les plus couramment utilisés correspondent à des modèles de ruissellement/érosion. Le second type de modèles plus couramment utilisés, sont les modèles statistiques. On distinguera parmi ces modèles, les modèles probabilistes et les méthodes utilisant l'apprentissage statistique.

MODELE PHYSIQUES

La description des modèles physiques ne sera pas effectuée. En effet, de nombreux modèles physiques cherchant à modéliser la quantité de matières en suspension à la sortie d'un système (souterrain ou de surface) se basent sur des modèles de ruissellement et d'érosion. Malheureusement, ces modèles ne sont généralement appliqués qu'à des bassins de petite taille (généralement inférieur à 1 km² même s'ils peuvent atteindre 100 km²). A notre connaissance, aucun modèle physique ne propose de modéliser directement la relation pluie-turbidité. Ainsi nous pourrons noter le modèle SHESE (SHE hydrological modelling system) correspondant à une adaptation du modèle SHE (Système Hydrogéologique Européen) utilisé par Wicks, J. M., & Bathurst, J. C. (1996) l'appliquant sur un bassin de 0.06 km². Cette méthode se base sur l'utilisation d'équations visant à modéliser les processus tels que :

- l'arrachement de particules par l'impact des gouttes de pluies (qui utilise l'intensité de la pluie et la proportion argile/limons/sable du sol),
- l'arrachement des particules par ruissèlement (qui utilise les coefficients d'érosion des sols),
- le maintien des particules en solution (qui utilise la hauteur de la lame d'eau, la concentration en sédiments, la porosité de la surface du sol et l'élévation du sol),
- le transport de l'eau et des sédiments dans des conduits (qui utilise la section du conduit, la concentration en sédiment, la porosité des parois, la vitesse de transport des sédiments, le coefficient de dispersion longitudinal et le débit d'alimentation du conduit).

On notera également les travaux de Kothyari, U. C., & *al.* (1997) appliquant leur modèle kinetic sur des bassins allant de 0,002 km² à 92,5 km². Le modèle utilisé durant leur étude est composé de deux équations visant à transcrire :

- l'érosion brute du sol (qui utilise l'érosivité des pluies, le facteur d'érodibilité du sol, le degré d'inclinaison et la longueur des pentes et le facteur de couverture du sol),
- le ratio d'apport de sédiments (qui utilise la longueur du flux, la pente et l'utilisation du sol).

L'application de telles méthodes à des bassins de plus grande taille aura pour impact de grandement augmenter la complexité du modèle en diversifiant et multipliant le nombre et la taille des différentes structures intervenant dans la réponse hydrologique du système. Ces structures ne pouvant pas toutes être identifiées et caractérisées sur de grandes superficies, l'utilisation de ces modèles en est compromise.

MODELES PROBABILISTES

Towler, E. & *al.* (2010) ont cherché à modéliser, pour quatre périodes différentes (périodes sèches, très sèches, humides et très humides), la turbidité au niveau d'une source d'eau potable située dans le Nord-Est des États Unis. Les auteurs avaient à leur disposition des mesures mensuelles de précipitation et des mesures journalières de vitesses d'écoulement et de turbidité. La modélisation est effectuée en 3 étapes. Dans un premier temps, les auteurs ont modélisé la probabilité d'avoir des pluviométries supérieures à la normale, proches de la normale et inférieures à la normale, pour chaque période. Puis, via cette pluviométrie, ils ont effectué la prévision des vitesses d'écoulement à partir des résultats précédents. Finalement, ils ont utilisé une méthode probabiliste basée sur la régression logistique locale afin de prévoir le dépassement d'un seuil de turbidité. Après avoir appliqué cette méthode sur des seuils de turbidité allant de 1 NTU à 5 NTU, ils ont montré que leur méthode apportait un avantage dans la prédiction mensuelle d'un dépassement de seuil dans 60% des cas.

Tramblay, Y. & *al.* (2007) a appliqué différentes distributions de probabilité sur un total de 149 bassins versants (49 au Canada et 100 au États Unis) pour lesquels il disposait de 10 ans de mesures au pas de temps journalier des concentrations en matières en suspension, qu'ils a ré-échantillonnées en effectuant la moyenne mensuelle. Après avoir ajusté ces différentes distributions sur ses données, un critère bayésien d'information a permis de sélectionner la distribution s'ajustant le mieux pour chaque bassin versant. Ainsi 80% des sites se sont retrouvés le mieux adaptés à des distributions de types : lognormal, exponentiel, gamma et Weibull. Les 20% restants se sont trouvés mieux adaptés à des distributions de types : GEV, log-Pearson de type III et Pareto généralisé. À partir de ces lois statistiques, il est alors possible de déterminer des probabilités de dépassement de seuils.

MODELES UTILISANT L'APPRENTISSAGE STATISTIQUE

Beaudeau, P. & *al.* (2001) ont utilisé le couplage d'un modèle hydrologique conceptuel pour modéliser le ruissellement et d'un réseau de neurones pour prédire la turbidité. Cette étude a été appliquée au bassin versant d'Yport situé en Normandie (France) en utilisant des données de turbidité mesurées dans un forage proche de l'exutoire du bassin et captant l'eau dans un drain karstique, les données de piézométrie dans ce même forage et les données de pluviométrie recueillies par un piézomètre situé au centre du bassin d'alimentation du captage. La base de données comporte des mesures de décembre 1989 à novembre 1991 au pas de temps de 15minutes. Ils ont ensuite comparé les résultats via la prévision par réseau de neurones (couplé à un modèle de ruissellement) avec la prévision obtenue par un modèle linéaire de type ARMAX (Auto-Regressive Moving-Average with Exogenous Inputs). Ils ont ainsi obtenu des résultats pour lesquels 28% des variations de turbidité étaient mieux représentées par les valeurs simulées dans le cas du modèle linéaire et de 74% dans le cas du model par réseau de neurones.

Rajaee, T. & al. (2009) ont effectué un travail sur la concentration de matières en suspension (SSC pour *Suspend Sediment Concentration*) dans la *Little Black River* et la *Salt River* (Etats Unis). Pour cette étude, ils avaient à leur disposition des données journalières de SSC (sortie des modèles) et de débits (entrées des modèles). Les modèles choisis sont des réseaux de neurones, des NF (pour Neuro-Fuzzy), des MLR (pour Multi Linear Regression) et SRC (pour Sediment Rating Curve). Ils ont ainsi montré que le meilleur modèle était le modèle de type NF avec un coefficient de détermination (R²) de 0,7. Pour les modèles de type réseaux de

neurones, MLR et SRC, le R² était respectivement de 0,46, 0,26 et 0,23. Ils ont ainsi montré l'utilité des réseaux de neurones dans la prédiction journalière de la concentration de matières en suspension.

Kisi, O. & *al.* (2012) ont tenté de modéliser la concentration journalière en matières en suspension de deux stations (*Bardouville* et *Pineville*) situées sur la rivière de Cumberland (Etats-Unis) en utilisant les débits journaliers mesurés en ces deux mêmes points. Ils ont ainsi comparé les résultats de modèles de type ANFIS (Adaptative Neuro-Fuzzy Interference System), réseaux de neurones, SVM (Support Vector Machine) et GEP (Genetic Expression Programming). C'est le GEP qui a obtenu les meilleurs résultats avec un R² de 0,963 (Pineville) et 0,938 (Bardouville). Les modèles de type SVM, réseau de neurones et ANFIS ont quant à eux, donné des résultats avec un R² de 0,9 pour les deux stations. Avec ces travaux, Kisi, O., & al. (2012) ont montré l'efficacité des modèles GEP mais également la capacité des trois autres modèles (SVM, réseau de neurones et ANFIS) pour la modélisation des concentrations de matières en suspension journalières.

Concernant la prévision, Mulia, I. E. & *al.* (2013) ont appliqué un réseau de neurones (de type perceptron multicouche) sur deux stations situées dans le détroit de Singapour afin de prévoir la turbidité et la concentration en chlorophylle-a avec un horizon de prévision de 14h. Ils ont utilisé les données horaires de turbidité (sortie du modèle), de concentration en chlorophylle-a (entrée du modèle) et de hauteur d'eau (entrée du modèle) mesurée pendant l'année 2010 pour le premier point et les données allant du 1^{er} novembre 2009 au 1^{er} janvier 2010 pour le second point. Afin de sélectionner la meilleure architecture pour le réseau de neurones, ils ont utilisé un algorithme génétique pour déterminer l'architecture et le meilleur jeu de paramètres. Pour la comparaison des résultats obtenus, ils ont utilisé la racine de l'erreur quadratique moyenne (ou RMSE pour Root Mean Square Error), ce qui rend les comparaisons inter bassins difficiles. Le modèle le plus performant pour la prévision de la turbidité a donné des résultats de RMSE de l'ordre de 1,5 sur l'ensemble des tests, et des résultats de RMSE globalement supérieurs à 2 pour la prédiction des concentrations en chlorophylle-a (RMSE allant de 1,5 à 5).

Plus récemment Houria, B. & *al.* (2014) ont utilisé en conjonction 2 méthodes de l'apprentissage statistique (réseaux de neurones et optimisation par essaims particulaires) pour prévoir la concentration en matières en suspension du bassin *Ksob* en Algérie. Ils partent du concept de « courbe de tarage » entre le débit de la rivière et la concentration en matières en suspension. Le modèle est ainsi alimenté par le seul débit du cours d'eau et délivre une prévision de la concentration de matières en suspension ayant un coefficient de corrélation linéaire de 0.57 à 0.85. La base de données et son découpage en apprentissage et en test ne sont pas très bien définis ; il est donc difficile d'évaluer la qualité des performances annoncées.

Iglesias, C. & *al.* (2014) ont appliqué des réseaux de neurones pour prédire la turbidité dans la rivière *Nalón*. Pour cette prédiction, ils ont utilisé les données de turbidité (sortie du modèle), de concentration en ammonium (entrée du modèle), de conductivité (entrée du modèle), d'oxygène dissous (entrée du modèle), de pH (entrée du modèle) et de température (entrée du modèle) au pas de temps horaire. En utilisant ces données, ils ont créé des variables issues de la combinaison de plusieurs données entre elles. Grâce à un perceptron multicouche (PMC), ils ont obtenu des prédictions avec un R² de o,8 pour l'utilisation de la variable combinée « oxygène dissous*Température » contre un R² de o,7 dans le cas de l'utilisation des variables simples. Malheureusement, ils ne font pas mention de leur horizon de prédiction, ce qui ne nous permet pas de savoir si leur modèle effectue une prédiction ou seulement une simulation.

Sur le même site d'étude (rivière *Nalón*), Garcia Nieto, P.J. & al. (2014) ont mis en place des modèles hybrides PSO (particle swarm optimization) et SVM. Les données utilisées pour cette étude sont les données de turbidité (sortie du modèle), de conductivité (entrée du modèle), de concentration en ammonium (entrée du modèle), d'oxygène dissout (entrée du modèle) et de pH (entrée du modèle). Ils ont divisé leur base de données en trois périodes en se basant sur la hauteur d'eau. Ils ont ainsi effectué des prédictions pour des périodes de hautes, moyennes et basses eaux. Les modèles ont alors permis l'obtention de prédiction avec des résultats de R² de 0,9 pour la période de hautes et moyennes eaux et de 0,87 pour la période de basses eaux.

Il apparait que la modélisation de la turbidité a été étudiée au travers de différents modèles. Dans une première catégorie, les modèles physiques utilisés se basent sur la modélisation du ruissellement et de l'érosion. De par le nombre de paramètres intervenant dans leur processus, ces derniers sont généralement appliqués à des bassins homogènes, donc en général de très petite taille. Leur application à des bassins hydrogéologiques de plus grande taille (plus de 200 km²) a pour effet une forte dégradation des résultats des modèles due à l'augmentation de la complexité du système (Berthet, L., & al. 2008, November). En se référant au Tableau 1 effectuant la synthèse des modèles statistiques, on observe tout d'abord que la première approche que l'on peut noter est l'utilisation de la relation entre la turbidité et le débit (Twoler, E. & al. 2010 ; Houria, B. & al. 2014 ; Kisi, O., & al. 2012 et Rajae, T. & al. 2009). Ensuite, on note l'exploration de la prédiction en se basant sur les propriétés chimiques de l'eau effectuée par Garcia Nieto, P.J. & al. (2014) et Iglesias, C. & al. (2014) sur le site de la rivière Nalón. Pour finir, Mulia, I.E. & al. (2013) se sont basés sur l'état du système. D'un point de vue hydrologique, ces travaux ont utilisé des données d'entrée correspondant à des variables ayant subi l'influence du système souterrain et de surface. D'autres travaux quant à eux ont tenté de modéliser la turbidité ou la concentration en matière en suspension en utilisant des variables n'ayant pas subi les modifications dues aux systèmes souterrains et de surface. Ainsi, on notera les travaux de Beaudeau, P. & al. (2001) et Tramblay, Y. & al. (2007) qui ont utilisé les pluviométries comme variable d'entrée de leurs modèles afin de prédire la turbidité à la sortie de leur système. Il apparait alors peu de modèles effectuant la prédiction de la turbidité au pas de temps infra-journalier en utilisant les données de pluviométries. Il ressort également que les modèles utilisant l'apprentissage statistique (de type réseaux de neurones) sont généralement plus performants en termes de qualité de prédiction/modélisation. On peut également noter, pour finir, le faible nombre d'études s'intéressant aux systèmes souterrains.

Auteurs	Towler, E. & <i>al.</i> (2010)	Tramblay, Y. & <i>al.</i> (2007)	Beaudeau, P. & <i>al</i> . (2001)	Rajaee, T. & <i>al</i> . (2009)	Kisi, O. & al. (2012)	Mulia, I. E. & <i>al.</i> (2013)	Houria , B. & <i>al.</i> (2014)	Iglesias, C. & <i>al.</i> (2014)	Garcia Nieto, P.J. & al. (2014)
Modèle(s)	MP	Distribution de probabilité	NN + Runoff model	NN, NF, MLR, SRC	ANFIS, SVM, GEP	NN	NN, PSO	NN	PSO + SVM
Pas de temps	М	М	Н	J	J	Н	Н	Н	?
Application	Eau de surface	Eau de surface	Eau souterraine < 50 NTU	Eau de surface	Eau de surface	Eau de surface	Eau de surface	Eau de surface	Eau de surface
Données d'entrée	Pluie	Pluie	Pluie	Pluie, débit	Débit	[Chlorop hylle-a], hauteur d'eau	Débit	[NH ⁺ ₄], Conducti vité, Oxygène dissous, pH, Tempéra ture	[NH ⁺], Conducti vité, Oxygène dissous, pH, Tempéra ture

Tableau 1 : Tableaux récapitulatif des travaux effectués sur la prévision de la turbidité par des modèles statistiques. Pour les modèles, les abréviations; MP, NN, NF, MLR, SRC, SVM, GEP, PSO correspondent respectivement a; Modèle probabiliste, Neural Network, Neurone Flous, Multilinear Regression, Sediment Rating Curve, Support Vector Machine, Genetic Expression Programming, Particul Swarm Optimisation. Pour les pas de temps, les abréviations M, J et H représentent respectivement des pas de temps Mensuel, Journalier et Horaire.

1.3.2 LES PRODUITS PHYTOSANITAIRES

Les produits phytosanitaires sont des substances ayant pour objectif d'agir sur le développement des organismes vivants jugés comme nuisibles. Suivant leurs actions, on peut distinguer 4 types de produits phytosanitaires :

- Les herbicides
- Les fongicides
- Les insecticides
- Les autres (acaricides, molluscicides, taupicides,...)

Les produits phytosanitaires peuvent être utilisés pour des traitements préventifs (pour empêcher le développement de nuisibles) ou en traitement curatif (pour se débarrasser des nuisibles lorsque ces derniers sont apparus).

DEGRADATION ET TRANSPORT LIBRE

Les produits phytosanitaires, une fois dans le milieu naturel et outre leur dégradation propre, peuvent subir 3 modes de diffusion :

- l'absorption par les plantes,
- la diffusion gazeuse,
- le transfert par l'eau.

L'absorption par les plantes permet dans un premier temps d'empêcher les produits phytosanitaires de subir un transport dans le sol ou à sa surface. En prenant l'exemple du maïs, Tasli, S. (1995) a montré que seulement 10% de la dose épandue est absorbée par les plantes. Raveton, M. & *al.* (1997) ont cependant montré que les plantes jouaient, dans le cas de l'atrazine, un rôle de concentration. On notera que le taux d'absorption des produits phytosanitaires par les plantes dépend de nombreux paramètres tels que la disponibilité des composés, la solubilité dans l'eau, les constituants lipidiques racinaires (Kow) et l'activité biologique de la plante au travers de paramètres tels que la température et l'humidité (Schroll, R. & *al.* 1992).

Comme l'a montré Abbott, D. C. & *al.* (1965), Lode, O. & *al.* (1995) et Millet, M. & Bedos, C. (2017), la diffusion gazeuse joue un rôle important dans la mise en évidence de produits phytosanitaires dans l'eau de pluie. Même si ce phénomène fut beaucoup étudié (Glotfelty, D. E. & *al.* 1984; Cooper, R. J. *al.* 1990; Jaunky, A. 2000; Foy, C.L. 1964; Kearney, P. C. & *al.* 1964), la diffusion gazeuse reste une voie de transport des produits phytosanitaires mal identifiée. En effet, chaque produit possède des taux de volatilisation différents et les propriétés de l'environnement rendent la modélisation complexe.

Pour le transfert par l'eau, il est possible de différencier deux mécanismes majeurs ; le lessivage/lixiviation et le ruissellement. Le lessivage/lixiviation se traduit par l'infiltration de l'eau dans le sol entrainant la migration des produits phytosanitaires dans le sol. Cette infiltration est régie par des paramètres tels que la perméabilité du sol ou la vitesse d'écoulement de l'eau. On aura alors une absorption et une dégradation des produits phytosanitaires dans le sol (Miller, J.L. & *al.* 1997). Ainsi, Aderhold, D. & Nordmeyer, H., (1995) ont montré que le sol jouait un rôle important de filtre et évitait la contamination de l'eau souterraine. Le ruissellement, quant à lui, est un mécanisme qui intervient dans un second temps, lorsque la capacité d'infiltration du sol est arrivée à saturation. La capacité d'infiltration du sol peut fortement varier au cours du temps. On aura ainsi une infiltration maximale après que le sol ait été labouré et une infiltration minimale lorsque les composés du sol se seront réarrangés après une pluie en formant une croûte imperméable à la surface du sol appelée "croûte de battance" (Boiffin, J. & *al.* 1986). Le transport des produits phytosanitaires peut dans ce cas se trouver sous deux formes. On peut avoir un transport dissout des produits phytosanitaires solubles ou un transport physique des produits phytosanitaires solubles ou un transport physique des produits phytosanitaires liés aux particules en suspension. Ces deux modes de transport seront alors caractérisés par les propriétés physico-chimiques des composés (Klöppel, H. & *al.* 1994).

Pour finir, nous n'avons pour le moment parlé que des modes de dispersion. La dégradation des produits phytosanitaires, quant à elle, peut se faire suivant deux modes ;

- la dégradation abiotique,
- la dégradation biologique.

La dégradation abiotique est le mécanisme le moins représenté dans les mécanismes de dégradation. Il correspond à la dégradation chimique et/ou photochimique des produits phytosanitaires (Beestman, G. B. & Deming, J. M. 1974; Fusi, P. 1991; Ristori, G.G. & Fusi, P. 1995; Koriko, M. & *al.* 2014; El Boukili, A. 2016). Les mécanismes de dégradation abiotique peuvent varier suivant les composés et avoir un effet inverse entre deux composés. On notera l'exemple de l'atrazine dont la dégradation par hydrolyse augmente avec la teneur en acide humique et fulvique (Chesters, G., 1989) contrairement à la furanone pour laquelle, l'effet est inversé (Miettinen, I., & al. 1993).

La dégradation biologique peut être effectuée par certains micro-organismes utilisant les produits phytosanitaires comme source d'énergie et de carbone (Audus, L. J., 1949). L'avantage de cette dégradation est sa perpétuelle adaptation. Les micro-organismes, de par leur forte reproduction et mutation, vont permettre l'apparition de souches capables d'utiliser les produits phytosanitaires dans leur métabolisme (Rousseaux, S., 2001). Cette voie de dégradation est donc en perpétuelle évolution.

Au travers les différents modes de transport et de dégradation des produits phytosanitaires, on notera la diversité pouvant être observée suivant les propriétés hydro-climatiques, physico/chimique du sol et bactériologique de la zone d'épandage des produits phytosanitaires. Ces derniers auront alors un impact plus ou moins important sur l'eau souterraine et une étude approfondie de la zone devra être effectuée avant de pouvoir estimer les risques de contamination de l'eau souterraine par les produits phytosanitaires.

LIEN ENTRE TURBIDITE ET PRODUITS PHYTOSANITAIRES

Dans la partie précédente, nous avons proposé une analyse des modes de transport et de dégradation des produits phytosanitaires. Nous allons nous intéresser dans cette partie à la capacité de fixation des produits phytosanitaires par le sol. En effet, les particules en suspension responsables de la turbidité de l'eau sont majoritairement issues de l'arrachement de particules du sol lors du ruissèlement de l'eau. Notre objectif est alors de voir si la turbidité peut servir d'indicateur/PROXY de la présence de produits phytosanitaires dans les eaux naturelles. Suivant la capacité de fixation du sol, les matières en suspension pourront constituer un vecteur plus ou moins important de transport des produits phytosanitaires. Ainsi, Comme décrit sommairement dans la partie précédente, le transport des produits phytosanitaires est régi par leur taux d'absorption par le sol. L'absorption des produits phytosanitaires par le sol est régi par deux paramètres : la solubilité et le coefficient de partage eau/sol (K_d). Cependant, parmi ces deux paramètres, seul le coefficient de partage eau/sol est un bon indicateur de la mobilité de produits phytosanitaires (Gril, J. J. & al. 1999). En montrant que la concentration des produits dans l'eau de surface est inférieure à leur solubilité, il apparait que des molécules, même très peu solubles, de par leur coefficient de partage eau/sol peuvent se retrouver présentes. Le coefficient de partage eau/sol apparait donc comme l'indice majeur pour l'étude de la fixation des produits phytosanitaires sur le sol et les particules en suspension dans l'eau. Ce coefficient est propre à chacune des molécules et sera fonction de leur nature chimique, des caractéristiques de la phase solide et de la teneur en matières organiques (Barriuso, E. & al. 1996). On notera également l'influence, comme dans toute réaction chimique, des paramètres cinétiques (El Azzi, D. 2012) tels que:

- le temps de contact,
- la concentration initiale du polluant,
- la nature du substrat,
- la force ionique,
- la température,
- le pH.

En se concentrant sur l'absorption des produits phytosanitaires par les matières en suspension, Voltz, M., & Louchart, X. (2001) ont proposé des courbes théoriques de transport des produits phytosanitaires en fonction de quatre coefficients de partage eau/sol différents (Figure 1).

Si l'on porte maintenant notre attention sur les produits de dégradation, il apparait que certains de ces résidus, de par leur propriétés physico-chimiques, sont non extractibles des matières en suspension et du sol. C'est en 1970 que les premières observations ont montré qu'une partie de la radioactivité des pesticides marqués avec du ¹⁴C ne pouvait plus être extraite du milieu naturel (Grébil, G., & *al.* 2001). Divers travaux ont ensuite visé à mieux identifier ce phénomène avec la dégradation des produits phytosanitaires (Khan, S. U., & Hamilton, H. A. 1980 ; Capriel, P., & *al.* 1985 ; Schiavon, M., & *al.* 1990 ; Winkelmann, D. A., & Klaine, S. J. 1991).

Avec les informations précédentes en notre connaissance, il apparait que la fixation des produits phytosanitaires et de leurs produits de dégradation est régie par de nombreux paramètres. Ces paramètres pouvant varier au cours du temps et de l'espace, l'absorption de ces composés est un mécanisme complexe. On peut imaginer que, dans le cas d'un mélange de l'eau de surface avec l'eau souterraine, les changements de conditions tels que le pH, la température, la variation de concentration en matière en suspension, peuvent jouer un rôle dans la rétention des produits phytosanitaires par les matières en suspension.



Figure 1: Courbes théoriques de variation du pourcentage de substances transportées sur les MES pour différentes valeurs de K_d (Voltz, M., & Louchart, X. 2001).

Ainsi, de par ces fortes interactions existantes entre matières en suspension et produits phytosanitaires, l'utilisation de la turbidité (résultat de la présence de matières en suspension dans l'eau) comme PROXY des produits phytosanitaires apparait comme prometteuse. Malheureusement, à notre connaissance, il n'existe pas d'études effectuant l'analyse de la relation entre produits phytosanitaires et turbidité/matières en suspension.

1.3.3 LE TRANSPORT DES PARTICULES DANS LE KARST

Ayant évoqué la possibilité de transport des produits phytosanitaires et de leurs métabolites par les matières en suspension, il apparait important de s'interroger sur le transport des matières en suspension de la surface à l'exutoire des systèmes karstiques. Pour débuter, le principal facteur d'écoulement de l'eau à la surface et en souterrain est la gravité. En surface, l'eau va ruisseler le long des pentes jusqu'à atteindre une structure lui permettant d'atteindre une altitude de plus en plus faible. Lors de cet écoulement de surface, suivant sa vitesse et la stabilité du sol, l'eau va se charger en particules en suspension. Cette eau chargée va alors atteindre plus ou moins rapidement deux structures principales. Dans le premier cas, l'eau va atteindre des chenaux d'écoulement préférentiels qui, à force de se combiner, vont donner lieu à des cours d'eau/rivières/fleuves. Dans le deuxième cas, les écoulements d'eau vont atteindre des zones d'infiltrations préférentielles telles que les bétoires/dolines (en ce qui concerne notre site d'étude). C'est ce deuxième cas qui nous intéresse car il sera responsable de l'infiltration de l'eau dans le système souterrain. Durant son voyage dans le milieu souterrain, l'eau va traverser différents niveaux structuraux. En ce qui concerne le karst plusieurs de ces niveaux ont pu être définis conceptuellement. Nous les décrivons ci-après.

L'EPIKARST

L'épikarst est la partie faisant la transition entre la surface et le karst. Bien que proche du sol, cette zone reste mal définie : on notera les travaux de A. klimchouck (2004) qui a référencé douze définitions différentes de l'épikarst dans la littérature. Cette zone d'infiltration a la capacité de stocker l'eau et ainsi de former une réserve d'eau disponible pour l'évapotranspiration, mais également permettant un retard des potentielles contaminations d'eau provenant de la surface. Au niveau de cette zone, l'eau peut alors décanter et un dépôt de matières en suspension peut se former. Ce dépôt peut ensuite être remis en suspension lors d'un évènement de crue résultant de fortes précipitations et d'une infiltration rapide.

Par la suite, l'eau plus ou moins chargée en matières en suspension atteindra, par des écoulements verticaux, la zone non-saturée et la zone saturée.

LA ZONE NON SATUREE

Concernant les karsts d'origine gravitaire, cette zone correspond à la zone principale de présence des conduits karstiques. On notera cependant, l'existence d'exceptions telles que celle des karsts méditerranéens dont la formation a été affectée par la crise messénienne, et pour lesquels un grand nombre de conduits karstiques peuvent être observés dans la zone saturée. Pour ces karsts, la fermeture du détroit de Gibraltar et ainsi l'asséchement de la mer méditerranée ont induit un creusement des conduits karstiques en dessous du niveau actuel de la mer (Audra, P., & *al.* 2004). L'écoulement dans la zone non saturée peut être de deux types (Bakalowicz, M. 2008, May) :

- écoulement diphasique (écoulement de l'eau en contact avec l'air dans les microfissures),
- écoulement monophasique (écoulement de l'eau dans les conduits saturés).

L'écoulement diphasique étant responsable de la dissolution de la roche, cet écoulement peut être à l'origine de l'apport de matières en suspension dans l'eau des écoulements monophasiques.

Les écoulements monophasiques sont les principaux écoulements intervenant dans le transport rapide (crues) de l'eau et des matières en suspension.

On notera que la structure des conduits peut amener la présence de zones de stockages de l'eau responsables de la décantation et du dépôt des particules en suspensions. Les écoulements rapides (résultants d'une crue) pourront, par la suite, remettre en suspension les matières préalablement déposées lors de périodes d'écoulements plus lents (telles que les périodes d'étiages).

LA ZONE SATUREE

La zone saturée est caractérisées par une zone noyée dont la limite supérieure est délimitée par la surface piézométrique de la nappe phréatique. Cette limite supérieure peut donc varier en fonction de l'état hydrique de la nappe. On peut ainsi retrouver des karsts dans cette zone de battement que l'on peut considérer appartenant à la zone saturée et à la zone non saturée. Dans la zone saturée, les écoulements dans les conduits noyés sont régis par les gradients hydrauliques et non par les pentes des galeries, comme on pouvait l'observer pour les écoulements de surfaces et dans la zone non saturée. Il arrive donc que les directions d'écoulements dans la zone non saturée et saturée soit différentes. Même si les avis sont partagés sur les processus des stockages, cette zone est considérée comme pouvant être la zone principale de stockage de l'eau. En effet, selon Mangin. A. (1975), ce sont des cavités qui forment un système annexe au drain principal qui sont responsables du stockage. De son côté, Drogue. C. (1974) avance l'hypothèse selon laquelle le stockage s'effectue par la porosité matricielle. Il est vraisemblable que chacune des hypothèses soient vraie selon le type de karst étudié et le régime météorologique.

CONCLUSION SUR LE TRANSPORT DE PARTICULES DANS LE KARST

Mangin. A. (1998) a proposé un schéma récapitulatif des écoulements dans le karst présenté en Figure 2. Au travers de ce schéma, représentant plusieurs chemins possibles entre les pluies et l'exutoire, et des informations précédemment recueillies, il est clair que l'origine des matières en suspension à l'exutoire d'un système karstique peut être très variable. Les travaux de Lacroix, M., & *al.* (2000) ou encore de Massei N. (2001) ont d'ailleurs mis en évidence une variation des origines des matières en suspension suivant si l'on est en période d'étiage ou de crues.



Pluie– contribution au système

Figure 2 : Schéma fonctionnel du karst d'après Mangin. A. (1998)

Dans la littérature, Il est admis que la turbidité est un traceur naturel des transports de particules en suspension lorsque la turbidité est utilisée comme traceur des éléments dissous (Fournier, M., & *al.* 2007). Cependant, la possible utilisation de la turbidité en tant que PROXY des produits phytosanitaires, n'est pas à exclure. En effet, comme nous avons vu dans le paragraphe 1.3.2, la variation des propriétés physico-chimiques de l'eau et de la concentration en particules en suspension peut impacter le taux d'absorption des produits phytosanitaires. Il apparait donc intéressant d'approfondir l'étude de la turbidité en tant que PROXY des produits phytosanitaires.

1.4 CONCLUSION DU CHAPITRE 1

Au travers de ce chapitre nous avons pu voir que la qualité de l'eau est un sujet devenant de plus en plus crucial pour la protection des populations et de la biodiversité. De nombreuses réglementations à l'échelle européenne et nationale ont vu le jour au fils des années dans le but, dans un premier temps, d'identifier puis de limiter l'impact de l'activité humaine sur la ressource en eau. Par la suite, en s'intéressant à la turbidité, représentant l'un des critères de qualité de la ressource en eau, nous avons vu que la turbidité pouvait être observée tout simplement de manière visuelle. Cependant, ce dernier mode d'observation, par son manque de précision, a laissé la place à de nombreuses méthodes de mesure plus complexes. Tout en décrivant ces méthodes, les difficultés de mesure provenant des propriétés physiques et chimiques des particules en suspension responsables de la turbidité, ont été mises en avant. Parmi ces méthodes, la transmissométrie est ressortie comme la méthode la plus facile à mettre en place, avec une incertitude acceptable liée aux variations des propriétés des matières en suspension.

Par la suite, un état des travaux de modélisation de la turbidité a été effectué pour permettre un positionnement scientifique de nos travaux. Ainsi, nous avons montré que la diversité des modèles utilisés pour prédire la turbidité est relativement importante. Cependant, la plupart des modèles s'attachent à prédire la turbidité de l'eau de surface dont les mécanismes sont très divers en comparaison des mécanismes intervenant sur les flux de l'eau souterraine. Les rares modèles s'intéressant aux systèmes hydrologiques souterrains n'effectuent généralement qu'une prédiction avec un pas de temps à minima journalier. De plus, il est apparu que les modèles à réseaux de neurones apparaissent comme les plus appropriés pour modéliser la relation pluie-turbidité, en grande partie grâce à leur capacité à modéliser des systèmes fortement non linéaire.

Pour finir, nous avons vu que les produits phytosanitaires pouvaient, de par les mécanismes responsables de leur transport dans les eaux, être associés à la turbidité. Ainsi, à la fin de cette partie, nous avons proposé d'étudier plus en profondeur l'utilisation de la turbidité comme PROXY des produits phytosanitaires, et ainsi augmenter l'intérêt de la mesure de la turbidité pour le suivi de la qualité de l'eau souterraine.
Chapitre 1 : Importance de la mesure de la turbidité pour la qualité de l'eau souterraine

Chapitre 2 : Présentation du Site d'Yport et des données associées

CHAPITRE 2 : PRESENTATION DU SITE D'YPORT ET DES DONNEES ASSOCIEES

Chapitre 2 : Présentation du Site d'Yport et des données associées

2.1 INTRODUCTION

Ce chapitre a pour vocation de présenter le site d'étude sélectionné pour nos travaux. Dans un premier temps, une description globale du climat normand est effectuée. S'en suit la description géologique et hydro(géo)logique de cette région.

Par la suite, la troisième partie centrera notre étude sur le site d'Yport où se situe le captage d'alimentation en eau potable constituant la cible de notre travail. Nous verrons ainsi que les périmètres de protection ont subi de nombreuses modifications au cours des années, au fur et à mesure que les connaissances géologiques et hydrogéologiques amenaient des précisions dans la description du bassin d'alimentation d'Yport. Nous verrons également la cause de la forte vulnérabilité de l'eau souterraine de notre zone, autant d'un point de vue naturel qu'anthropique. Pour finir, les données à notre disposition seront décrites et critiquées afin d'évaluer leur intérêt pour nos travaux et leur possible impact dans le cadre de leur utilisation dans des modèles prévisionnels.

Pour finir, nous proposerons des alternatives aux pratiques actuelles, afin d'augmenter la précision et la qualité des mesures des produits phytosanitaires effectuées actuellement.

2.2 CONTEXTE REGIONAL

2.2.1 CLIMAT DE LA SEINE MARITIME

Le site d'Yport est localisé en Normandie, à l'Ouest de la Seine Maritime dans le Pays de Caux (Bec de Caux). Il est principalement sous l'influence d'un climat océanique avec une température mensuelle moyenne comprise entre 6,6° et 14,4°C (1981-2010), et des précipitations mensuelles moyennes variant entre 600 et 1100 mm/an (source : Météo France). Durant l'année, on note deux périodes avec de fortes précipitations (Figure 3). La première, et la plus importante, s'étend de début octobre à fin janvier. Cette période regroupe à elle seule 50% des précipitations annuelles (Delay, F. & *al.* 1997). La deuxième période s'étend de mai à juin. Les précipitations durant cette période se manifestent par des orages. Durant la première période, les précipitations peuvent varier du simple au triple en fonction des années et du simple au sextuple durant la seconde période (Hauchard, E. 2002). Le cycle annuel se trouve alors plus ou moins marqué en fonction des années.



Figure 3 : Précipitations et températures en Normandie relevées durant l'année 2015. Source : Météo France.

Sur la Figure 4, on peut noter la présence d'un gradient de précipitation provenant de la localisation côtière de la Haute Normandie. On observe globalement un gradient décroissant sur un axe nord-ouest/sud-ouest avec des précipitations annuelles variant entre 1120mm et 560mm. En se concentrant sur notre zone d'étude

(rectangle de contour vert), on peut noter des précipitations annuelles de l'ordre de 800 mm puis une augmentation de la valeur allant jusqu'à 1120 mm lorsque l'on s'éloigne du front de mer.



Figure 4 : Carte de répartition des cumuls de précipitations quotidiennes (normales calculées sur la période de 1971 à 2000). Source : Météo France.

2.2.2 GEOLOGIE

La principale formation géologique retrouvée en Seine Maritime correspond à un plateau de craie, karstifié, formant la partie Ouest du Bassin Parisien. Cette formation présente un pendage d'environ 10° vers l'Est. Audessus de ce plateau, différentes formations superficielles peuvent être identifiées telles que des argiles à silex, des lœss, des dépôts tertiaires, des colluvions et des alluvions. Une coupe schématique de ces formations a été proposée par Laignel, B. (2003) mais une révision du schéma a récemment été réalisée par Chédeville, S. & *al.* (2014a) afin de représenter les réseaux karstiques de manière plus juste (Figure 5).

Chapitre 2 : Présentation du Site d'Yport et des données associées



Figure 5 : Coupe schématique des formations superficielles de la Seine Maritime (Chédeville, S. & al. 2014a).

LA CRAIE

Pour la craie, Il est possible d'observer des affleurements de terrains datant majoritairement du Crétacé inférieur. On retrouve ensuite des terrains datant du Jurassique au nord-est le long du pays de Bray principalement constitué du Portlandien. On trouve ainsi 5 étages stratigraphiques (Laignel, B. 1997) de craies plus ou moins riches en silex qui sont les suivants (en partant du haut vers le bas) : le Sénonien formé par le Campanien (de 30 à 40 m d'épaisseur), le Santonien (de 30 à plus de 60 m d'épaisseur), le Coniacien (de 20 à plus de 60 m d'épaisseur), puis le Turonien (de 70 à 90 m d'épaisseur) et le Cénomanien moyen et supérieur (de 30 à 35m d'épaisseur).

Pour finir, à la base des formations crayeuses du Cénomanien, on retrouve le Cénomanien dans une couche composée de sable calcaire glauconieux et l'Albien qui se compose d'argiles grises.

LES FORMATIONS SUPERFICIELLES

Les formations superficielles sont composées de différentes couches :

- Les dépôts sableux ou sablo-argileux Tertiaires, situés au sommet des argiles à silex souvent localisées dans des fossés d'effondrement ou des poches karstiques (cônes de dissolution, racines d'altération) mais rarement en couches continues (Laignel, B. 2003; Hauchard, E. & Laignel, B. 2008)
- Des argiles à silex Plio-quaternaires issues de la dissolution de la craie
- Des lœss quaternaires correspondant à des sédiments détritiques d'origine éolienne mis en place en périphérie des zones glacières de l'aire du Quaternaire et qui constituent une barrière fortement imperméable. Ces formations, au contact de la craie, forment une structure où l'on peut cependant observer des zones de pénétration des couches superficielles dans le manteau crayeux.
- Les alluvions que l'on retrouve en fond de vallées.

LE CONTEXTE STRUCTURAL DE LA SEINE MARITIME

D'un point de vue structural, le Pays de Caux se trouve affecté par de nombreuses déformations et accidents. On trouve de grandes failles de direction armoricaine N150-160 et N120-130 (Figure 6 : failles F1, F5, F6 et F7) et des failles secondaires de direction varisque N30°-50° (Figure 6 : failles F3 et F10).

Parmi ces failles, on note trois grandes failles qui sont les failles : (*i*) de Fécamp-Lillebonne (Figure 6 : faille F1), (*ii*) de la Seine (Figure 6 : Faille F5) et (*iii*) de Bray (Figure 6 : Faille F7) (Hauchard, E. & Laignel, B. 2008), qui vont grandement affecter les écoulements souterrains. On note pour finir une direction majeure de failles situées au nord du Pays de Caux comportant une orientation N90°-100° (Figure 6 : Failles F2 et F4) (Hauchard, E. 2001, Hauchard, E. & *al.* 2002; Hauchard, E. & Laignel, B. 2008).

Chapitre 2 : Présentation du Site d'Yport et des données associées



Figure 6 : Schéma structural du Nord-Ouest du Bassin de Paris. (Hauchard, E. & Laignel, B. 2008).

2.2.3 HYDROGEOLOGIE REGIONALE

Les circulations d'eau en surface et en profondeur vont subir l'influence des failles recensées précédemment, ainsi que l'influence des différentes couches superficielles telles que les alluvions ou les sables verts de l'Albien ; mais c'est la craie, de par sa présence généralisée, qui va jouer le rôle le plus important et ainsi former le principal aquifère.

LA NAPPE DES ALLUVIONS DE LA SEINE

La nappe alluvionnaire, principalement présente dans les fonds de vallées, est composée de deux grands ensembles d'épaisseur variable représentés par : (*i*) les alluvions fines en surface formant une couche de faible perméabilité et (*ii*) les alluvions anciennes en profondeur avec une plus forte perméabilité.

La nappe des alluvions est considérée comme captive du fait de la présence des alluvions fines. C'est la nappe des alluvions anciennes qui constitue la plus grande partie de la nappe des alluvions, et qui se trouve au contact de la nappe de la craie sous-jacente.

De par la présence de la couche faiblement perméable des alluvions fines et de sa localisation au-dessus de la nappe de la craie, la nappe des alluvions se trouve généralement alimentée directement par l'eau en provenance de la nappe de la craie par débordement, ce qui amène à s'interroger sur sa différentiation en tant qu'unité hydrologique (Hanin, G. 2010).

LA NAPPE DE LA CRAIE

La nappe de la craie se trouve dans un aquifère karstifié, cette nappe est libre et délimitée en profondeur par le toit de la nappe de l'Albien. On note le recouvrement ponctuel de cette nappe par la nappe des alluvions présente dans le fond des vallées. Malgré cela, la nappe de la craie est tout de même considérée comme libre. On observe deux écoulements principaux ; un vers le nord pour les bassins versants littoraux et un vers le sud pour les bassins versants alimentant la Seine.

Les écoulements dans cette nappe sont gouvernés par les propriétés hydrauliques de trois types de structures: (*i*) la matrice (via sa porosité et sa conductivité hydraulique), (*ii*) les fissures et fractures (via leur densité, étendue horizontale, profondeur), et (*iii*) les drains karstiques.

D'après Juignet, P. (1980) et Maqsoud, A. (1996), la porosité de la craie présente des valeurs de l'ordre de 30 à 40% et Laignel, B. (2003) a montré que cette dernière pouvait varier de 15,4% à 47% suivant les étages

considérés. Cependant, la connexion des pores étant très faible, Calba, F. (1980) et Mégnien, C. (1979) ont montré que la porosité contribuant à la perméabilité n'est que d'environ 0,5%. Par la suite, Bracq, P. & *al.* (1992) ont montré que la conductivité hydraulique de la craie était de l'ordre de 10⁻⁸ m/s. La capacité transmissive de la craie est renforcée par ses fractures de sorte que la fracturation confère à l'aquifère une conductivité hydraulique de l'ordre de 10⁻⁴ à 10⁻⁶ m/s ; plus localement, la conductivité hydraulique peut atteindre 0,1 m/s à la faveur des drains karstiques (traçage IDDEA de la perte de Mirville, 2012). En effet, bien que non comparable, en terme de grandeur, à des karsts tels que les karsts des Grand Causses du sud de la France, le développement du karst de la craie n'en n'est pas moins important dans certaines zones (Rodet, J. 1993). D'après une étude menée par Rodet J. (1993), si l'on ne tient pas compte des racines d'altération composant l'épikarst, 89,1% de la karstification est localisée dans le littoral cauchois et le Val de Seine, 10% dans les vallées des affluents de la Seine et 0,9% au niveau des plateaux.

Les aquifères karstiques, de par leur rapidité de transfert, correspondent à la principale cible pour les besoins d'alimentation en eau potable. Les formations superficielles imperméables, combinées aux conduits karstiques, vont entrainer un ruissellement important vers des zones d'infiltration préférentielles, dolines de dissolution nommées les bétoires, et connectées directement avec le réseau karstique de drainage. L'eau s'infiltrant dans ces bétoires va contribuer en partie au rechargement de la nappe par ces fissures, et s'évacuer par les sources qui drainent le massif. Ce transfert rapide influence fortement la qualité de l'eau souterraine (Massei, N. 2001). Les bétoires et le karst vont donc jouer un rôle important dans le transport de matière en suspension, potentiellement associée à des polluants (produits phytosanitaires, contaminations fécales,...) transférés ainsi vers des zones de captage de l'eau souterraine.

LA NAPPE DES SABLES VERTS DE L'ALBIEN

A l'exception de sa partie comprise dans l'estuaire de la Seine où les sables verts de l'Albien affleurent en surface ; la nappe des sables verts de l'Albien est considérée comme une nappe captive généralement localisée sous l'aquifère de la craie. On note la séparation entre la nappe de la craie et la nappe des sables verts de l'Albien par la présence d'une couche imperméable allant de 10 à 50 m composée d'argiles de Gault. Cependant, cet isolement des deux nappes ne semble pas total car l'analyse des pressions relevées dans ces deux nappes laisse supposer de possibles échanges (Valdes, D., 2005).

2.3 TERRITOIRE D'ETUDE : LE BASSIN D'ALIMENTATION DU CAPTAGE D'YPORT

2.3.1 DELIMITATION DU BASSIN D'ALIMENTATION

Le captage d'alimentation en eau potable d'Yport, exploité par la Communauté d'Agglomération Havraise (CODAH) capte son eau dans un drain karstique. Le drain capté est également responsable de l'alimentation des sources du littoral composant les Fontaines d'Yport. Ces fontaines ont été exploitées depuis de nombreuses années pour l'alimentation en eau potable de la ville d'Yport. C'est en 1896 que furent entrepris les premiers travaux d'aménagement d'un captage. Des travaux plus importants furent entrepris à la suite d'une étude des sources qui a montré des débits estimés à 1,5 m³/s (Hole. J.P. & Roux. J.C. 1978) puis ré-estimé récemment à 2,8 m³/s (Gaillard. T. & al. 2012). Cependant, la localisation des sources le long des falaises d'Yport dans la partie tidale de la côte entraine une forte vulnérabilité des installations de captage (éboulements, marées,...). C'est dans ce contexte qu'une étude hydrogéologique des fontaines d'Yport fut réalisée par le BRGM 1968 dans le but d'améliorer l'alimentation en eau d'Yport mais également la recherche de nouvelles ressources pour l'alimentation en eau potable de la ville du Havre. Cette étude a défini un bassin d'alimentation des sources de l'ordre de 18 km². C'est en 1994 que les travaux de création d'un captage au niveau du talweg du Bois de la Vierge ont été réalisés. Ce captage correspond au captage encore utilisé à l'heure actuelle. Ce fut 20 ans après que des études concernant le captage du Bois de la Vierge ont révélé que les fontaines d'Yport ne pouvaient pas fournir les débits observés avec un bassin d'alimentation de seulement 18 Km² (Gaillard & al. 2012). Ainsi, l'étude du bassin d'alimentation du forage d'Yport réalisée par la SAFEGE en 2012 a permis de délimiter le bassin

d'alimentation connu de nos jours. Dans un premier temps, un bassin d'alimentation continu de 225 km² a été délimité en tenant compte de la piézométrie relevée par le BRGM en 2011. Par la suite, du fait de la présence de nombreuses formes karstiques dans la zone d'étude, l'interprétation de la topographie associée à l'interprétation des résultats de traçages effectués de 1974 à 2013 au niveau de pertes et de bétoires (Annexe 5) ont permis de délimiter un bassin d'alimentation dit "discontinu" de l'ordre de 140 km². Ils ont également mis en évidence que ce bassin d'alimentation discontinu du captage d'Yport était alimenté par l'eau provenant d'autres bassins superficiels tels que le bassin d'Etretat ou encore le bassin d'alimentation du captage de Radicatel, lui-même responsable d'une grande partie de l'alimentation en eau potable de la ville du Havre. C'est ainsi que deux bassins d'alimentation ont été délimités ; un bassin d'alimentation dit "continu" délimité par l'étendue de la nappe (basé sur la piézométrie), et un bassin d'alimentation "discontinu" basé sur l'assemblage des bassins topographiques identifiés par les traçages comme étant en connexion avec le captage d'Yport. En assemblant les superficies de ces deux bassins, on obtient alors un bassin d'alimentation "global" estimé à 230km² (Figure 7).





Figure 7 : Représentation des bassins d'alimentation "continu" et "discontinu" du captage d'Yport.



Carte de localisation des bétoires comprises dans le bassin d'alimentation "continu" du captage d'Yport

Figure 8 : Carte de recensement des bétoires sur le bassin d'alimentation "continu" d'Yport.

2.3.2 IMPACT DES PRATIQUES AGRICOLES

Une fois ce bassin d'alimentation identifié, on peut observer que la majeure partie du territoire compris dans le bassin d'alimentation d'Yport se trouve être recouverte par des terres agricoles (Figure 9). Ce sont en tout 83% des terrains du bassin qui sont utilisés pour l'agriculture. On trouve ainsi parmi ces terrains agricoles : 43% de cultures de blé, 17% de prairies, 10% de maïs, 9% de colza, 8% de plantes à fibres, 5% de cultures industrielles, 5% de prairies temporaires, 4% de légumes/fleurs, 3% d'orge et 1% de protéagineux. De par les pratiques culturales présentes, et la présence des bétoires (Figure 8), l'utilisation de produits phytosanitaires engendre un risque important de contamination de l'eau de surface et de l'eau souterraine. Il est important de noter qu'en Normandie, l'application de produits phytosanitaires n'est pas soumise à un calendrier d'utilisation. Les règles entourant l'utilisation des produits phytosanitaires se basent sur les conditions météorologiques (interdiction de traiter avec un vent supérieur à 19km/h). Il existe également des Zones de Non Traitement (ou ZNT) identifiées par rapport aux milieux aquatiques pour lesquelles une réglementation s'applique différemment pour chaque produit phytosanitaire. A des fins de suivis des exploitations, toute utilisation de produits phytosanitaires devra faire l'objet d'une entrée dans le registre phytosanitaire de l'exploitation dans lequel il est obligatoire de mentionner : le nº de l'îlot PAC ou le nom de la parcelle, la culture produite, le nom commercial du produit utilisé, la quantité ou la dose de produit utilisé, la date de traitement et la date de récolte (source : Chambre d'agriculture Seine-Maritime).

Si l'on met en relation l'utilisation de produits phytosanitaires et les formations géologiques, il apparait que l'eau souterraine présente un fort risque de contamination. La Figure 8 nous permet de voir que le risque de contamination de l'eau souterraine est fortement présent car l'ensemble du bassin d'alimentation est parsemé de bétoires responsables de l'infiltration rapide de l'eau dans les formations géologiques. De plus, la Figure 8 présente les bétoires recensées par le BRGM, mais il existe des bétoires qui ne sont pas visibles à la surface du sol et qui ne peuvent donc pas être recensées. On notera de plus l'absence de cours d'eau pérenne, ce qui souligne l'importance de l'infiltration sur notre site d'étude. Pour finir, l'association des bétoires avec les formations d'une croute de battance imperméable va limiter l'infiltration lente au travers du sol et empêcher la décontamination de l'eau par le sol est les micro-organismes.

Il apparait donc que l'association du captage de l'eau dans un drain karstique avec une couverture agricole importante et la présence des bétoires entraine un risque important de contamination de l'eau captée pour l'alimentation en eau potable.



Carte d'occupation des sols dans le bassin d'alimentation "continu" du captage d'Yport

Figure 9 : Carte d'occupation du sol sur le territoire du bassin d'alimentation du captage d'Yport (Pereira-Machado, V. 2017).

2.3.3 PIEZOMETRIE DE LA NAPPE D'YPORT

D'un point de vue régional, de nombreux travaux ont mis en évidence une forte diversité dans les processus de recharge et de stockage d'eau des aquifères (El Janyani. S. 2013, d'El Janyani. S. & *al.* 2014, Slimani, S. 2009). Ces disparités sont généralement dues à des variations dans la densité des bétoires, du taux de karstifications, de la transmissivité. On notera pour cela une partie des travaux d'El Janyani. S (2013) qui a mis en évidence la possibilité de regrouper les piézométries de la région Normandie selon 3 groupes avec pour caractéristiques un fort stockage de l'eau et un faible écoulement souterrain pour le premier groupe, un second groupe présentant des zones de fortes transmissivité hydraulique et pour finir un troisième groupe où l'on retrouve des zones de tamponnage des niveaux piézométriques associées à une faible capacité de stockage. On notera également les travaux d'El Janyani. S. & *al.* (2014) qui ont mis en évidence l'importance du karst d'introduction, du karst de restitution (Rodet. J. 1992) et de leurs interactions dans le processus de recharge de la nappe.

En ce qui concerne Yport, une étude piézométrique de la nappe d'Yport a permis en 2001 de dresser une carte des hauteurs d'eau moyennes au voisinage du captage d'Yport (Figure 12). On notera en premier lieu la présence de deux points de crête à l'est (90m NGF) et à l'ouest du bassin (80m NGF). Par la suite, les écoulements d'eau au sein de cette nappe s'effectueront tout d'abord par une concentration des écoulements vers une ligne de basses eaux localisée au centre du bassin d'alimentation du captage et présentant une orientation nord-nord/ouest. Une fois que les écoulements ont atteint cette ligne de basses eaux (entre 50 m NGF au sud et 0 m NGF au nord), l'écoulement principal s'effectuera dans une direction nord-nord/ouest pour atteindre la zone côtière au niveau de laquelle l'eau atteindra le niveau o m NGF. En prenant en compte les travaux d'El Janyani. S (2013 et 2014), on peut penser que le karst se trouve plus développé au centre de notre bassin sur l'axe nord-nord/ouest. Ainsi, il est possible d'émettre l'hypothèse selon laquelle les eaux s'écouleraient plus rapidement et le stockage serait moins important sur cet axe. Les niveaux piézométriques importants à l'est et à l'ouest pourraient s'expliquer par une zone présentant une karstifications et un stockage d'eau moins important.



Carte piézométrique au voisinage du captage d'Yport

Figure 10 : Carte piézométrique moyenne au voisinage du captage d'Yport (Khammari B. A. 2013)

2.4 BASE DE DONNEES

La CODAH a fourni des données mesurées au pas de temps horaire depuis 1993. Cette base de données comporte les mesures des variables suivantes : turbidité, conductivité et hauteur d'eau dans le forage d'Yport, ainsi que les débits de pompages de la station d'Yport.

Nous disposons également, grâce à cet organisme, des données collectées par 11 pluviomètres. Les mesures possédées par la CODAH sont effectuées à un pas de temps infra-horaire ; cependant, afin d'adapter le pas d'échantillonnage à la dynamique temporelle d'évolution des signaux, toutes les données ont été rééchantillonnées à un pas de temps horaire. Comme le traitement de la base de données se fait en vue d'une utilisation en prévision, nous avons également imposé que le processus soit causal. De cette manière c'est seulement l'information contenue dans les signaux durant l'heure précédant l'instant discret *k* qui est prise en compte pour calculer la valeur de l'échantillon à l'instant *k*. Le ré-échantillonnage a consisté, pour les données de précipitation, à effectuer le cumul des mesures de pluviométrie durant l'heure précédente, ou bien, pour les autres données à calculer la moyenne des valeurs de l'heure précédente.

Également, concernant les mesures des produits phytosanitaires au niveau de deux points situés à proximité du captage, nous avons réalisé un export des données à partir du site de l'ADES (<u>http://www.ades.eaufrance.fr/</u>). Les données et les traitements effectués sont présentés dans la partie suivante.

2.4.1 TURBIDITE

MESURES PAR LA SONDE 1.

Il s'agit de la mesure de turbidité effectuée à l'entrée de la station d'Yport. Cette mesure est réalisée dans la conduite de pompage, juste avant que l'eau n'entre dans la chaine de traitement. Cette installation ne fournit malheureusement pas des informations en continu sur les valeurs réelles de turbidité dans le karst. En effet, lorsque la pompe est en marche, la sonde mesure la turbidité d'une eau très proche voir identique à celle présente dans le conduit karstique. En revanche, lorsque le pompage n'est pas actif, la sonde ne peut pas mesurer l'évolution de la turbidité dans le karst et ne mesure alors que l'évolution de la turbidité due à une décantation de l'eau stagnante dans le conduit métallique de la station de pompage.

Le contrôle d'activité du pompage (actif/inactif) ne s'effectue pas par rapport au niveau de turbidité mais par rapport à un seuil de niveau d'eau dans un réservoir situé en sortie de station. Si le niveau d'eau du réservoir passe en-dessous d'un seuil, le pompage est remis en marche et vice-versa. De ce fait, la turbidité ne joue un rôle qu'au niveau de la chaîne de traitement de l'eau en déclenchant la mise en place d'un décanteur lorsqu'elle atteint une valeur supérieur ou égale à 15 NTU.

Avec un tel positionnement de la sonde 1, on obtient des courbes de turbidité présentant des diminutions progressives même pendant un pic de turbidité au sein du drain karstique (Figure 11).

Les mesures ainsi collectées par la sonde 1 débutent au 1^{er} janvier 1993 (date de sa mise en place) jusqu'à aujourd'hui.

En plus des interruptions de pompage causant un défaut de mesure, il est important de noter qu'avant le 1^{er} janvier 2000, la mesure de la turbidité était plafonnée à 100 NTU. De plus, on note de fortes lacunes pour les années 1994, 1995, 1996, 1997, 1998 et 2000. Il a donc été décidé de ne pas tenir compte; dans notre travail, des données antérieures au 1^{er} janvier 2001.





Figure 11 : Turbidité dans le forage et turbidité à l'entrée de la station de traitement durant la période allant du 11 janvier 2015 au 25 janvier 2015.

Lorsque l'on classe les données de turbidité en fonction de leur amplitude durant les années allant de 2005 à 2015, on observe que la majorité des mesures sont comprises entre o et 29 NTU (Figure 12)(44 441 entre o et 9 NTU et 3 154 mesures entre 10 et 29 NTU).Une diminution importante du nombre d'occurrences intervient ensuite lorsque l'on passe de 30 à 39 NTU (seulement 981 occurrences) suivie d'une décroissance à l'allure exponentielle. Il apparait donc que le seuil de déclenchement de la décantation, réglé à 15 NTU, se trouve régulièrement atteint. Cette rapide analyse sommaire suggère que le seuil de décantation pourrait être optimisé en fonction des coûts générés.



Figure 12 : Classes des mesures de turbidités pour les données allant de 2009 à 2015.

MESURES PAR LA SONDE 2

Une deuxième sonde mesure également la turbidité. Il s'agit de la sonde 2, qui est installée dans un réservoir alimenté par une pompe puisant en continu l'eau dans le drain karstique au lieu du forage. Cette sonde permet donc d'avoir une information sur la turbidité de l'eau souterraine à tout instant.

La sonde 2 a été installée durant l'été 2013 et fournit des données depuis le 2 juillet 2013 jusqu'à aujourd'hui. On note sur la Figure 11 que les deux mesures de turbidité présentent des différences : déphasage, amplitude des pics différents et décroissance artificielle pour la première mesure.

On note cependant que la chronique de turbidité est globalement similaire pour les deux signaux et que le pic majeur intervient au même moment. On note également que pendant la phase de diminution de la turbidité mesurée par la sonde 1, cette dernière diminue plus doucement et par palier, ce qui peut s'expliquer par le fait que le pompage n'est pas constamment actif.

Il est donc visible que, même si le pompage influence la représentativité de la mesure de la sonde 1 par rapport à la valeur de la sonde 2, la sonde 1 représente assez fidèlement la réalité.

2.4.2 CONDUCTIVITE ELECTRIQUE

La conductivité électrique est mesurée dans le drain grâce à une sonde installée depuis le 14 mars 2013. On peut noter sur la Figure 13 que la conductivité mesurée dans le forage évolue de manière inverse par rapport à la turbidité : la conductivité diminue globalement lorsque la turbidité augmente. On peut également observer que la conductivité subit des variations. On peut alors émettre l'hypothèse que la mise en route du pompage induit des arrivées d'eau différentes, par exemple venue de fissures, qui ne sont pas sollicitées lorsqu'il n'y a pas de pompage.



Figure 13 : Evolution de la conductivité en fonction de la turbidité à l'entrée de l'usine et dans le forage.

2.4.3 PIEZOMETRIE

La piézométrie est disponible du 1^{er} janvier 2015 à aujourd'hui. On peut noter que le niveau piézométrique du forage est fortement affecté par le phénomène de la marée de période 12h (Figure 14). Cependant cette chronique comporte de nombreuses lacunes, et particulièrement durant des évènements turbides ce qui limitera nos possibilités d'analyses par la suite.



Figure 14 : Niveau piézométrique mesuré au niveau du captage d'Yport.

Cependant, grâce au site de l'ADES, il a été possible de récupérer des données piézométriques sur une plus grande durée afin de disposer de l'évolution du niveau d'eau dans le milieu souterrain. Les données proviennent d'un point situé sur la commune d'Auberville-La-Renault au niveau du point référencé par le code BSS 00751X0004/S1. Ces données de piézométrie vont du 26 janvier 1982 à aujourd'hui (Figure 15). Nous pouvons ainsi grâce à ces données voir que durant les périodes allant de 1999 à 2004, de 2007 à 2011 et de 2013 à début 2015, le niveau piézométrique était haut. Nous verrons donc par la suite dans le chapitre 4 si ces différentes périodes peuvent être mises en relation avec des variations de turbidité. Le cas échéant, cela nous permettra de juger de la nécessité d'intégrer un tel paramètre dans la modélisation.



Figure 15 : Niveau piézométrique du point 00751X0004/S1 de la commune d'Auberville-La-Renault.

2.4.4 PLUVIOMETRIE.

Les mesures de précipitation mises à notre disposition proviennent de 11 pluviomètres répartis sur le bassin d'alimentation du captage (BAC) et à proximité de la zone d'étude (Figure 16). Nous avons ainsi 3 pluviomètres sur le bassin d'alimentation du captage (stations de Froberville, de Goderville et Anglesqueville), 3 stations pluviométriques proches du bassin (stations de Gonneville, Annouville et Ethainus respectivement éloignées de 500m, 600m et 1 km) et pour finir 5 stations éloignées du bassin (stations de Mannevillette, St Romain, St Nicolas, Octeville et du Havre respectivement éloignées de 5 km, 6 km, 7 km, 12 km et 14 km).





Figure 16 : Carte de répartition des pluviomètres par rapport au bassin d'alimentation du captage d'Yport.

Chapitre 2 : Présentation du Site d'Yport et des données associées

Toutes les données des pluviomètres débutent à partir du 1^{er} juillet 2009. Comme nous basons notre étude sur la relation entre pluviométrie et turbidité, c'est donc cette date qui nous servira de date de départ pour toutes les autres chroniques.

On peut noter sur la Figure 17 que les précipitations suivent un cycle saisonnier avec les maxima de pluviométrie atteints pour les mois d'octobre, novembre et décembre et que les minimums de précipitations sont présents pendant la période allant de mars à août.



Figure 17 : Précipitations mesurées par la station de Froberville de juillet 2009 à juillet 2015.

Pour finir, il est possible de noter sur la Figure 18 que la majorité des mesures appartient à la catégorie o à 1 mm d'eau. On observe également une diminution importante de la fréquence lorsque l'on passe de 1 à 1,5 mm de précipitation puis une diminution de type exponentiel de l'occurrence des classes d'intensité lorsque l'intensité des précipitations augmente. On observe donc une évolution similaire à celle de la turbidité.



Figure 18 : Représentation du nombre de mesures pour une classe d'intensité de pluviométrie.

2.4.5 VARIATION GLOBALE DES DONNEES DURANT LA PERIODE D'ETUDE

D'un point de vue global sur les données de précipitation, turbidité et hauteur d'eau dans le forage, on peut noter sur le Tableau 2 que les plus faibles variations concernent les données de hauteur d'eau. En effet, les minimums et maximums de hauteur d'eau ne varient que très peu durant les années concernées. Son interprétation sera tout de même effectuée en relation avec la turbidité pour tenter de trouver un lien avec la turbidité. Concernant les données de précipitation, on observe une faible variation de la moyenne annuelle. Pour les maximums de précipitation, on observe une faible valeur pour l'année 2010 (14,5 mm) et des valeurs augmentant au fur et à mesure pour atteindre 24,1 mm en 2013. Il en va de même pour le maximum de la turbidité qui est de 53,9 pour 2014 et qui atteint 405,3 en 2013. Pour plus de détails, les données mensuelles pour chaque année sont présentées en Annexe 6.

		Turbidité (NTU)	Précipitation (mm/h)	Hauteur d'eau (m)
2010	Minimum	0	0	57
	Maximum	245,4	14,5	57
	Moyenne	7,5	0,9	57
2011	Minimum	0,3	0	56
	Maximum	307,9	17,1	57
	Moyenne	6,0	0,9	57
2012	Minimum	0,2	0	55
	Maximum	262,9	20,6	57
	Moyenne	10,5	0,9	55
2013	Minimum	0,2	0	57
	Maximum	405,3	24,1	58
	Moyenne	8,8	о	57
2014	Minimum	0,2	0	57
	Maximum	53,9	23,2	58
	Moyenne	4,2	0	58

Tableau 2 : Tableaux récapitulatif des valeurs maximales, minimales et moyennes de turbidité, de précipitation et de hauteur d'eau dans le forage, observées de 2010 à 2014.

2.4.6 PRODUITS PHYTOSANITAIRES.

Les données concernant les produits phytosanitaires proviennent de la base de données de l'ADES. Deux points situés à proximité du forage nous fournissent des informations. Nous avons le point avec le code BSS 00568X0045/P et le point avec le code BSS 00568X0061/P. Cependant, même si ces deux points possèdent des référencements différents, il s'agit d'un seul et même point référencé sous deux codes BSS. Cependant, on note des différences dans les périodes de mesure et dans les analyses effectuées, c'est pourquoi, nous traiterons ce point comme deux entités.

Pour le point 00568X0045/P, le tri des données a fait ressortir des mesures de 15 produits phytosanitaires qui peuvent être utilisées. Les autres produits ne sont analysés que trop peu de fois (nombre de mesures inférieur à 20) et ne peuvent pas nous fournir des informations utilisables.

Les produits retenus sont donc les suivants :

Chapitre 2 : Présentation du Site d'Yport et des données associées

- Atrazine
- Atrazine déséthyle
- Atrazine déisopropyl
- Atrazine déisopropyl déséthyl
- Bentazone
- Chlortoluron
- Isoproturon
- Métazachlore

- AMPA
- Oxadiazon
- Cyproconazole
- Oxamyl
- Imidactopride
- Diflufenicanil
- Glyphosate

Pour le point avec le code BSS 00568X0061/P, environ 14 produits phytosanitaires possèdent un nombre de mesures suffisamment important (nombre de mesure supérieur à 20) pour en faire une analyse. Les produits ayant été retenus pour l'étude sont donc les suivants :

- Atrazine
- Atrazine déséthyl
- Chlortoluron
- Isoproturon
- Métazachlore
- AMPA
- Simazine

- Alachlore
- Cyanazine
- Dinoterbe
- Diuron
- Lénacile
- Diflufenicanil
- Glyphosate

Malheureusement, toutes ces données ne seront pas forcément utilisables en termes de concentration en produits phytosanitaires. En effet, si l'on prend comme exemple l'AMPA dans la Figure 19, il est visible que la mesure est constante au début puis que l'on assiste à une diminution brusque vers fin 2007 avec un établissement à un niveau constant. Cette évolution de la concentration est similaire pour de nombreux produits phytosanitaires. Ceci peut être expliqué par l'évolution des techniques de mesures qui ont permis de diminuer les seuils de quantification (seuil pour lequel 20% des mesures sont justes) et les seuils de détection (seuil pour lequel 50% des mesures sont justes).



Mesures inférieurs au seuil de détection
 Mesures valides

Figure 19 : Mesures de concentration en AMPA effectuées au point 00568X0061/P

En revanche, certaines données présentent des valeurs dépassant ces seuils. C'est par exemple le cas de l'atrazine, reportée en Figure 20. On observe également dans cet exemple la diminution des seuils de quantification et de détection. Cependant de nombreuses mesures présentent des concentrations supérieures aux seuils et pourront alors être utilisées pour identifier l'existence ou non d'un lien entre la turbidité et les produits phytosanitaires.





2.5 CONCLUSION DU CHAPITRE 2

Ce chapitre portant sur le bassin d'alimentation d'Yport nous a permis de présenter le contexte climatique et hydrogéologique dans lequel ce dernier s'inscrit. En effet, d'un point de vue climatique, la Normandie présente une variabilité des pluviométries annuelles bien plus faible que d'autres régions de France. Par exemple, à l'opposé des régions du sud de la France, les pluies en Normandie sont globalement de faible intensité mais avec une durée bien plus importante.

En ce qui concerne notre zone d'étude correspondant au bassin d'alimentation du captage d'Yport, cette dernière, de par sa localisation côtière, présente un comportement hybride côtier/continental avec un gradient de précipitation de la côte vers les terres (Nord-Ouest/Sud-Est). D'un point de vue géologique, ce bassin d'alimentation présente de nombreuses failles et un système karstique entrainant un écoulement rapide de l'eau vers les fontaines d'Yport. Il est alors ressorti qu'une telle configuration géologique, associée à ; (*i*) une couche de surface composée d'argiles affectées par de nombreuses bétoires et (*ii*) une occupation du sol majoritairement agricole, entraine un risque de pollution (turbidité, produits phytosanitaires, contamination bactériologique,...) de l'eau souterraine non négligeable.

Ainsi, afin de préparer le travail de modélisation, le Chapitre 4 présentera les travaux d'analyse des données afin de mieux caractériser le fonctionnement hydrogéologique de notre bassin.

Cependant, avant d'entrer plus en amont dans la présentation de nos résultats, il apparait nécessaire de décrire les méthodes utilisées afin que le lecteur puisse comprendre l'intérêt de ces dernières mais également les résultats que nous présenterons par la suite.

Chapitre 3 : Outils et Méthodes

CHAPITRE 3 : OUTILS ET METHODES

Chapitre 3 : Outils et Méthodes

3.1 INTRODUCTION

Le chapitre 3 a pour objet une description des méthodes employées dans notre étude. Dans les premières parties, nous présenterons des méthodes appliquées aux données : analyses corrélatoires, analyses multirésolution et décomposition en ondelettes. Puis nous aborderons la présentation des réseaux de neurones et des différentes méthodes permettant de les utiliser correctement (arrêt précoce, validation croisée, ...). Pour finir, nous nous focaliserons sur les méthodes d'échantillonnage et d'analyse des produits phytosanitaires que nous utiliserons dans le chapitre 7.

3.2 ANALYSES CORRELATOIRES

Du fait de l'hétérogénéité et de l'anisotropie structurale du karst, il est en pratique impossible de déterminer avec exactitude, par la mesure directe, la structure tridimensionnelle du système physique. L'étude fonctionnelle de l'aquifère ne peut donc pas être fondée sur la représentation physique certaine de ce dernier. En raison de cette difficulté, d'autres approches ont vu le jour, permettant d'étudier le fonctionnement du karst dans sa globalité, sans chercher à décrire son fonctionnement physique tel que les approches systémiques.

L'aquifère karstique comprend plusieurs fonctions différentes (infiltration rapide, infiltration lente, stockage,...). Il peut être décrit comme l'assemblage de fonctions qui forment un système. C'est ce qu'a proposé Mangin, A. (1975). Dans l'approche systémique, les mesures d'entrée et de sortie sont considérées comme des informations suffisantes à la compréhension globale du système. Ainsi ce dernier est interprété comme étant un opérateur transformant un signal d'entrée en un signal de sortie. De ce fait, l'appréhension des caractéristiques fonctionnelles du système passe par des méthodes d'analyse des signaux d'entrée et de sortie. Au final, l'étude des séries chronologiques des entrées et des sorties (dans notre travail, la pluie et la turbidité), permet d'obtenir des informations sur les caractéristiques fonctionnelles du système de réponse,...). Pour ce faire nous allons nous intéresser aux méthodes d'analyses corrélatoires (autocorrélation et corrélation croisée) introduites en hydrogéologie par Mangin, A. (1975).

3.2.1 ANALYSES CORRELATOIRES

Les calculs de corrélation sont des calculs statistiques qui estiment la coïncidence de l'apparition d'événements décrits par des variables. Les calculs de corrélation ont été utilisés par Mangin, A. (1975) dans le domaine de l'hydrologie en considérant que les variables sont des signaux temporels.

Ainsi, pour deux signaux temporels discrétisés **x** (x_i ; i=1, n) et **y** (y_i ; i=1, n), de moyennes respectives \bar{x} et \bar{y} , et d'écart type respectifs σ_x , σ_y le corrélogramme est défini par :

$$C_{k} = \frac{\frac{1}{n} \sum_{i}^{n-k} (x_{i} - \overline{x}) (y_{i+k} - \overline{y})}{\sigma_{x} \sigma_{y}}$$
(1)

Avec k un décalage temporel (en nombre de pas de temps), C_k le coefficient de corrélation pour le décalage k, et n le nombre de pas de temps que comporte la chronique considérée.

Deux limites peuvent être soulignées par rapport à cette définition : d'une part, le calcul de coïncidence n'est pas une preuve de causalité, alors que les corrélations sont souvent utilisées pour identifier des causalités¹

¹ Par exemple, lorsqu'il pleut de manière homogène sur un bassin, on peut noter que les niveaux piézométriques augmentent suite à la pluie. Cela ne signifie pas que les deux réservoirs dans lesquels sont mesurées les hauteurs d'eau sont connectés, mais simplement ces réservoirs reçoivent des précipitations très dépendantes l'une de l'autre : la corrélation des entrées se propage vers les sorties.

; d'autre part, l'approche utilisée effectue un simple produit des variables considérées, sans transformation nonlinéaire de ces variables. Ce calcul ne prend donc pas en compte le fonctionnement non linéaire des processus considérés. Néanmoins, en l'absence d'autres méthodes opérant sur les systèmes non-linéaires, ces approches se révèlent très utiles ; nous les présentons donc dans ce qui suit. La fonction de corrélation peut être calculée sur une seule série temporelle, on parle alors de corrélogramme simple ou d'autocorrélation, ou entre deux séries différentes ; on parle alors de corrélation croisée.

Les méthodes corrélatoires ont été utilisées dans de nombreux travaux. Outre leur présentation par A. Mangin (1975) pour être utilisées sur les hydrosystèmes karstiques, Eisenlhor, L. & al. (1997) ont cherché à valider ces méthodes d'autocorrélation et de corrélation croisée. Afin d'effectuer cette validation, Eisnenlhor, L. & al. (1997) ont conçu un modèle numérique représentant les conduits d'un karst en 2D puis 3D leur permettant de simuler le fonctionnement du système avec des paramètres connus, et ensuite d'appliquer aux chroniques simulées l'analyse corrélatoire. Pour les auteurs, une des difficultés de ce type d'approche vient du fait qu'elle prend en compte un système très global car la relation pluies-débit dépend d'une multitude de paramètres. L'interprétation des résultats obtenus en termes de caractéristiques simples et intégratrices s'avère être un exercice difficile, que ces dernières soient hydrodynamiques ou structurales. L'étude effectuée sur le modèle numérique fait ressortir le fait que la forme du corrélogramme dépend non seulement du taux de karstification et de la densité de drainage² du karst considéré, mais aussi de la proportion de l'écoulement rapide par rapport à l'écoulement de base, de la fréquence des événements pluvieux durant la période considérée et du type d'infiltration régissant le système (infiltration diffuse ou concentrée). Ainsi, d'après les auteurs, il est nécessaire de posséder des connaissances hydrogéologiques sur le système étudié avant de tirer des interprétations sur les caractéristiques structurelles et hydrodynamiques de ce système.

Néanmoins ces méthodes seront utilisées dans notre travail comme outils d'estimation de la durée de corrélation entre les signaux utiles pour nos modélisations. Ceci nous permettra d'appréhender simplement la dynamique des écoulements dans notre bassin d'étude : le bassin d'alimentation du captage d'Yport, et ainsi, pour une première approche, de réaliser un pré-dimensionnement des modèles que nous utiliserons.

AUTOCORRELATION

Lorsque l'on s'intéresse à identifier des structures au sein d'un seul signal, on peut utiliser la fonction de corrélation qui est définie par l'Eq.1, en remplaçant **y** par **x**. On peut tracer cette fonction graphiquement en fonction du décalage k (Figure 21). Comme la fonction est, par construction, symétrique par rapport à l'axe des ordonnées, seules les valeurs pour les décalages positifs sont représentées. Lorsque l'on effectue cette analyse sur des hauteurs d'eau ou les débits d'un système hydrologique, l'autocorrélation nous permet de déterminer l'effet mémoire du système étudié. Mangin, A. (1975), a défini l'effet mémoire comme étant la durée de décalage k nécessaire pour atteindre la valeur d'autocorrélation au-delà de laquelle on considère, par un test statistique, que la série temporelle n'est plus corrélée à elle-même. L'effet mémoire représente l'inertie du système. Dans ses travaux de 1975, Mangin, A. a déterminé le seuil de corrélation définissant l'effet mémoire en calculant l'erreur sur l'estimation du coefficient de corrélation C_k . Ce dernier est défini comme étant la durée nécessaire pour que l'autocorrélation atteigne une valeur de o, 2 (Mangin, A., 1981; Larocque, M. & *al.*, 1998).

Lorsque l'on travaille sur les chroniques de débit, un effet mémoire court signifie que le système dissipe rapidement l'information. Dans le cas d'hydrosystèmes karstiques, Mangin, A. a alors émis l'hypothèse que l'on pouvait associer un effet mémoire court à une karstification importante et à un stockage de l'eau faible, corollaire de l'évacuation rapide des débits des crues successives. On observe ce phénomène par exemple au niveau du karst d'Aliou, dans les Pyrénées ariégeoises dont l'effet mémoire est de seulement 4 jours. A l'opposé, le karst de Fontestorbes, dans les Pyrénées ariégeoises également, présente un long effet mémoire (60 jours) et peut indiquer la présence de réserves importantes (Mangin, A., 1984).

² La densité de drainage correspond à la longueur du réseau hydrographique par unité de surface du bassin.

Sur la Figure 21, nous avons représenté l'autocorrélation de la turbidité à la station d'Yport. Dans cet exemple, l'utilisation du seuil de 0,2 peut être envisagée pour déduire l'effet mémoire de 162 heures soit environ 7 jours.





CORRELATION CROISEE

La corrélation croisée est utilisée sur deux séries chronologiques différentes (par exemple, pour notre travail, la pluie et la turbidité) ; comme présenté dans l'équation 1.

Ainsi on calcule le corrélogramme croisé représentant la corrélation entre les deux signaux **x** et **y** en fonction d'un décalage temporel. Un exemple de représentation graphique du corrélogramme entre des pluies et une turbidité est donné dans la Figure 22. Contrairement à l'autocorrélation, le corrélogramme de la corrélation croisée n'est pas symétrique. Comme nous nous intéressons à la partie causale des relations entre la pluie et la turbidité, nous n'avons tracé que la partie droite, pour les *k* positifs du diagramme (c'est la pluie qui créé la turbidité et non l'inverse). Ainsi, en observant l'effet mémoire dans le cas d'un corrélogramme croisé, il est possible de déterminer la durée pour laquelle les deux séries sont dépendantes, c'est-à-dire l'historique de pluie permettant de calculer la turbidité.



Figure 22 : Fonction de corrélation croisée (zoom des valeurs pour les décalages allants de 0 et 2000) entre la pluviométrie relevée à Goderville et la turbidité mesurée à l'entrée de la station durant la période allant du 1^{er} juillet 2009 au 1^{er} juillet 2015.

Dans l'exemple présenté sur la Figure 22 on peut déterminer plusieurs informations intéressantes ; tout d'abord la valeur maximale du corrélogramme croisé n'est pas égale à 1 comme c'était le cas pour l'autocorrélation. Plus cette valeur maximale est importante, plus les signaux sont linéairement dépendants. Il peut arriver que la valeur maximale soit toujours inférieure à 0,2, auquel cas la relation linéaire entre la pluie et le débit est très faible. La première information pouvant être extraite de cette représentation est la durée associée au maximum de réponse du pic observé. Nous appellerons ce temps « temps de réponse », car il traduirait physiquement le temps que mettrait l'eau associée au maximum de pluie pour produire la turbidité. Dans la Figure 22, le temps de réponse est de 37 heures.

3.3 ANALYSES EN ONDELETTES ET ANALYSES MULTIRESOLUTION

Au contraire de l'analyse de Fourier utilisant une décomposition du signal sur une base orthogonale via des exponentielles complexes (à durée infinie dans le temps), la décomposition en ondelettes utilise des fonctions représentant une oscillation finie (localisée dans le temps). En mathématique, une ondelette se décrit généralement par une fonction de carré sommable le plus souvent de type oscillante avec un certains nombres de moments nuls dans l'espace euclidien $\mathbb{R} \times \mathbb{R}$ +*. Plusieurs ondelettes sont utilisables, on notera en particulier les ondelettes de Haar (Haar, A., 1910), de Morlet, de Daubechies (Daubechies, I., 1992) et les dérivées de gaussiennes, comme le chapeau mexicain. Il est à noter que certaines ondelettes sont adaptées à la transformée en ondelettes continue (Morlet par exemple), et d'autres plutôt à l'analyse multirésolution (Daubechies et Haar par exemple).

La propriété principale des analyses en ondelettes par rapport à la transformée de Fourier réside dans leur propriété de conservation de l'information temporelle. Elles peuvent donc s'appliquer à des signaux non stationnaires et permettre d'accéder à une description des non stationnarités. Pour cela, les ondelettes, grâce à leur fenêtre localisée dans le temps, peuvent êtres compressées et dilatées temporellement par un facteur qui permet de déterminer l'échelle temporelle de la représentation. Une échelle temporelle est une grandeur temporelle équivalente à une période dans l'analyse de Fourier. Ainsi, grâce à la dilatation temporelle de la fenêtre, on peut décomposer le signal selon plusieurs échelles et obtenir une analyse multi échelles. Initialement décrite par J. Morlet dans ses travaux en géophysique de 1983, la théorie sur les ondelettes continues fut développée par des travaux tels que les travaux de Grosmann, A. et de Meyer, Y. sur les fondements mathématiques des ondelettes (Hubbard, B. B., 1995).

On peut distinguer deux types de transformations en ondelettes : la transformée en ondelettes continue et la transformée en ondelettes discrète. Les analyses multirésolution sont la version discrète des analyses en ondelettes continue. Ces analyses discrètes sont de plus en plus utilisées dans de nombreux travaux pour l'étude des relations entre deux paramètres hydrologiques comme les pluies et le débit d'une source karstique ou d'un cours d'eau (Labat, D. & *al.*, 2002 ; Labat, D. & *al.*, 2004), comme les précipitations et les débits d'un cours d'eau, la hauteur d'eau dans un puits et les concentrations en nitrate, chlorure et sodium (Kang, S. & *al.*, 2007) ou bien la seule analyse des variabilités annuelles des débits d'une rivière (Coulibaly, P. & *al.*, 2004). Les méthodes d'analyses multirésolution ont également montré leur utilité dans le but de décomposition du signal en vue d'une utilisation des composantes dans des processus de modélisation. On notera ainsi les travaux couplant analyse multirésolution et réseaux de neurones effectués par Adamowski, J. & *al.* (2010) pour la prédiction des débits d'une rivière ou encore les travaux de Salermo, F. & Tartari, G. (2009) utilisant le couplage d'un modèle hydrologique de surface et l'analyse multi résolution et Johannet, A. & *al.* (2012) pour la prévision des crues karstiques. Afin de comprendre leur intérêt dans les études hydrogéologiques, il est important d'appréhender le fonctionnement des analyses en ondelettes.

3.3.1 TRANSFORMEE EN ONDELETTES CONTINUE

La transformée en ondelettes continues correspond au calcul des coefficients d'ondelettes pour tous les décalages temporels et dilatations de la fenêtre temporelle de l'ondelette.

La transformée en ondelettes se calcule en utilisant l'expression suivante :

$$g(s,\tau) = \int_{-\infty}^{+\infty} f(t) \psi_{s,\tau}^* dt$$
⁽²⁾

Où, * représente l'ondelette complexe conjuguée de l'ondelette $\Psi_{s,\tau}$, τ le paramètre de translation temporel et S le facteur de dilatation de la fenêtre. Ainsi, le facteur S permet de générer les différentes échelles du signal guand le paramètre τ permet le balayage temporel du signal.



Figure 23 : Signal composé dans sa première partie d'une sinusoïde de période 30 et d'amplitude 1 et dans sa deuxième partie d'une sinusoïde de période 15 et d'amplitude 1.

Pour illustrer, prenons par exemple un signal sinusoïdal composé successivement de 2 sinusoïdes finies dans le temps et de périodes différentes, la première période de 30 s, la seconde période de 15 s. Ce signal est représenté en Figure 23 et permet d'illustrer une non stationnarité assez simple. Après transformation en ondelettes continue en prenant une ondelette de Morlet, on obtient la représentation temps-échelle qui est

habituellement représentée comme en Figure 24: le temps est en abscisse, les échelles en ordonnée et l'amplitude de la puissance contenue dans le signal en niveaux de couleur.



Figure 24 : Décomposition en ondelettes continue (ondelette de Morlet) du signal présenté en Figure 23. La couleur rouge représente les composantes de faible puissance et la couleur jaune/blanc représente les signaux de forte puissance.

La limitation principale de la transformée en ondelettes continue est qu'elle n'est pas la mieux adaptée à la reconstruction du signal d'origine, ou à l'extraction de ses composantes en raison de la redondance inhérente à la transformée continue. Ceci induit donc des erreurs lors d'opérations de filtrage.

3.3.2 ANALYSE MULTIRESOLUTION

L'analyse multirésolution consiste en la décomposition d'un signal original en « n » niveaux de décomposition. La première étape consiste à effectuer la convolution du signal avec une ondelette (Daubechies, Harr,...). De la convolution résulte le premier niveau de décomposition et un résidu. Le processus est répété sur le résidu jusqu'à la décomposition complète du signal, après avoir mis en œuvre toutes les dilatations possibles compte tenu de la durée temporelle finie du signal. On obtient ainsi « n » niveaux de décomposition et 1 niveau de décomposition « n+1 » qui correspond au résidu final. Les niveaux de décomposition sont alors la représentation des basses ou hautes fréquences présentes dans le signal original. L'intérêt principal de cette analyse est que la transformation inverse permet d'obtenir exactement le signal initial : ce dernier est la somme de décompositions et du résidu. Il est ainsi possible de retrouver le signal original en effectuant la somme de tous les niveaux de décomposition. La fenêtre temporelle d'analyse est définie comme dyadique (en puissance de 2). Ainsi, par exemple, la décomposition d'un signal de 1024 pas de temps ($1024=2^{10}$) conduit à 11 niveaux de décomposition.

On a ainsi une décomposition suivant le schéma de la Figure 25.



Figure 25 : Schéma opérationnel de la décomposition en ondelettes discrète.

Dans la Figure 25, les filtres h_n et g_n sont liés respectivement à la fonction d'échelle et à l'ondelette. $r_j(k)$ représente le résidu (ou approximation) et $d_j(k)$ la décomposition (ou détail) ; ils sont définis pour chaque échelle *j* selon les équations d'échelle suivantes :

$$r_{j}(k) = \sum_{i=0}^{l} h[i]r_{j-1}[2k-i]$$

$$d_{j}(k) = \sum_{i=0}^{l} g[i]r_{j-1}[2k-i]$$

$$(3)$$

$$(4)$$

Afin de poursuivre notre illustration sur le signal de la Figure 23, la décomposition multirésolution de ce signal conduit aux approximations et détails présentés en Figure 26 pour la décomposition du signal artificiel de la Figure 23. Avec cette méthode, le nombre de composantes extraites dépend de la longueur du signal.



Figure 26 : Décomposition en ondelettes discrète du signal artificiel de la Figure 23.

3.4 MODELISATION HYDRODYNAMIQUE PAR RESEAUX DE NEURONES

3.4.1 MODELISATION PLUIE-TURBIDITE

RAPPELS SUR LES MODELES HYDROLOGIQUES.

La modélisation hydrologique ou hydrogéologique consiste en l'utilisation d'un modèle dans le but de reproduire, à partir de la pluie tombée sur le bassin versant, la réponse du système (traduite en hauteur d'eau, débits ou bien en turbidité dans notre cas d'étude). Les modèles sont le plus souvent exprimés sous forme mathématique, ils peuvent être séparés en deux catégories suivant les grandeurs servant à les concevoir.

Le modèle est dit *stochastique* dans le cas où les grandeurs l'influençant sont des variables aléatoires représentées par une fonction de distribution de probabilité. Ces modèles mettent l'accent sur la compréhension de la relation structurelle régissant les variables de sortie du bassin (turbidité) et les variables d'entrée (pluies).

Le modèle est dit déterministe si aucune des grandeurs n'est considérée comme aléatoire. La grande majorité des modèles hydrologiques est de type déterministe. L'approche est basée sur la simulation physique du système hydrologique et c'est cette dernière hypothèse qui constitue leur principale limite. En effet, dans le cas d'un modèle physique, décrit par une équation différentielle assortie de paramètres à déterminer, il est supposé que le calcul exact, ou la mesure, des paramètres est possible. Cependant, dans de nombreux cas, les phénomènes naturels sont très variables et leur mesure très incertaine. Ainsi, les modèles physiques sont finalement tous limités par le grand nombre de paramètres à mesurer et par les limites des connaissances des systèmes naturels complexes (Kurtulus, B., 2005).

Parmi ces deux grandes familles, on pourra retrouver différents modèles tels que :

- les modèles *empiriques* pour lesquelles l'objectif est de reproduire les variations des sorties en fonction des entrées sans tenter d'identifier les processus intrinsèques. Ces modèles sont également appelés *boîtes noires*.
- les modèles *conceptuels* se basant sur une discrétisation du système étudié en différents sousensembles.

Encore une fois, ces modèles peuvent être différentiés en deux grandes familles suivant la discrétisation temporelle effectuée :

- les modèles *évènementielles* se baseront sur la modélisation d'un système lors d'un évènement (pluie, crue, turbidité,...).
- les modèles *continus* permettront une modélisation à plus long terme. Ces modèles sont connus pour leurs fortes dépendances aux variables d'état des bassins afin d'expliquer les variations des stocks d'eau en fonction des années.

Ainsi, comme justifié dans le chapitre 1, en suivant cette classification, notre étude sera effectuée en utilisant des modèles empiriques de type boîte noire et avec une modélisation de type évènementielle.

LE MODELE « BOITE NOIRE »

Lorsque l'on effectue une modélisation de type « boîte noire », on considère les écoulements de manière globale en assimilant la structure hydrogéologique considérée à un système mettant en relation des données d'entrée (historique de précipitation), et des données de sortie (historique de turbidité). Un tel modèle permet la conceptualisation d'une fonction reliant l'entrée à la sortie. Cette fonction reflète bien évidement des caractéristiques internes du système mais celles-ci ne sont pas toujours identifiables au travers des paramètres du modèle.



Figure 27 : Représentation de la réponse impulsionnelle d'un système linéaire.

On cherche, en l'absence de connaissance sur la structure physique, à étudier la relation entre l'entrée et la sortie. Dans le cas de la modélisation de processus linéaires, on résout ce problème en recherchant un opérateur linéaire *h* (réponse impulsionnelle) dont l'application sur l'entrée entraine le calcul de la sortie (Figure 27). C'est par exemple le cas de la méthode de l'hydrogramme unitaire (Jacquet, J., 1960). Cependant en hydrologie, du fait, a minima, des saisons et de l'évapotranspiration, tous les systèmes sont non linéaires; l'identification de la réponse impulsionnelle ou de sa transformée de Fourier : la fonction de transfert conduit donc à de fortes approximations que l'on peut pallier en utilisant plusieurs fonctions de transfert qui varient en fonction de l'état du système étudié : schématiquement crue ou étiage.

Dans notre cas d'étude, la relation pluie-turbidité est fortement non linéaire, et grandement inconnue, comme nous l'avons présenté dans le chapitre 1, il convient donc de la modéliser par une méthode permettant de prendre en compte cette non linéarité et pouvant s'appliquer sans connaissance complète *a priori* des phénomènes physiques intervenants. C'est pourquoi nous nous sommes orientés vers une modélisation de type réseaux de neurones formels comme nous allons l'expliciter ci-après.

3.5 LES RESEAUX DE NEURONES FORMELS

LE NEURONE FORMEL

Un réseau de neurones formels est défini par l'assemblage de plusieurs neurones connectés entre eux suivant une architecture particulière. Un neurone formel ou appelé dans la suite plus simplement neurone, est un opérateur mathématique qui calcule deux actions (Figure 28):

- Le calcul de son potentiel

$$v = \sum_{j=1}^{n} w_j \ e_j \tag{5}$$

Où v est le potentiel, \mathcal{W} le vecteur des paramètres, \mathcal{C} le vecteur des entrées, j un entier naturel (j=1,n), n est le nombre maximal d'entrées.

- Le neurone calcule sa sortie par une transformation généralement non linéaire :

$$s = f(v) \tag{6}$$

Où s est la sortie et f(.) est la fonction d'évaluation. Deux types de fonctions d'évaluation sont utilisées : la fonction identité lorsque le neurone est linéaire et la fonction « sigmoïde » lorsque le neurone est non-linéaire :

$$f(v) = \frac{1}{1 + e^{-\beta v}} \tag{7}$$

Avec β la pente de la sigmoïde.

Avec la fonction sigmoïde la sortie évolue quasi linéairement pour des potentiels faibles, et atteint des valeurs de saturation (o et +1) pour des potentiels importants.



LE PERCEPTRON MULTICOUCHE

ARCHITECTURE

Le perceptron multicouche (Figure 29) est un ensemble de neurones associés en couches. Les neurones d'une couche reçoivent les sorties de tous les neurones de la couche précédente ; les neurones d'une même couche ne sont pas connectés les uns aux autres. La couche d'entrée n'est pas une couche de neurones car celleci n'effectue aucune modification de l'information. On distingue deux types de neurones ; les neurones dits « cachés » qui participent aux calculs mais dont on ne connait pas de valeur de sortie mesurée, et les neurones de sortie dits « visibles », dont on connaît les valeurs mesurées ou *désirées*.



Figure 29 : Représentation schématique du perceptron multicouche.

APPROXIMATION UNIVERSELLE ET PARCIMONIE

Le perceptron comportant deux couches, dont la première de neurones avec une fonction d'évaluation sigmoïde, et la couche de sortie de neurones linéaires, possède 2 propriétés remarquables : (*i*) l'approximation universelle : il peut approximer toute fonction non-linéaire différentiable avec une précision arbitraire qui dépend du nombre de neurones de la couche cachée (Hornik, K. & *al.*, 1989), et (*ii*) de parcimonie, c'est à dire que pour approcher la fonction désirée avec une précision arbitraire il a besoin de moins de paramètres qu'un modèle statistique dont la sortie dépend linéairement des paramètres (Barron, A.R., 1993).

LES DIFFERENTS RESEAUX UTILISES FONDES SUR LE PERCEPTRON MULTICOUCHES

Nous allons maintenant décrire trois architectures de perceptrons multicouches différentes, qui correspondent chacune à un filtre numérique non-linéaire. On distinguera en particulier les modèles statiques (ou à réponse impulsionnelle finie) dont la sortie ne dépend que d'entrées extérieures, appelées souvent exogènes, des modèles dynamiques (ou à réponse impulsionnelle infinie) dont la sortie k dépend également de leur propre état interne au instant précédents.

Le modèle non bouclé pour identifier un processus statique
Ce modèle associe des entrées exogènes à la sortie du processus. Par exemple, la Figure 30 présente un réseau statique à deux couches. Ce type de réseau correspond à un filtre numérique non récurrent. Il effectue le calcul suivant:



Figure 30 : Schéma d'un réseau de neurones non bouclé

Le modèle non bouclé est qualifié de statique car, lorsque ses entrées exogènes ne varient plus (par exemple la pluie), sa sortie reste constante. Or ce phénomène n'est pas observé en hydrologie. Il faut donc un modèle qui puisse prendre en compte la dynamique du système hydrologique. Ceci s'obtient en utilisant en entrée une variable décrivant l'état du système que l'on modélise, comme par exemple la sortie du processus.

Le modèle statique pour l'identification d'un processus dynamique

Le modèle suivant est statique car ses entrées sont toutes des entrées mesurées sur le processus à identifier. Cependant, ce modèle reçoit la valeur observée de la grandeur de sortie aux pas de temps précédents. Cette grandeur apporte une information que l'on qualifie d'état du système. En ce sens, on peut dire que le modèle est « dirigé » (Figure 31) par les observations. Il effectue le calcul suivant:

$$y_{c}(k) = w_{s,0}c + \sum_{j=1}^{m} \left[w_{s,j}f_{j}\left(\sum_{i=0}^{n} w_{j,i}x_{i}(k)\right), w_{s,j}f_{j}\left(\sum_{i=0}^{n} w_{j,i}y_{d}(k-1)\right) \right]$$
(9)



Figure 31 : Représentation schématique d'un réseau bouclé dirigé à deux couches

> Le modèle récurrent pour l'identification d'un processus dynamique

Dans les réseaux bouclés, la sortie de certains neurones cachés ou celle du neurone de sortie est renvoyée en entrée du réseau. Ce bouclage donne un comportement dynamique au modèle, la sortie ne dépend plus seulement des entrées exogènes du réseau, mais également de sa sortie aux instants précédents. Ce modèle est appelé par la suite « non dirigé », car il n'est pas dirigé par les observations de la variable de sortie. Il est représenté en (Figure 32) pour un ordre (nombre de connexions récurrentes) de 1. Ce dernier effectue le calcul suivant:

$$y_{c}(k) = w_{s,0}c + \sum_{j=1}^{m} \left[w_{s,j}f_{j}\left(\sum_{i=0}^{n} w_{j,i}x_{i}(k)\right), w_{s,j}f_{j}\left(\sum_{i=0}^{n} w_{j,i}y_{c}(k-1)\right) \right]$$
(10)
$$x_{n}(k) = \text{variable d'entrée à l'instant } k$$
$$y_{c}(k) = \text{sortie calculée à l'instant } k-1$$
$$y_{d}(k) = \text{sortie calculée à l'instant } k$$
$$f_{m}(v) \text{ et } f_{s}(v) = \text{ fonction d'activation } c = \text{Entrée constante}$$
$$y_{c}(k-1) = \text{ fonction d'activation } c = \text{Entrée constante}$$

Figure 32 : Représentation schématique d'un réseau bouclé non dirigé d'ordre 1 à deux couches

3.5.2 L'APPRENTISSAGE

L'apprentissage correspond au calcul du vecteur W des paramètres du modèle afin de minimiser une fonction de coût correspondant généralement à l'erreur quadratique moyenne.

La fonction la plus couramment utilisée pour la minimisation de la fonction de coût est la fonction des moindres carrés:

$$j(w) = \sum_{i=1}^{k} \left[y_d^i - y_c^i(x^i, w) \right]^2$$
(11)

Avec j(w) la fonction de coût dépendant de la matrice des paramètres w. La sortie du modèle $y_c^i(x^i, w)$ est ainsi calculée pour les variables x^i . On notera également y_d^i la sortie désirée, qui est la sortie mesurée.

Le réseau va donc s'adapter jusqu'à ce que $y_c^i(x^i, w)$, la sortie du modèle, soit égale à y_d^i , pour tous les indices \mathbf{i} correspondant aux instants de la séquence d'apprentissage.

Cependant, les réseaux de neurones étant non linéaires en leurs paramètres, la méthode d'apprentissage des moindres carrés n'est pas applicable. De nombreuses règles d'apprentissages sont décrites par Dreyfus, G. & *al.* (2008) parmi lesquelles se trouve la règle de Levenberg-Marquardt (Hagan, M.T., & Menhaj, M.B. 1994) qui

correspond actuellement à la règle la plus performante pour le perceptron multicouches. La règle de levenberg-Marquardt correspond à une règle de second ordre utilisant l'inversion de la matrice Hessienne (Press, W.H., & *al.* 1996). L'avantage de cette méthode, contrairement aux autres méthodes de second ordre, réside dans le fait qu'elle n'utilise que des dérivées premières de la fonction de coût. En revanche, dans le cas de modèles de grande taille possédant un grand nombre de paramètres, cet algorithme ne peut être utilisé car il générerait une matrice Hessienne de trop grande taille. Dans notre cas d'étude, les modèles étant de petites tailles (moins de 1000 paramètres), nous utiliserons donc cet algorithme.

Il est cependant important de noter que la fonction de coût peut présenter des minimums locaux. Ainsi, l'initialisation des paramètres du modèle peut induire une localisation du modèle dans un minimum local, ce qui entraine que le minimum optimal de la fonction de coût ne sera pas atteint. En pratique, pour éviter cette convergence du modèle dans un minimum local de la fonction de coût, l'initialisation des paramètres s'effectue selon une suite aléatoire. Ainsi, la multiplication des apprentissages utilisant chacun des initialisations différentes permet d'augmenter la probabilité d'atteindre le minimum optimal de la fonction de coût. Ceci a été mis en avant dans les travaux de Darras, T. (2016) pour lesquels, l'utilisation de 10 modèles dont les paramètres sont initialisés aléatoirement a permis d'obtenir un modèle moyen robuste dépendant peu de l'initialisation.

3.5.3 LA GENERALISATION

DILEMME BIAIS-VARIANCE

La généralisation d'un modèle est son aptitude à bien fonctionner sur des évènements ne faisant pas partie de sa base de données d'apprentissage. La généralisation se mesure donc sur des ensembles disjoints de ceux de l'apprentissage que l'on appelle ensembles de test ou ensembles de validation. En général l'ensemble de test est utilisé à l'issue de la conception du modèle pour mesurer sa qualité sur des données qui n'ont servi ni à sa calibration (apprentissage) ni à son dimensionnement. Geman, S. & al. (1992) a montré que l'erreur de sortie du modèle pouvait se décomposer en la somme de 3 termes :

- La variance du bruit (présent dans les observations, par exemple provenant du matériel de mesure)
- Le biais (la différence entre l'espérance mathématique des valeurs calculées par le modèle et l'espérance mathématique des valeurs mesurées)
- La variance (représente la variance des valeurs calculées par des modèles obtenus avec tous les ensembles d'apprentissage possibles)

Comme la variance ne peut être négative, l'erreur de prédiction moyenne ne peut être inférieure à l'erreur présente dans les données observées. Le biais et la variance possèdent donc une variation inverse l'un envers l'autre. Comme la variance augmente avec le nombre de paramètres du modèle (sa complexité), ceci se traduit par une diminution du biais avec une augmentation du nombre de paramètre du modèle. Kong A Siou, L. & *al.* (2012)

Le dilemme biais-variance implique donc de rechercher un équilibre entre biais et variance. Cet équilibre correspond à une valeur minimale de la somme entre biais et variance au-delà de laquelle, le modèle est considéré comme « *surajusté* ».

On peut voir pour l'exemple de la Figure 33, que le biais diminue au départ lors de l'augmentation du nombre de paramètres libres du modèle alors que la variance augmente dans cette même partie. Pour la partie oscillante du biais au début de la courbe puis croissante à la fin de la courbe, Johannet, A. (2011) émet l'hypothèse de l'influence des incertitudes et du bruit de mesure qui entache le biais d'une erreur liée à la variance.



Figure 33 : Evolution du biais (trait pointillé) et de la variance (trait continu) en fonction du nombre de paramètres libres du modèle

(Johannet, A. 2011).

Considérant l'architecture du perceptron multicouche, le nombre de paramètres libres est dépendant du nombre d'entrées, de l'ordre du modèle et du nombre de neurones cachés. Ces trois « hyperparamètres » de l'architecture doivent donc être sélectionnés de manière à optimiser la complexité du modèle pour atteindre l'optimum entre biais et variance.

METHODES DE REGULARISATION

Afin d'éviter le sur-ajustement, deux méthodes de régularisation peuvent être utilisées en combinaisons. Pour ces deux méthodes, la base de données doit faire l'objet d'un découpage en différents ensembles. Ainsi, outre l'ensemble de test, nous définissons des ensembles dit d'apprentissage, un/des ensembles de validation (pour la méthode de la validation croisée) et un/des ensembles d'arrêt (pour l'arrêt précoce).

LA VALIDATION CROISEE

La validation croisée permet d'estimer la qualité du modèle par rapport à l'ensemble des données disponibles (hormis l'ensemble de test). Elle est utilisée dans notre travail afin d'optimiser l'architecture du modèle en sélectionnant les variables les plus pertinentes (Dreyfus, G. & *al.* 2008, Kong-A-Siou, L. 2011).Pour cela la base d'apprentissage est divisée en *e* sous-ensembles. Les e-1 ensembles servent à l'apprentissage tandis qu'un d'entre eux est utilisé pour le calcul de validation. L'erreur effectuée sur l'ensemble de validation,

notée J_e est conservée, puis le processus est itéré pour chaque ensemble de validation précédemment défini.

À l'issue de la validation croisée, le score dit de « validation croisée » s'obtient ensuite en calculant la moyenne des *K* scores précédemment obtenus sur chacun des ensembles de validation :

$$S = \frac{1}{K} \sqrt{\sum_{e=1}^{K} (J_e(C))^2}$$
(12)

Avec S, le score de validation, K, le nombre d'ensembles, et $J_e(C)$, la fonction de coût calculée sur l'ensemble e (Figure 34).

Une fois calculé le score de validation pour chaque ensemble, on effectue la moyenne et/ou la médiane des scores de validation afin d'obtenir un score global de validation croisé.



Figure 34 : Illustration du principe de validation croisée.

Des essais sont alors réalisés pour sélectionner la meilleure architecture possible en définissant les entrées, ordre et nombre de neurones cachés. Pour une bonne comparaison des résultats, les mêmes caractéristiques de réseau sont adoptées lors de chaque simulation. Un seul hyperparamètre à la fois sera modifié pour une meilleure observation du changement apporté par sa modification. La Figure 35 présente un exemple de variation du score de validation croisée en fonction du nombre de neurones de la couche cachée pour le dimensionnement du perceptron multicouche dirigé effectuant une prédiction à 24h avec des données au pas de temps 6h qui sera présenté dans le chapitre 5.





L'ARRET PRECOCE

La deuxième méthode de régularisation utilisée est l'arrêt précoce. Cette méthode, appliquée au fur et à mesure que l'apprentissage se déroule, permet de garder en mémoire l'itération correspondant à la meilleure capacité de généralisation du modèle. Pour cela, à chaque itération d'apprentissage, la fonction de coût est calculée sur deux ensembles sélectionnés dans la base de données. Le premier ensemble, l'ensemble d'apprentissage, permet de calculer la valeur de la fonction de coût sur cet ensemble d'apprentissage, qui diminue généralement au fur et à mesure que le réseau apprend. Le deuxième ensemble, l'ensemble d'arrêt, est utilisé pour arrêter l'apprentissage. En effet, généralement, l'erreur sur l'ensemble d'arrêt commence par diminuer lorsque l'apprentissage se déroule puis, au bout d'un certain temps, se met à augmenter, signifiant par-là que le modèle n'est plus capable de généraliser à des exemples inconnus. On arrête ainsi l'apprentissage

lorsque l'erreur calculée sur l'ensemble d'arrêt atteint sa valeur minimale. En pratique, si sur 100 itérations (une itération correspond à une modification des paramètres du modèle), l'itération 20 s'avère la meilleure pour l'ensemble d'arrêt, nous retenons le modèle avec les paramètres obtenus à l'issue de la 20^{ième} itération. Il est à noter que la sélection de l'ensemble d'arrêt se fait sur des évènements représentatifs des besoins du futur modèle. Si le modèle a pour objectif de prédire des valeurs extrêmes, l'ensemble d'arrêt est composé d'un ou de plusieurs évènements pour lesquelles la généralisation du modèle est la meilleure.

3.5.4 CONCEPTION D'UN MODELE

La conception d'un modèle consiste à sélectionner le modèle et à calculer ses paramètres :

La sélection du modèle consiste à définir :

- L'ordre du modèle (modèles récurrents)
- Les données d'entrée
- Le nombre de neurones dans les couches cachées.

Il existe d'autres hyperparamètres à fixer tels que :

- La suite aléatoire utilisée pour l'initialisation des paramètres avant l'apprentissage

C'est après avoir sélectionné tous les hyperparamètres et calculé les paramètres du modèle que les résultats sur l'ensemble de test peuvent être obtenus.

RNFPRO : SIMULATEUR DE RESEAUX DE NEURONES FORMELS

Pour modéliser les relations pluie-turbidité par les réseaux de neurones formels, le logiciel « RNFPro » a été utilisé. Il a été développé en Java grâce à une collaboration entre l'Ecole des Mines d'Alès et la société Geonosis (Bertin, D.).

3.5.5 CRITERE DE QUALITE

CRITERE DE NASH (R^2)

Le critère de Nash (Nash, J.E., 1970), autrement appelé *coefficient de détermination* en statistique, est un critère normé habituellement utilisé en hydrologie comme critère de qualité car il permet de comparer l'écart quadratique moyen des mesures rapporté à leur variance.

$$R^{2} = 1 - \frac{\sum_{k=1}^{n} (y_{o}^{k} - y^{k})^{2}}{\sum_{k=1}^{k} (y_{o}^{k} - \overline{y_{o}})^{2}}$$
(13)

Avec y_o^k la valeur désirée au temps discret k, y^k la valeur calculée au temps discret k et y_o^k la valeur moyenne de la chronique considérée.

Un critère de Nash proche de 1 signifie que l'erreur entre les valeurs mesurées et les valeurs calculées est réduite. En hydrologie, un résultat proche de 0,8 est considéré comme moyen ou acceptable en fonction de la difficulté de la modélisation.

On note que, pour un jeu de données fixé, quand on cherche à minimiser l'erreur quadratique lors de l'apprentissage, cela revient à maximiser le critère de Nash.

RETARD DU PIC (PEAK DELAY: P_D)

Le critère de Nash permet de juger de la qualité du modèle sur l'ensemble d'une chronique, par exemple, d'un événement de crue. Cependant, le critère de Nash est assez peu sensible au décalage temporel entre la

sortie du modèle et la chronique observée en particulier au niveau d'un pic. Dans le cas de la prévision des crues éclair, Artigue, G. (2012) a mis en évidence la nécessité d'utiliser un critère prenant en compte le décalage des deux pics. Ce critère correspond au retard observé entre la valeur maximale désirée et la valeur maximale calculée. Il répond à l'équation suivante :

$$P_D = k_o^{\max} - k^{\max} \tag{14}$$

Avec P_D le retard du pic, k_o^{max} le temps d'occurrence de la valeur maximale observée et k^{max} le temps d'occurrence de la valeur maximale calculée.

Si le P_D est positif, cela signifie que la sortie est en retard par rapport à l'occurrence du pic et au contraire, si la valeur de P_D est négatif, cela signifie que la sortie est en avance par rapport à l'occurrence du pic (Figure 36). Ainsi, lorsque $P_D = 0$ les pics mesuré et simulé sont synchrones.

LE POURCENTAGE DE RESTITUION DU PIC (PEAK PERCENTAGE; P_P)

Ce critère tout comme le « *Retard du Pic* » est un critère se concentrant sur la qualité de prévision des pics. Le « *Pourcentage de Restitution du Pic* » vise à quantifier la différence d'amplitude du pic calculé par rapport au pic mesuré en effectuant le ratio de ces deux valeurs. Ce critère se calcule ainsi :

$$P_P = 100 \frac{Max(y)}{Max(y_o)} \tag{15}$$

Avec P_p le pourcentage de restitution, fonction de $_{Max(y)}$ la valeur maximale calculée par le modèle est

 $_{Max(y_o)}$ la valeur maximale observée. Plus le P_p est proche 100%, plus la valeur maximale calculée est proche de la valeur maximale désirée. Cependant, ces deux valeurs ne sont pas nécessairement synchrones (Figure 36).



Figure 36: Représentation des critères du P_P et P_P en fonction de la prédiction (ligne en tirets) et de l'observée (ligne en pointillée).

3.6 ANALYSE DES PRODUITS PHYTOSANITAIRES

3.6.1 ECHANTILLONNAGE

ECHANTILLONNAGE ACTIF

L'échantillonnage actif est un échantillonnage ponctuel qui ne peut donner d'informations que relatives à l'instant de prélèvement. Il consiste à prélever un volume d'eau (1L dans notre cas d'étude pour effectuer deux réplicas de 500mL). Le contenant doit être un flacon ou une bouteille permettant de limiter les interactions avec les éléments présents dans l'échantillon. Une fois le prélèvement effectué, l'échantillon n'est pas stable et des réactions physico-chimiques ou même biologiques peuvent avoir lieu. On notera des phénomènes tels que des réactions photochimiques (dégradation ou transformation des composés chimiques), la biodégradation/biotransformation ou des phénomènes d'oxydoréduction / complexation / précipitation.

Cette méthode d'échantillonnages est régie par de nombreuses normes AFNOR (Tableau 3).

Il existe ainsi de nombreuses normes entourant cette méthode car il s'agit de la méthode d'échantillonnage préconisée par la DCE dans le processus de surveillance de l'eau de surface et souterraine. On notera que pour les pesticides, la DCE préconise de réaliser :

- 12 prélèvements par an pour les pesticides classés comme substances dangereuses et/ou prioritaires,
- 4 prélèvements par an pour les pesticides classés comme substances pertinentes.

Les données de concentration en produits phytosanitaires présentées dans le chapitre précédent ont été obtenues grâce à un échantillonnage de type actif. Cependant, comme l'ont montré Criquet, J. & al. (2017), cette méthode d'échantillonnage convient à l'étude des variations en produits phytosanitaires sur le court terme. Au regard de leurs résultats, il ressort que la prise d'échantillons d'eau à raison d'une prise par mois peut engendrer la prise en compte d'évènements extrêmes ne représentant pas la concentration moyenne dans les milieux. Leurs travaux ont montré que l'utilisation d'échantillonneurs passifs de type POCIS (Polar Organic Chemical Device) fournit des résultats plus robustes lorsque l'on cherche à étudier les concentrations moyennes en produits phytosanitaires.

NF FN ISO 5667-1 de 2007-02-01	Qualité de l'eau - Échantillonnage - Partie 1 : lignes directrices pour la				
	conception des programmes et des techniques d'échantillonnage				
NF EN ISO 5667-3 de 2004-06-	Qualité de l'eau - Échantillonnage - Partie 3 : lignes directrices pour la				
01	conservation et la manipulation des échantillons d'eau				
150 5657 (12087 do 1087 or 15	Qualité de l'eau. Échantillonnage. Partie 4 : guide pour l'échantillonnage				
130 500/-4.198/ 08 198/-04-15	des eaux des lacs naturels et des lacs artificiels				
	Qualité de l'eau - Échantillonnage - Partie 5 : lignes directrices pour				
ISO 5667-5:2006 de 2006-04-15	l'échantillonnage de l'eau potable des usines de traitement et du réseau				
	de distribution				
	Qualité de l'eau - Échantillonnage - Partie 6 : lignes directrices pour				
130 5007-0:2005 de 2005-07-15	l'échantillonnage des rivières et des cours d'eau				
150 5667 0:1002 do 1002 10 15	Qualité de l'eau. Échantillonnage. Partie E9 : guide pour				
130 5007-9.1992 de 1992-10-15	l'échantillonnage des eaux marines				
150 -667 40 4000 40 4000 44 45	Qualité de l'eau. Échantillonnage. Partie 10 : guide pour				
150 5667-10:1992 de 1992-11-15	l'échantillonnage des eaux résiduaires				
150 -56 - 11 1000 do 1000 00 15	Qualité de l'eau. Échantillonnage. Partie 11 : guide général pour				
150 566/-11:1993 de 1993-03-15	l'échantillonnage des eaux souterraines				
	Qualité de l'eau - Échantillonnage - Partie 18 : lignes directrices pour				
150 500/-10:2001 de 2001-04-15	l'échantillonnage des eaux souterraines sur des sites contaminés				

ISO 5667-20:2008 de 2008-03-	Qualité de l'eau - Échantillonnage - Partie 20 : lignes directrices relatives					
	à l'utilisation des données d'échantillonnage pour la prise de décision -					
15	Conformité avec les limites et systèmes de classification					

Tableau 3 : Normes AFNOR régissant l'échantillonnage dans le cadre du suivi de la qualité des masses d'eau (Ibrahim, I., 2013).

ECHANTILLONNAGE PASSIF

C'est en 1987 que les premiers échantillonneurs passifs furent utilisés dans le but d'analyser les molécules organiques hydrophobes contenues dans l'eau (Soedergren, A., 1987). Les échantillonneurs passifs étaient auparavant utilisés pour l'analyse de la qualité de l'air (Kot-Wasik, A. & *al.*, 2007).

Les échantillonneurs passifs fonctionnent tous sur le principe de l'intégration des molécules organiques et inorganiques sur une période définie. La captation des molécules se fait par un processus de diffusion au travers d'une membrane vers une phase réceptrice. Pour chaque dispositif, le temps de déploiement est régi par leur taux d'accumulation. On définit le temps de déploiement maximum comme le temps limite $(T_{1/2})$ entre le régime intégratif et le régime curvilinéaire de la courbe d'accumulation (Figure 37)



Figure 37 : Courbe d'accumulation des molécules en fonction du temps de déploiement d'un échantillonneur passif (Vrana, B. & *al.*, 2006)

Il existe plusieurs échantillonneurs passifs dont les plus utilisés de nos jours sont :

- Les DGT (Diffusion Gradient in Thin-film) permettant l'analyse des métaux (Zang, H. & al., 1995)
- Les SPMD (Semi-Permeable Membrane Device) pour les molécules organiques apolaires (Huckins, J.N. & al., 1993).
- Les ceramic dosimeter pour les hydrocarbures aromatiques polycycliques (HAP et les composés organiques volatiles (COV) du type BTEX (Bopp, S., 2004)
- Les MESCO (Membrane Enclosed Sorptive Coating) pour les composés organiques hydrophobes (Paschke, A. & al., 2007).
- Les POCIS pour les molécules organiques polaires (Alvarez, D.A. & al., 2004).

Notre étude se concentrant sur les produits phytosanitaires polaires, le détail du fonctionnement des DGT et des SPMD ne sera pas abordé.

Breveté en 2002 sous le numéro U.S. patent 6478961, le POCIS a initialement été développé pour compléter l'analyse des eaux effectuée avec les SPMD afin d'apporter la capacité d'analyser les substances polaires (Alvarez, D.A. & al., 2004).

Ce dispositif est composé de deux anneaux métalliques maintenant deux membranes microporeuses (porosité de o.1µm) en polyéthersulfone (PES). On retrouve entre ces membranes une phase absorbante (Figure 38). La phase peut être de différents types suivant les molécules que l'on veut analyser. On notera ainsi la

possibilité d'utiliser une phase de type PHARMA® pour l'étude des produits pharmaceutiques ou bien les phases de type Pesticide® pour l'analyse des pesticides (Ibrahim, A., 2013).



Figure 38 : Schémas éclaté d'un POCIS (Source : AFFINISEP)

Les membranes en PES servent ainsi à maintenir la phase en place et également à filtrer les composés. Ainsi, les particules, macromolécules et microorganismes de plus de 0,1µm seront bloqués par la membrane. Il faut cependant noter que la diffusion des molécules au travers de la membrane peut être ralentie voire stoppée dans le cas d'un dépôt de sédiment à la surface de la membrane ou de la formation d'un biofilm. Il est donc recommandé de ne pas placer les POCIS dans une eau stagnante pour éviter ces phénomènes. Ainsi, en se basant sur les phénomènes précédemment cités et sur la courbe d'accumulation des molécules en fonction du temps de déploiement du POCIS, Ibrahim, A. (2013) en conclut que le temps optimal de déploiement des POCIS est de l'ordre de 20 jours.

3.6.2 PREPARATION ET EXTRACTION DES ECHANTILLONS

PREPARATION DES ECHANTILLONS D'EAU

La concentration en produits phytosanitaires dans l'eau étant généralement faible, il est nécessaire de concentrer les échantillons d'eau prélevés. Avant toute extraction, on injecte une concentration connue d'un traceur pour contrôler le rendement du processus d'extraction. Par la suite, l'analyse consiste en l'extraction des composés chimiques de l'échantillon. On utilise ainsi une méthode d'extraction liquide-solide. Pour cette extraction, le conditionnement d'une phase solide est effectué en fonction des propriétés des molécules que l'on cherche à analyser. Une fois conditionné, la totalité de l'échantillon est percolé au travers de la phase. Une fois la totalité de l'échantillon percolé, la phase est éluée avec un solvant (une fois encore sélectionné suivant les caractéristiques des molécules visant à être analysées). L'éluât ainsi obtenu est concentré pour permettre l'obtention d'une concentration en molécules assez importante pour dépasser les limites de détection des analyseurs.

PREPARATION ET EXTRACTION DES POCIS

Pour les POCIS, la procédure diffère des échantillons d'eau par le fait que les molécules sont déjà contenues dans la phase solide. La procédure de préparation consiste en la récupération de la phase contenue entre les deux membranes de PES. L'injection du traceur se fait alors avant de passer à l'extraction qui s'effectue selon la même procédure et avec le même solvant utilisé que pour les échantillons d'eau.

3.6.3 ANALYSE DES ECHANTILLONS

Les analyses des échantillons ont été effectuées au laboratoire LGEI (Laboratoire de Génie de l'Environnement Industriel) de l'IMT Mines d'Alès. Ce laboratoire dispose d'un chromatographe liquide de haute performance (Alliance HPLC, Waters2695) et d'un spectromètre de masse triple quadripôle (Micromass Quatro micro, Waters). Avant toute analyse, un traceur de rendement de l'analyse est incorporé dans chaque échantillon. Par la suite, l'analyse par HPLC-MS/MS suit deux grandes étapes en série. L'échantillon va dans un premier temps passer dans la colonne du chromatographe liquide qui sera responsable de la séparation des molécules selon la masse moléculaire. Une fois sorties de la colonne, les molécules seront dirigées vers le

spectromètre de masse qui va effectuer l'ionisation des molécules et ainsi classer ces dernières selon leur propriété. Cette étape d'analyse consiste en la vaporisation du liquide une fois sorti de la colonne sous forme de microgouttelettes. Ceci favorise l'évaporation du liquide qui libérera des molécules ionisées qui seront ensuite analysées par le spectromètre de masse. On aura ensuite une étape d'atomisation ou de nébulisation au moyen d'un gaz inerte (azote) afin d'améliorer la qualité des pulvérisations dans le cas d'un fort débit. Pour finir, un logiciel d'interprétation va permettre d'analyser les résultats du spectromètre et ainsi obtenir des concentrations pour chacune des molécules. Ainsi, en connaissant les temps de transferts des molécules d'interprétation de rattacher chaque pic de signal enregistré par le spectromètre de masse. En effectuant l'intégrale de chaque pic et en le comparant au pic relatif au traceur, il est alors possible de trouver la concentration du composé dans l'échantillon analysé.



Figure 39 : Equipement analytique du LGEI. Chromatographe liquide (à gauche) et spectromètre (à droite).

3.7 CONCLUSION DU CHAPITRE 4

Ce chapitre nous a donc permis de présenter les méthodes et outils que nous avons utilisés au cours de nos travaux. Ainsi les outils d'analyse des données telles que les analyses corrélatoires seront la base que nous utiliserons dans nos travaux décrits dans le chapitre 4 visant à extraire des séries chronologiques à notre disposition, les informations nécessaires pour la phase de modélisation.

Par la suite, deux outils qui nous serviront pour les travaux de prédiction de la turbidité ont été présentés ; (i) Les analyses multirésolution (ii) la décomposition en ondelette.

Les différentes méthodes citées précédemment constituent des outils puissants pour, d'une part, analyser les séries chronologiques afin de mieux connaitre les processus et leurs variables explicatives et d'autre part, simuler ces processus. Ces méthodes seront appliquées sur notre site d'étude afin d'effectuer la prévision de la turbidité au captage d'Yport. Chapitre 4 : Analyse des données

CHAPITRE 4 : ANALYSE DES DONNEES

Chapitre 4 : Analyse des données

4.1 INTRODUCTION

Les travaux présentés dans ce chapitre portent sur l'analyse préliminaire des données en vue de mieux comprendre le système étudié pour mieux concevoir les modèles. Tout d'abord, des analyses qualitatives et visuelles sur les séries de données ont été effectuées. La première partie présente les corrélations simples entre les signaux relatifs aux différentes variables (précipitation et turbidité).

La deuxième partie consiste en des analyses de corrélations. Ces dernières auront pour but de représenter les relations linéaires existantes entre les différentes variables. Ce travail aura pour double objectif d'identifier certaines caractéristiques du système (temps de réponse) et de sélectionner les données les plus propices à notre modèle mais également leurs caractéristiques (pas de temps d'échantillonnage, regroupement de données). Pour finir, cette partie se concentrera sur la validation/invalidation de l'utilité d'un modèle événementiel par rapport à un modèle en continu.

4.2 COMPARAISONS SIMPLES

4.2.1 ANALYSE DE LA SPATIALISATION DES PRECIPITATIONS

Comme il a été présenté dans le chapitre 2, la région d'Yport est caractérisée par un gradient de précipitation croissant du nord-nord/ouest vers le sud-sud/est. Ce gradient est notable sur les précipitations annuelles comme il est visible sur la Figure 40. Si l'on prend les stations se trouvant le long de la côte, on peut observer sur la Figure 40 (rectangles verts) un gradient décroissant de précipitations lorsque l'on longe la côte depuis Froberville vers le Havre. On note l'exception des stations d'Anglesqueville et Mannevillette. Ces exceptions s'expliquent par la localisation plus continentale de la station d'Anglesqueville et de la présence d'évènements pluviométriques extrêmes pour la station de Mannevillette. Nous verrons cependant dans la suite que la station d'Anglesqueville et Manevillette présente un coefficient de corrélation non négligeable avec la turbidité au niveau du forage d'Yport. Ces deux stations ne seront alors pas mises de côté pour la suite de l'étude. Concernant le gradient de précipitations continentales, on peut voir sur la Figure 40 (rectangle jaune) que les précipitations cumulées présentent une augmentation lorsque l'on s'éloigne de la côte (Annouville, Goderville et Etainhus) puis une diminution lorsque l'on se rapproche de la Seine (St Romain et St Nicolas). Si l'on se réfère à la Figure 4 du chapitre 2, on identifie bien sur la Figure 40 les gradients de précipitation présentés dans le chapitre 2.



Figure 40 : Précipitations cumulées pour chaque station de mesure sur la période allant de 2009 à 2015.

On notera sur la Figure 40 un écart entre les cumuls des stations de Gonneville et d'Anglesqueville, qui, bien que proches en distance, enregistrent des précipitations cumulées sensiblement différentes (143mm de différence). Même si cette différence peut être expliquée par l'erreur de mesure (15%), au regard des précipitations cumulées sur les autres stations, ainsi que sur le gradient de précipitation identifié sur la zone d'étude, il apparait que la station de Gonneville présente des précipitations trop faibles. Il est apparu dans nos travaux ultérieurs que les mesures enregistrées à la station de Gonneville sont en partie identiques à celles du Havre (Figure 41) :

- Entre juillet 2009 et janvier 2014, la différence est nulle, les chroniques pour ces deux stations sont identiques.
- Janvier 2014 à mars 2014 : pas de mesures pour la station du Havre, la différence ne peut être calculée,
- Mars 2014 à avril 2015 : les chroniques de pluies pour les stations ne sont plus identiques.

Ceci explique donc cette observation que nous avons estimée "anormale" et nous conduit à supprimer la station de Gonneville de notre étude.





Pour ne pas multiplier les graphiques et parce que nous pensons que le premier niveau d'analyse ne dépend pas de la distribution spatiale des précipitations, nous avons effectué le reste des analyses sur la pluie moyenne du bassin.

En effectuant la même analyse sur une période plus courte (1 mois) on voit que les variations présentées sur la Figure 42 sont similaires aux variations identifiées sur une période de plusieurs années. On observe pour cette période que la station de Mannevillette présente un cumul inférieur aux autres stations et transcrit donc bien la diminution du gradient de précipitations identifié le long de la côte en direction de l'embouchure de la Seine. On notera cependant, une légère ré-augmentation des cumuls pour les stations d'Octeville et du Havre. Cette ré-augmentation peut être attribuée à l'influence de précipitations provenant de la vallée de la seine n'ayant pas affecté les pluviomètres plus proches de notre zone d'étude. Pour les pluviomètres continentaux, on observe l'augmentation précédemment identifiée des cumuls de précipitations. En revanche, la diminution du cumul de précipitation lorsque l'on se dirige vers la vallée de la Seine n'est pas observable sur la Figure 42. En effet, on observe une augmentation du cumul de précipitations pour la station de St Romain, ce qui confirme notre hypothèse de l'influence d'un évènement pluvieux localisé dans la vallée de la Seine. On notera un très faible cumul pour la station de St Nicolas, qui s'explique par une défaillance dans la mesure de la pluviométrie. Ce

pluviomètre comme nous le verrons par la suite sera exclu de nos travaux de par son défaut de fiabilité et les nombreuses lacunes dans les données.



Figure 42 : Précipitations cumulées pour chaque station de mesures pour l'évènement allant du 01/12/2011 au 31/12/2011.

On peut donc voir au travers de l'analyse spatiale des précipitations que la variabilité spatiale des précipitations n'est pas grandement significative par rapport à la variabilité spatiale pouvant être observée dans d'autres régions telles que le sud de la France (plus de 100 mm de différence en une heure sur une distance de 1km). Cependant, la non linéarité des systèmes tels que les aquifères pourrait amplifier l'impact de cette variation spatiale. L'étude de l'impact de la spatialisation dans le processus de modélisation sera donc étudiée.

4.2.2 RELATIONS ENTRE LA TURBIDITE ET LES PRECIPITATIONS RE-ECHANTILLONNEE

Afin d'étudier les variations comparées de la turbidité et des précipitations, nous les avons tracées sur un même graphique, afin d'avoir un premier niveau d'analyse visuelle. On peut ainsi noter sur la Figure 43, que les précipitations sont beaucoup moins variables sur la période considérée (2009-2015) que la turbidité. En effet, les précipitations sont présentes durant quasi toute l'année tandis que la turbidité apparait sous forme d'évènements pour les périodes allant de novembre à mars.



Figure 43 : Comparaison entre la turbidité à l'entrée de la station d'Yport (sonde 1) et les précipitations horaires moyennes.

Comme pour toute modélisation il est important de manipuler les signaux aux bonnes échelles temporelles. Pour cela, nous avons procédé à un ré-échantillonnage des précipitations en étudiant l'échantillonnage des précipitations aux pas de temps 10 jours et 1 mois. Lorsque les pluies sont ré-échantillonnées, il est important de conserver le cumul. Ceci se fait en affectant au nouveau pas de temps considéré le cumul des précipitations tombées aux pas de temps précédents avant échantillonnage. Par exemple si l'on note, à un pas de temps donné, *P*10 et *P*1 les précipitations aux pas de temps 10h et 1h respectivement, et k_{10} et k_1 les pas de temps discrets variant toutes les 10h pour k_{10} et toutes les heures pour k_1 . Si k_{10} et k_1 ont leur période temporelle qui se termine au même instant (par exemple midi) ; on calcule :

$$P_{10}(k_{10}) = \sum_{i=0}^{9} P_1 (k_1 - i).$$

(16)

Ce calcul a l'avantage d'être causal, c'est-à-dire qu'il n'utilise pour le calcul présent (temps *k*) que les valeurs passées et non pas les valeurs futures. Ceci est une condition nécessaire lorsque l'on travaille en prévision. Un inconvénient de ce calcul est qu'il introduit un déphasage dans les données. En effet si l'on échantillonne à 10h, l'information du début de la période d'échantillonnage n'est prise en compte qu'au bout de 10h. Il convient donc de prendre le plus petit pas de temps possible, afin de réduire ce déphasage, pourvu que celui-ci ait un sens au niveau de la modélisation réalisée.

Dans la Figure 44, nous traçons les précipitations ré-échantillonnées selon un pas de temps de 10 jours, la largeur de chaque barre représentant la pluie est donc plus importante que sur le graphique de la Figure 43.

En comparant visuellement la Figure 43 et la Figure 44, on note que la relation entre les précipitations et la turbidité est plus apparente lorsque les précipitations sont cumulées sur 10 jours.

On peut alors émettre l'hypothèse que la turbidité est plus en relation avec des évènements pluvieux de plusieurs jours.



Figure 44 : Comparaison entre la turbidité au forage d'Yport et les précipitations moyennes cumulées sur 10 jours.

Si l'on renouvelle l'exercice avec les précipitations moyennes du bassin, représentées en cumulé temporel sur un mois, visualisé sur la Figure 45. Le même phénomène est apparent.



Figure 45 : Comparaison entre la turbidité au forage d'Yport et les précipitations moyennes cumulées sur 1 mois.

Cependant on peut noter sur la Figure 45 que le lien entre le cumul des précipitations moyenne sur le bassin, cumulées temporellement sur un mois et les épisodes de turbidité n'est pas systématique, comme pour le mois de juillet 2011 ou de juillet 2014. Durant ces deux périodes, on observe des cumuls de pluies notables, mais beaucoup moins de pics de turbidité que sur les autres périodes pluvieuses. Ceci est d'ailleurs généralisable à tous les étés de la base de données : la turbidité y est systématiquement plus faible que durant le reste de l'année, qu'il pleuve peu ou plus nettement. Ceci est vraisemblablement dû à l'effet de l'évapotranspiration qui limite l'impact des précipitations tant sur l'humidité du sol que sur les infiltrations souterraines. Il apparait ainsi clairement que l'humidité du bassin joue un rôle important dans la génération d'un épisode turbide. C'est donc bien l'état hydrique du bassin, stimulé par l'intensité des précipitations qui déclenche l'occurrence d'évènements turbides.

4.2.3 RELATIONS ENTRE LA TURBIDITE ET LA PIEZOMETRIE

Comme nous l'avons présenté au chapitre 1, la turbidité est liée à la vitesse de l'eau qui mobilise les particules, soit en surface par ruissellement, soit dans le milieu souterrain, en régime noyé ou vadose. Si l'on considère qu'il y a une relation entre la vitesse de l'eau et le débit, ce dernier étant lié à la hauteur de l'eau dans le conduit, on peut aussi rechercher des relations entre la turbidité et la hauteur d'eau dans le drain souterrain. Pour cela, nous avons reporté en Figure 46 les évolutions temporelles de la turbidité et de la hauteur d'eau depuis juillet 2013 jusqu'à janvier 2015 dans le forage d'Yport.



Figure 46 : Évolution de la hauteur d'eau dans le forage d'Yport et de la turbidité à l'entrée de l'usine (sonde 1) durant la période allant du 3 juillet 2013 au 5 février 2015.

On peut tout d'abord noter de nombreuses oscillations sur la Figure 46. Ces dernières correspondent à la marée et peuvent facilement être supprimées par un filtrage de type « moyenne mobile » avec une période de 24 h. Avec cette méthode, les coefficients de marée qui varient sur plusieurs jours ne sont pas supprimés, mais la hauteur d'eau ainsi filtrée est visible grâce à la courbe verte.





Il apparait sur la Figure 47 que l'augmentation de la hauteur d'eau précède l'arrivée d'eau trouble au droit du forage. Cela correspond à un effet de "chasse" : l'eau déjà présente dans les conduits souterrains et chassée par les arrivées d'eau trouble de surface. Cet effet est généralement identifié en mesurant la température ou la conductivité de l'eau. Nous l'observons ici sur la turbidité lorsque cette dernière est causée par des apports en provenance de la surface.



Figure 48 : Évolution de la hauteur d'eau dans la nappe et de la turbidité mesurée à l'entrée de l'usine d'Yport (sonde 1).

Si l'on observe l'évolution de la piézométrie sur une plus longue période, telle qu'elle est représentée en Figure 48, on voit également que sa valeur sur le long terme n'est pas liée de manière simple à l'intensité des événements turbides. En revanche, ce serait plutôt les variations de la piézométrie journalière de la nappe mesurée à Auberville-la-Renault qui influencerait la génération des événements turbides. Cependant, le lien observé entre les deux variables n'est pas un lien simple. En effet, ce sont les précipitations qui influencent ces deux variables et qui créent un lien entre elles. Ceci est particulièrement visible pour la période de fin 2012 début 2013, durant laquelle un grand nombre de précipitations ont été mesurées (Figure 45). Ces précipitations ont alors été d'une part, responsables d'un rechargement de la nappe et d'autre part, d'un apport conséquent de matières en suspension. On notera par la suite un maintien du niveau piézométrique durant la période fin 2013- début 2014 qui est dû à nouveau à une période de fortes précipitations mensuelles. Le plus faible nombre d'évènements de turbidité peut s'expliquer par le lessivage important des sols s'étant produit en début 2013.

4.2.4 RELATIONS ENTRE LA TURBIDITE ET LA CONDUCTIVITE ELECTRIQUE

La conductivité électrique est plus importante pour une eau saturée en minéraux, elle est donc généralement associée à une eau ayant séjourné longtemps au sein du milieu souterrain. On pourra donc s'intéresser à la relation entre la turbidité et la conductivité de l'eau pour analyser si l'eau qui apporte la turbidité est une eau peu minéralisée (pouvant correspondre à une eau de surface), ou très minéralisée (pouvant correspondre à une eau de surface), ou très minéralisée (pouvant correspondre à une eau souterraine). Concernant la relation entre la turbidité et la conductivité électrique, il ne nous est pas possible de l'étudier sur une grande période pour le site d'Yport. En effet, la conductivité n'est mesurée et archivée dans le drain d'Yport que depuis janvier 2013, et cette chronique présente de nombreuses lacunes pour les périodes antérieures à janvier 2015. La période la plus complète comportant un épisode de turbidité est représentée sur la Figure 49.



Figure 49 : Évolutions de la conductivité dans le forage en fonction de la turbidité à l'entrée de l'usine (sonde 1).

On peut noter sur la Figure 49 que lorsque la turbidité augmente, la conductivité électrique diminue, signifiant que l'on a un apport d'eau de surface. Si la turbidité varie avant la diminution de la conductivité, nous avons une turbidité provenant de la remise en suspension des sédiments présents dans le karst. Cette remise en suspension est provoquée par l'effet de chasse provoqué par l'augmentation de la charge hydraulique à l'entrée du système. Ainsi, la conductivité peut ne pas varier alors que la turbidité augmente.

4.2.5 ANALYSE DES DONNEES CONCERNANT LES PRODUITS PHYTOSANITAIRES

Comme nous l'avons indiqué dans les objectifs de cette thèse, nous allons aussi nous intéresser au lien entre la turbidité et les produits phytosanitaires : la turbidité pouvant servir de transport à ces produits polluants. Pour cela, nous allons nous intéresser aux données concernant les produits phytosanitaires au travers de l'interprétation des mesures en atrazine. Nous avons choisi l'atrazine car c'est pour cet élément que les données récoltées apparaissent comme les plus intéressantes. En effet, les mesures en atrazine sont les mesures les plus fréquemment effectuées de par l'interdiction d'utilisation de ce produit au travers de l'avis aux opérateurs du 27 novembre 2001 (Miquel, M. G., 2003). De plus, c'est pour l'atrazine que le nombre de mesures supérieures au seuil de détection est le plus important. Ainsi, nous visualisons sur la Figure 50 la superposition des mesures de la turbidité et des mesures en atrazine durant la période allant de juillet 2009 à janvier 2015.



Figure 50 : Comparaison entre la concentration en atrazine et la turbidité à l'entrée de la station (sonde 1).

Au premier regard, on peut distinguer deux périodes différentes :

- La période entre le 1^{er} juillet 2009 et le 1^{er} janvier 2012 présentant un seuil de 0,02 μg/L pour lequel les mesures inférieures sont référencées dans la base de données comme "*inférieures au seuil de quantification*".
- La période allant du 1^{er} janvier 2013 au 1^{er} janvier 2015 présentant un seuil de 0,01 μg/L pour lequel les mesures inférieures sont référencées dans la base de données comme "*inférieures au seuil de quantification*", mais pour lequel les valeurs supérieures sont indiquées comme "*valides*".

Les différences durant ces deux périodes peuvent s'expliquer par une évolution du matériel de mesure dont le seuil de quantification serait passé de 0,02 μ g/L à 0,01 μ g/L.

De plus, sur la Figure 51, nous avons reporté les mesures de produits phytosanitaires et de turbidité afin d'évaluer leur éventuelle synchronisation.

Cependant il est clair qu'avec un pas de temps mensuel les mesures de concentration en produits phytosanitaires peuvent ne pas être effectuées au moment du passage des pollutions ou même du passage des pics de turbidité.

Cependant, les analyses servant à mesurer les concentrations en produits phytosanitaires sont très coûteuses et il n'est pas possible d'en faire aussi fréquemment que la mesure de la turbidité. Il serait donc intéressant d'envisager une nouvelle méthode d'analyse telle que les échantillonneurs passifs (exemple : Les Polar Organic Compounds Integrative Samplers, ou POCIS) permettant de mesurer l'intégration des concentrations en produits phytosanitaires.



Figure 51 : Effet du ré-échantillonnage de la turbidité sur le modèle de l'échantillonnage des produits phytosanitaires.

On note sur la Figure 51 que des concentrations en produits phytosanitaires sont observées lorsqu'il n'y a pas d'évènement turbide. Cette constatation laisse préjuger de difficultés pour identifier une relation entre la turbidité et la présence de produits phytosanitaires. Nous verrons donc dans le chapitre 7 les résultats d'une méthode alternative pour l'analyse des produits phytosanitaires dans le cas du forage d'Yport.

4.3 ANALYSES CORRELATOIRES

4.3.1 CORRELATION CROISEES ENTRE LES PRECIPITATIONS ET LA TURBIDITE AU PAS DE TEMPS HORAIRE

Les corrélations croisées entre, d'une part les différentes mesures des précipitations (au pas de temps horaire) à différents postes pluviométriques, et d'autre part les différentes mesures de précipitations et la turbidité mesurée à Yport, ont été calculées. Les corrélations simples et corrélations croisées sont calculées par année civile, puis la moyenne de ces corrélogrammes est réalisée sur toutes les années afin d'obtenir un corrélogramme moyen. Dans la suite nous omettrons le terme « moyen » pour désigner les corrélogrammes simples ou croisés afin de simplifier l'expression, mais ce calcul de moyenne sera toujours effectué.

ANALYSE DES MAXIMAS DE CORRELATION

A partir du corrélogramme, on peut extraire plusieurs informations simplifiées. Parmi celles-ci la valeur du maximum du corrélogramme permet d'estimer la simplicité, ou le caractère direct, de la relation liant les deux variables étudiées. Dans nos calculs concernant les corrélations entre pluviomètres au pas de temps horaire, ce maximum des corrélogrammes dépasse rarement 0,5, ce qui est assez faible (Tableau 4). Ceci signifie que la pluie au pas de temps horaire est assez variable spatialement. La valeur de 0,88 entre les stations du Havre et de Gonneville a attiré notre attention, c'est grâce à cela que nous avons pu remarquer que les 2 chroniques sont identiques sur quasiment toute la période d'observation (indiqué en section 4.2.1).Ce maximum de 0,88 est donc artificiellement élevé. Nous le portons cependant dans le (Tableau 4).

	Turb.	Pluviométrie										
	Yport	Frob.	Annou.	Goder.	Angles.	Gonne.	Manne.	Octe.	Etain.	St Rom.	St Nic.	Le Hav.
Turb. Yport	1											
Frob.	0,19	1										
Annou.	0,20	0,50	1									
Goder.	0,25	0,49	0,59	1								
Angles.	0,19	0,48	0,48	0,60	1							
Gonne.	0,15	0,41	0,42	0,43	0,46	1						
Manne.	0,21	0,38	0,38	0,42	0,48	0,41	1					
Octe.	0,14	0,40	0,38	0,41	0,48	0,49	0,48	1				
Etain.	0,19	0,42	0,46	0,52	0,53	0,50	0,45	0,50	1			
St Rom.	0,17	0,37	0,42	0,46	0,42	0,46	0,38	0,43	0,61	1		
St Nic.	0,11	0,30	0,37	0,38	0,31	0,34	0,28	0,35	0,42	0,47	1	
Le Hav.	0,14	0,39	0,38	0,39	0,42	0,88	0,38	0,48	0,49	0,47	0,36	1

Tableau 4 : Récapitulatif des maxima des corrélations croisées interpluviomètres, au pas de temps horaire, sur la période allant de juillet 2009 à janvier 2015.

Concernant l'analyse croisée pluie-turbidité, nous pouvons déterminer 2 catégories de résultats :

- La première catégorie (repérée en rouge dans le Tableau 4) concerne les stations de Froberville, Anglesqueville, Gonneville, Etainhus, St Romain, St Nicolas et le Havre pour lesquelles le maximum de corrélation croisée est inférieur à 0,2. On peut donc considérer que ces stations n'apportent pas d'informations significatives pour expliquer la turbidité à Yport. Pour la station de Froberville (0,19), ceci peut s'expliquer par sa proximité avec le forage. Ainsi, les pluies localisées à Froberville vont influencer la partie aval du conduit concerné par le captage et donc ne vont pas influencer la turbidité au niveau du forage. Pour la station d'Ethainus (0,19), cela peut s'expliquer par sa localisation au niveau de la zone de partage entre le bassin d'alimentation d'Yport et de Radicatel.
- La seconde catégorie (repérée en orange dans le Tableau 4), concerne les stations d'Annouville, Goderville, Mannevillette pour lesquelles les maxima des corrélations croisées sont supérieurs à 0,2 mais restent très faibles. Parmi ces stations, seule la station de Goderville est considérée comme appartenant au bassin d'alimentation du captage.

Le fait que nous ne puissions pas établir de relations claires entre l'appartenance au bassin d'alimentation et l'importance de la corrélation croisée pose des questions sur ; (i) la qualité des mesures, (ii) le choix de la période d'échantillonnage (nous avons montré au début que les précipitations apparaissaient comme mieux corrélées avec la turbidité au pas de temps décadaire), (iii) la complexité de la relation étudiée. Il est important de ne pas perdre de vue que la corrélation n'est pas la causalité ; ainsi les pluies des stations d'Annouville et de Mannevilette peuvent apparaitre comme corrélées à la turbidité à Yport, cependant cette corrélation peut être issue de leur propre corrélation avec les pluies de la station de Goderville, comprise dans le bassin d'alimentation du captage (maxima des corrélations croisées inter-pluviomètres de 0,59 et 0,42 respectivement).

L'analyse de ces données montre donc clairement que le travail de modélisation (ou prévision) ne sera pas facile à conduire.

ANALYSE DES TEMPS DE REPONSE AU PAS DE TEMPS HORAIRE

Nous avons reporté dans le Tableau 5 les temps de réponse relevés sur les corrélations croisées, obtenus comme décrit au Chapitre 3. On peut noter que les temps de réponses sont de l'ordre de 36 heures, excepté pour les stations de Manevillette, Octeville et Saint Nicolas qui présentent un temps de réponse de 30 heures.

Nous avons montré précédemment (section 4.2.2) que c'était le cumul de précipitations sur 10 jours qui paraissait visuellement le plus en relation avec les épisodes de turbidité. Au vu des temps de réponse, de l'ordre de grandeur (la trentaine d'heure) et de la relation pluie-turbidité, il semble que l'on peut supposer que les cumuls de précipitations continues sur 10 jours contribuent à un état favorable au déclenchement de l'épisode turbide, mais que cet épisode a une dynamique propre, plus rapide. On a donc bien deux phénomènes qui

interviennent à des échelles de temps différentes, la première étant liée à la notion d'état d'humidité du bassin, ce qui est une marque de non linéarité du processus. On notera le temps de réponse de 37h pour la station de Froberville. Comme indiqué précédemment, la station de Froberville est située à proximité du forage d'Yport. Il est donc peu plausible que cette valeur corresponde au temps de réponse de la turbidité dans le drain karstique lorsqu'il pleut à Froberville. Une fois de plus ce temps de réponse est vraisemblablement dû aux corrélations entre les différents pluviomètres entre eux. Pour obtenir le temps de réponse entre les précipitations et la turbidité mesurée à Froberville, il serait nécessaire d'identifier un évènement pluvieux localisé à Froberville. Malheureusement, un tel évènement n'est pas présent dans la base de données.

	Turbidité Yport
Froberville	37
Annouville	35
Goderville	37
Anglesqueville	37
Gonneville	40
Mannevillette	30
Octeville	30
Etainhus	37
St Romain	38
St Nicolas	30
l e Havre	40

Tableau 5 : Récapitulatif des temps de réponses sur la période allant de juillet 2009 à janvier 2015. A noter que les chroniques de pluie de Gonneville et du Havre ont une majeure partie commune.

CHOIX DES STATIONS AU PAS DE TEMPS 1H

En synthèse il apparait que l'analyse des corrélations croisées au pas de temps horaire n'apporte pas d'éléments très tranchés permettant de choisir sereinement les stations pluviométriques puisque les chroniques enregistrées sur ces dernières ont quasiment toutes des maxima des corrélations croisées pluie-turbidité très faibles, de 0,11 à 0,25 (Tableau 4). Nous avons également mis en évidence la non-linéarité de la relation pluie-turbidité puisque le système possède différents fonctionnements selon son état (Figure 48), estimables grâce au cumul de précipitations sur 10 jours. On peut donc en déduire que la méthode de corrélation croisée, limitée à la recherche de coïncidences entre les variables sans traitement ne permet de décrire que les relations linéaires, et se trouve donc limitée dans cette étude. En revanche nous pouvons toujours exploiter les connaissances de terrain sur la zone d'étude présentée au Chapitre 2 et proposer de conserver, pour un travail au pas de temps horaire, les chroniques disponibles sur le bassin d'alimentation du captage pour les stations de Froberville, Goderville et d'Ethainus. Nous retiendrons que c'est la station de Goderville qui a le plus fort coefficient de corrélation croisée (0,25).

4.3.2 INFLUENCE DE LA PERIODE D'ECHANTILLONNAGE SUR LES ANALYSES CORRELATOIRES

Compte tenu des différences de comportement mises en évidence selon que l'on considère les entréessorties aux pas de temps horaire ou décadaire, il paraît intéressant d'étendre cette étude à une variation de pas de temps plus étendue : 1 jour, 2 jours, 5 jours, 10 jours, 20 jours et 1 mois. La méthode de ré-échantillonnage des précipitations correspond à la méthode classique consistant à faire le cumul des précipitations. En revanche, pour la turbidité, c'est la valeur du maximum qui sera conservée car le ré-échantillonnage utilisant le cumul ou

Chapitre 4 : Analyse des données

la moyenne sur 1 mois pourrait gommer complètement les pics de turbidité³. Pour ce ré-échantillonnage nous choisissons les stations ayant un maximum de corrélations croisées, proches ou supérieures à 0,2, afin de ne pas exclure trop de stations. Les résultats concernant les maxima des corrélations croisées observés sont présentés dans la Figure 52.

On peut noter sur la Figure 52 que les corrélations maximales sont fournies par les données rééchantillonnées avec un pas de temps de 5 jours ou 10 jours. Ceci confirme donc l'hypothèse émise précédemment selon laquelle, un évènement pluvieux intense n'est pas le principal phénomène entrainant les transports d'eaux troubles mais que l'état du bassin joue un rôle important.

On observe cependant un cas particulier, difficilement explicable, au niveau de la station de Manevillette qui présente une augmentation moins marquée du maximum de corrélations suite à un ré-échantillonnage pour 1 jour et 2 jours et que cette dernière présente une diminution du taux de corrélation pour les pas de temps de 5 jours, 10 jours et 20 jours. Rappelons que cette station n'appartient pas au bassin d'alimentation du captage.

Nous avons ensuite cherché à voir si l'utilisation des précipitations moyennes sur le bassin versant pouvait apporter de meilleurs résultats de corrélation. Pour ce faire, une chronique présentant la moyenne des précipitations horaires relevées au niveau des stations de Froberville, Annouville, Goderville, Anglesqueville, Manevillette et Etainhus a été effectuée sur la période allant de juillet 2009 à janvier 2015. On obtient alors un maximum de corrélation de 0,27 et un temps de réponse de 37 heures. Le temps de réponse est identique aux temps de réponse que nous avons observés précédemment pour les corrélations calculées avec les mesures de pluviomètre seules. En revanche, le maximum de corrélation est supérieur au maximum que nous avions observé pour les corrélations individuelles des pluviomètres avec la turbidité (maximum de 0,25 pour la station de Goderville).





Il apparait donc que l'utilisation d'une chronique de précipitation moyenne sur le bassin semble intéressante dans la suite des travaux, et permettra de simplifier les premiers travaux d'analyses.

4.3.3 APPROCHE EVENEMENTIELLE

Du fait des nombreuses lacunes présentes dans la base de données, nous avons été contraints de spécialiser notre analyse sur les évènements turbides. Pour ce faire, des évènements ont été sélectionnés dans la chronique

En effet ce ré-échantillonnage n'est rien d'autre qu'un filtre passe bas : il élimine les hautes fréquences.

allant de juillet 2009 à janvier 2015. La sélection des évènements a été réalisée en appliquant un seuil sur les précipitations et non sur les turbidités afin d'éviter la prévision de faux positifs en ne considérant que les précipitations ayant engendré une réponse de turbidité. Le seuil de précipitation défini pour la sélection des évènements est de 3,5 mm en 24 heures en utilisant une fenêtre glissante.



Figure 53 : Représentation des 22 évènements sélectionnés.

Ainsi, c'est un total de 22 évènements qui ont pu être sélectionnés (Figure 53). Le Tableau 6 présente les caractéristiques de chaque évènement.

On peut noter sur la Figure 53 que des évènements pluviométriques sont présents durant la période allant de novembre 2012 à mars 2013. Cependant, il a été décidé de ne pas tenir compte de ces derniers car on peut voir un arrêt de la mesure de la turbidité qui correspond à une période ou le capteur n'était plus en place. De nombreuses lacunes sont également observables durant cette période.

Le Tableau 7 présente les maxima des corrélations croisées évènementielles entre les précipitations moyennes sur le bassin et la turbidité au captage d'Yport.

	22	06/12/201	31/15/2014	624	INO	53,9
	21	τος/ζο/ζε	20/08/2014	600	NON	48,4
	20	56/06/2014	לדס∠/20/6ד	576	NON	4,4
	19	זל/סל/זסזל	οל/ο2/50	504	NON	8,4
	18	£τοz/zτ/£τ	τος/το/ττ	744	INO	86,7
	17	ετος/ττ/το	£τοz/zτ/το	744	INO	157,5
	16	ετος/οτ/So	ετος/οτ/τε	648	INO	405,3
	15	ετοz/60/90	£τοz/60/0£	600	NON	5,0
	14	ετος/ζο/9τ	ετος/ζο/τε	384	NON	3,7
	13	ετος/20/80	ετος/ςο/τε	576	NON	5,0
nent	12	ΖΤΟΖ/90/ΤΟ	ztoz/90/6t	456	NON	3,6
Evéner	11	ττος/ζτ/το	ττος/ςτ/τε	744	INO	307,9
	10	ττοτ/60/Ζο	ττοτ/60/0ε	576	INO	256,1
	6	ττος/80/9τ	ττοτ/80/9τ	384	NON	6,5
	8	ττος/80/το	ττος/80/ζτ	360	NON	26,9
	7	0102/21/10	οτος/ςτ/τε	744	INO	84,7
	9	0107/11/10	οτος/ττ/οξ	720	INO	245,4
	5	οτοτ/6ο/ζτ	0707/60/08	336	NON	7.7
	4	οτοτ/6ο/το	οτοz/6ο/9τ	384	NON	9,8
	3	0107/20/01	οτοτ/εο/ετ	1008	INO	135,0
	2	6007/07/82	6002/11/22	624	INO	302,5
	1	6007/07/70	6002/01/21	288	NON	7,1
		Date de début de l'évènement	Date de fin de l'évènement	Durée de l'évènement (en heures)	Présence d'un pic de turbidité	Maximum de Turbidité (NTU)

	22	L . C	0,45		31			37/126	
	21	r c	7'0	171			38		
	20	r c	7'n	40		58			
	19	0,4 87						12	
	18	0,31 29				48/122			
	17	r C	0,47		22		31		
	16	- Y - Y	10'0		35		34		
	15	r c	7'n		40			4	
	14	r c	7'0		33			4	
ent	13	Ċ	7'0		43			10	
énem	12		5'n		41			9	
Ε<έ	11	Ċ	5'0	40			67		
	10	, 66	00 ' 0	25			35		
	6		4'0	44			31		
	8	r c	5'n	36				13	
	7	y v	0 0	29				27	
	6	о бг	ςυ,υ	23			48		
	5	r c	5'0	50			9		
	4	Ĺ	<i>د</i> יט	41			œ		
	З	0,49		111		82			
	2	o,52		46			72		
	1 0,3 (5'0	66			32		
		Maximum de	corrélation	Temps de	réponse	(en heures)	Effet	mémoire (en	heures)
		Corrélation	entre	pluviométrie	moyenne et	turbidité	Corrélation	simple de la	turbidité

Tableau 7 : Synthèse des analyses corrélatoires au pas de temps horaire pour chaque évènement.

Dans le cas où les évènements ne présentent pas de pic de turbidité, il n'est pas possible de calculer un temps de réponse. En ce qui concerne le maximum de la corrélation croisée, on peut noter que la majeure partie du temps, lorsque l'évènement ne présente pas de pic notable de turbidité, le maximum de corrélation entre la turbidité et les précipitations moyennes est très rarement supérieur à 0,4. On note cependant un maximum de corrélation supérieur à 0,4 pour les évènements 4 et 9. Ceci s'explique par le fait que ces évènements possèdent un pic de précipitation localisé dans le temps et intense (2 mm en 1 heures). Concernant les corrélations pour les évènements durant lesquels un pic de turbidité important est présent, on peut noter que pour la majeure partie des évènements les maximums sont supérieurs à 0,4, à l'exception des évènements 11 et 18. Ceci s'explique par le fait que la durée du pic de turbidité est beaucoup plus importante que pour les autres évènements et donc qu'il ne recouvre pas bien l'évènement rapide de pluie. D'un point de vue de l'effet mémoire, il est visible que les résultats sont très différents pour chacun des évènements. On peut noter que:

- la plupart des durées relatives à l'effet mémoire sont proches d'une valeur de 30 heures.
- la durée maximale de l'effet mémoire est de 82 heures pour l'évènement 3, ce qui correspond à la décroissance lente du pic de turbidité.
- la durée minimale de l'effet mémoire est de 27 heures pour l'évènement 7, ce qui correspond à un pic de turbidité relativement court.
- pour l'évènement 22, le corrélogramme passe deux fois par la valeur de 0,2 en décroissant. Le premier passage intervenant après 37 heures et le deuxième passage intervenant après 126 heures. Ceci est dû à la présence de deux pics de turbidité dans l'évènement.

Au travers de ces analyses, et surtout aux maxima des corrélations plus élevés, il est visible que l'utilisation de données sous la forme d'évènements parait plus pertinente que sous la forme d'une chronique continue. On effectue ainsi des analyses plus spécifiques centrées sur les situations susceptibles de générer des évènements turbides au sein du forage.

4.4 CONCLUSION DU CHAPITRE 4

Au travers de ce chapitre, avant toute interprétation des données, une critique de ces dernières a été effectuée. Une fois les données critiquées et sélectionnées, leur analyse nous a permis d'identifier le fonctionnement du système hydrogéologique d'Yport. Pour l'analyse des données concernant les produits phytosanitaires, la critique nous a permis de comprendre l'impact de l'échantillonnage actuellement utilisé et ainsi d'envisager une alternative.

L'analyse de la spatialisation des pluies a permis de mettre en évidence une différence peu importante entre les pluviomètres, lorsque l'on étudie cette spatialisation sur plusieurs années. En revanche, dans le cas où l'on s'intéresse à des évènements de pluies (sur 1 mois), la spatialisation des précipitations se trouve bien plus prononcée. Ainsi, ces interprétations, sur le double aspect de la spatialisation nous ont guidé sur deux pistes de réflexion; (i) l'utilisation d'une pluviométrie moyenne pour les modèles (ii) l'utilisation d'une pluie spatialisée pour les modèles.

Par la suite, l'effet d'un ré-échantillonnage des données a été investigué. Nous avons alors pu mettre en évidence grâce à ce ré-échantillonnage le lien fort existant entre humidité du bassin (longue période de précipitations) et les évènements turbides.

L'étude de la piézométrie nous a permis d'écarter une relation possible entre la variation piézométrique causée par la marée et les variations de turbidité. Pour finir, l'étude de la conductivité a mis en évidence un lien fort de cette variable avec la turbidité. Cependant, l'historique faible des données dans le cas de la conductivité ne nous permet pas d'envisager l'utilisation de cette dernière pour la modélisation.

Dans la partie s'attachant à l'analyse des données concernant les produits phytosanitaires, nous avons mis en avant les problèmes engendrés par une analyse de l'eau en utilisant des échantillons ponctuels avec une fréquence proche du bi mensuel. Il a été montré que dans le cas d'un ré-échantillonnage de la turbidité basé sur les prélèvements d'eau, les pics de turbidité sont tous absents des nouvelles données extraites. Ces observations nous confortent alors dans le besoin de proposer et de tester un échantillonnage passif des produits phytosanitaires.

Dans la suite des travaux présentés dans le chapitre 4, les corrélations simples des précipitations ou bien croisées avec la turbidité nous ont permis d'identifier les paramètres du système d'Yport. Ainsi, il est ressorti un temps de réponse du bassin de l'ordre de 36h. Par la suite, le ré-échantillonnage des données à différents pas de temps et la réalisation de corrélation entre les précipitations et la turbidité ré-échantillonnée nous ont permis de confirmer l'hypothèse précédente concernant l'influence de l'humidité du bassin sur la réponse hydrologique du système d'yport. Pour finir, l'analyse des coefficients de corrélation nous a permis d'effectuer une sélection des meilleurs pluviomètres afin de ne garder que les données impactant notre site d'étude. C'est ainsi 6 stations sur 11 qui ont été retenues pour la suite de notre étude (Froberville, Annouville, Goderville, Anglesqueville, Manevillette et Etainhus).

Pour finir, la dernière partie de ce chapitre nous a permis de mettre en évidence que l'utilisation de données sous forme d'évènements paraissait plus pertinente pour les modélisations. Cela nous permet alors d'effectuer une modélisation plus spécifique des situations susceptibles d'entrainer des événements turbides au niveau du forage. Ce sont alors 22 événements qui ont été extraits de la totalité de notre base de données et que nous utiliserons dans les prochains chapitres pour effectuer la prévision de la turbidité.

Chapitre 4 : Analyse des données

Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance

CHAPITRE 5 : PREVISION DE LA TURBIDITE PAR RESEAUX DE NEURONES 12H ET 24H A L'AVANCE

Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance

5.1 INTRODUCTION

Ce chapitre présente notre travail effectué sur la prévision de la turbidité avec les réseaux de neurones. Les résultats des modèles prédictifs à 12h ont fait l'objet d'une première publication ; Savary, M. & al. 2016. Par la suite, une extension de ces résultats à l'horizon de prévision 24h ont fait l'objet d'un article soumis dans le journal "Journal of hydrology". Ce chapitre comprend l'article rédigé pour le journal remis en forme pour s'intégrer aux autres chapitres du présent manuscrit ainsi que quelques compléments.

Pour ces travaux, deux groupes de modèles ont été conçus avec 2 horizons de prévision respectifs : 12h et 24h et deux durées d'échantillonnage : 1h et 6h. Ces horizons de prévision ont été choisis pour leur intérêt opérationnel pour les gestionnaires du captage d'alimentation en eau potable Yport. Les deux périodes d'échantillonnage sont testées afin d'évaluer si la diminution du nombre de paramètres obtenue par l'échantillonnage à 6h procure un gain de qualité aux modèles. Comme présenté dans le chapitre 3, nous utilisons les modèles de type perceptron multicouche dirigé (feed-forward dans l'article) et non-dirigé (récurrent dans l'article). Une autre architecture a également été utilisée, le modèle à deux branches (dirigé et non-dirigé). Ce modèle a montré des résultats intéressants dans la prédiction des débits (Johannet, A. & al. 2008), et nous avons jugé intéressant de tester ses capacités dans le cas de la prédiction de la turbidité. Ainsi, avec l'utilisation de deux types d'architecture de réseaux de neurones, de deux types de modèles (dirigé et non-dirigé) et des objectifs de prévision à 12h et 24h, c'est en tout 32 modèles qui ont été conçus et testés. La qualité de prédiction de chacun de ces modèles a été évaluée selon deux types d'observation; la qualité de prévision des pics de turbidité (amplitude et synchronisation) et la prédiction d'un dépassement de seuil de 100 NTU. Ces critères de qualité ont été mis en perspective de la complexité des modèles conçus. Les résultats sur la totalité des évènements turbides de notre base de données sont présentés pour chacun des objectifs. Une illustration des courbes obtenues pour chaque modèle est présentée pour l'évènement 2.

5.2 RESUME

Près de 25% de l'alimentation en eau potable mondiale provient de l'exploitation de l'eau souterraine. Néanmoins, du fait de leurs faibles propriétés de filtration, les aquifères karstiques présentent une forte vulnérabilité aux pollutions et tout particulièrement à la turbidité. Dans ce contexte, du fait de la complexité des processus impliqués et des faibles connaissances opérationnelles sur le transport des particules dans les systèmes souterrains, l'approche de modélisation la plus prometteuse semble être l'approche boîte noire qu'offrent les réseaux de neurones. En dépit de la capacité d'approximation universelle du modèle perceptron multicouche, il semble difficile de prendre en compte les conditions hydrologiques du bassin. Ces conditions hydrologiques dépendent à la fois de l'état initial du bassin (globalement humide ou sec) et l'intensité des précipitations. Dans ce but, une architecture alternative au perceptron multicouche a été proposée dans des travaux précédents afin de prendre en compte les phénomènes à grande échelle temporelle (humidité du bassin), couplée avec les variables à plus faible échelle temporelle (pluviométrie). Cette architecture appelée "perceptron multicouche à deux branches" est comparée avec le modèle classique du perceptron à deux couches, avec pour chacun des modèles une utilisation en mode dirigé et non-dirigé. Ces différents modèles étant appliqués au cas particulier du forage d'alimentation en eau potable d'Yport (Normandie, France) avec un horizon de prévision de 12h et de 24h, il apparait que les deux modèles : dirigé et non-dirigé sont capables de prédire chacun une propriété du signal de turbidité: (i) l'amplitude et la synchronisation des variations sont mieux représentées par le modèle à "deux branches" dirigé, (ii) le dépassement d'un seuil de prédiction, est mieux prédit par le modèle classique du perceptron multicouche dirigé.

Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance

Operational Turbidity Forecasting in karst aquifers: comparisons between Recurrent and Feed-Forward *based* Multilayer Perceptrons. Case study on the *Yport* basin (Normandy, France).

Michaël Savary^{1,2}, Anne Johannet², Nicolas Massei¹, Jean-Paul Dupont¹, Emmanuel Hauchard³

¹ M2C Laboratory, Rouen University, Place E. Blondel, 76821 Mont-Saint-Aignan, France

 $\{michael.savary1, nicolas.massei, jean-paul.dupont\} @univ-rouen.fr$

² LGEI, IMT Mines Alès, Univ. Montpellier, 6 avenue de Clavières, 30 319 Alès Cedex, France anne.johannet@mines-ales.fr

³ Communauté d'Agglomération Havraise, 19 Rue Georges Braque, 76600 Le Havre, France emmanuel.hauchard@codah.f

Corresponding author: Anne Johannet, <u>anne.johannet@mines-ales.fr</u>

Highlights

- Rainfall-turbidity relation can be modeled for operational applications by neural networks
- Feed-forward and recurrent neural network models are complementary
- Re-sampling signals can be useful to manage bias-variance tradeoff

Abstract

Approximately 25% of the world population drinking water depends on karst aquifers. But karst aquifers are very sensitive to pollutant transport and specifically to turbidity because they are bad filters of water. Due to the non-linearity of the processes generating turbidity, and the lack of accurate physical description of the involved phenomena, physical modelling is not frequently used to model turbidity. Alternatively, as the significant physical information is difficult to measure *in situ*, it appears interesting to investigate other approaches based on a different paradigm: statistical modelling. Statistical models can address various global relations: e.g. velocity/turbidity, which is the most frequent approach. In an original way, neural networks were chosen in this study to address directly the rainfall-turbidity relation. In this context 32 models implementing rainfall-turbidity relation were designed and their properties were assessed using original criteria. It appears that rigorous model selection allows designing models able, with limited uncertainty, to predict the threshold overtaking and maximum of turbidity peak up to 24h ahead.

1. Introduction

1.1. Turbidity and karst aquifers

Approximately 25% of the world population drinking water depends on karst aquifers. Nevertheless, due to their poor filtration properties, karst aquifers are very sensitive to pollutant transport and specifically to turbidity. Turbidity is the ability of a liquid to interrupt a light beam. It is caused by the presence of suspend particles (Hardenbergh, W. A. 1938, Bin Omar, A. F. & al. 2009). The measurement unit for turbidity is the NTU (Nephelometric Turbidity Unit). Turbidity is a crucial problem in karst aquifers that are bad filters of water. Indeed, karst aquifers are discontinuous, heterogeneous and anisotropic medium. Water is first infiltrated through the epikarst, which is the interface between the karst itself and the surface. Water circulates afterward through the nonsaturated zone and joins finally the saturated zone (if it exists). Water circulates in fissures and conduits of very different sizes from micrometric to decametric dimensions. Large drains and fissures in the karst promote a rapid transfer of unfiltered water. The filtration property of karst aquifers depends thus essentially from the filtering ability of the interface soil-epikarst. If a direct connection is possible from the surface towards the aquifer (e.g. thanks to swallow holes), the turbidity at the outlet can be generated by two kinds of processes: (i) the surface water turbidity propagated through the aquifer up to the spring (exogenous turbidity), or (ii) turbidity coming from the re-suspension of matter yet situated in the karst network (endogen turbidity). Both processes have very different hydrodynamics properties. It is possible to discriminate these processes by the observation of the water features. Indeed, Massei, N. & al. (2003) proposed using tracer tests in conjunction with flow rates measurements to identify
endogen and exogenous incomes; Shevenell, L. & *al.* (2002) proposed to associate pH, TOC (Total Organic Carbon), temperature and DO (Dissolved Oxygen) measurement; Atteia, O. & Kozel, R. (1997) show the possibility to differentiate endogen and exogenous turbidity thanks to the observation of the particles size and water pH. To identify flush effect, Kneller, B. & Buckee, C. (2000), used combination of flow rate and turbidity events.

All studies converge to state that turbid responses can be very variable, even for the same hydrosystem, depending on the active processes. For this reason, modelling and predicting turbidity are currently difficult tasks.

1.2. Turbidity: a proxy of pollution?

Regarding, the exogenous turbidity generation process, it appears that depending on the land use of the alimentation basin, the surface water can be charged in organic particles (fecal bacteria, organic matter) or chemical compounds (nitrate, phytosanitary products). Two paths can promote pollution transport: suspended matter or dissolved matter. Regarding suspended matter, Dussart-Baptista, L. & *al.* (2003) proved the relation between turbidity and bacterial contamination; Pronk, M. & *al.* (2007) showed that bacteria (E. colie) have a greater chance to be associated with small particles ($0.9 \,\mu$ m to $2 \,\mu$ m) of matter; Mahler, B. J. & *al.* (2000) showed that the percentage of bacteria fixed in the suspended matter can vary between 50% (during dry season) and 100% (during wet season). Consequently, turbidity is usually considered as a proxy of micro-bacterial contamination (Auckenthaler, A. & *al.* 2002; Nebbache, S. L. M. A. & *al.* 1997; Ryan, M. & Meiman, J. 1996). Nevertheless, it must be underlined that the absence of turbidity is not an indicator of micro-bacterial absence. Considering dissolved matter, as it generally not modifies the turbidity, it is difficult to establish a link between both variables.

Due to the non-linearity of the processes generating turbidity, and the lack of accurate physical description of the involved phenomena, physical modelling is not frequently used to model turbidity. To estimate this quantity, the use of rainfall/runoff model is chosen for example in the study of Beaudeau, P. & *al.* (2001). However, we can highlight the works of Wicks, J. M. & Bathurst, J. C. (1996) and Kothyari, U. C. & *al.* (1997) focused on sediment transport and soil erosion that physically models the turbidity for small basins (0.06 km² for the first study and 92.5 km² maximum for the second study). Alternatively, as the significant physical information is difficult to measure *in situ*, it appears interesting to investigate other approaches based on a different paradigm: statistical modelling.

Statistical models exploit input-output information included in the database. Because they are not constrained to represent the exact physical relation, statistical models can address various global relations: velocity/turbidity, discharge/turbidity or even "hydrologic state"/turbidity. The most frequent approach consists in modelling the turbidity thanks to the discharge or the velocity of water (Towler, E. & *al.* (2010), Rajaee, T. & *al.* (2009), Kisi, O. & *al.* (2012) and Houria, B. & *al.* (2014)) investigated the flow-rate/turbidity relation. Alternatively, water quality variables as temperature, conductivity, pH, concentration of ammonium, chlorophyll-a concentration and dissolve oxygen can be used (Mulia, I. E. & *al.* (2013), Nieto, P. G. & *al.* (2014) and Iglesias, C. & *al.* (2014)). Predicting turbidity threshold overtaking was also done using the hydrologic state of the system, represented by 4 classes ("very dry", "dry", "wet" and "very wet" periods) by Mulia, I. E. & *al.* 2013), Towler, E. & *al.* (2010). The distribution of probability can be established (Tramblay, Y. & *al.* 2007). It appears that the rainfall-turbidity relation is rarely studied, (Beaudeau, P. & *al.* 2001 and Towler, E. & *al.* 2010) while rain is the triggering element of the turbidity. This is due to the complexity of the rainfall-runoff relation.

1.3. Machine learning for turbidity and water quality forecasting

Machine learning refers to a set of technics that uses databases to design a black-box model, generally nonlinear and implementing an auto-calibration process. Neural networks, genetic algorithms fuzzy logic, ... belong to this discipline. Thanks to the flexibility of these models, a great diversity of approaches can be proposed. Synthetically, it was shown that nonlinear models work better than linear ones (Beaudeau, P. & *al.* 2001). Various technics were used: association of a conceptual runoff model with a neural networks model to predict turbidity, and compare the result with a linear ARMAX model (Auto-Regressive Moving-Average with eXogenous Inputs): (Beaudeau, P. & *al.* 2001); multilayer perceptron whose training and models optimization was done by genetic algorithms: (Mulia, I. E. & *al.* 2013); neural network and Particule Swarm Optimization Houria, B. & *al.* (2014); hybrid Particle Swarm Optimization and Support Vector Machines Garcia (Nieto, P. G. & *al.* 2014); Adaptative

Neuro-Fuzzy Interference System (ANFIS), Artificial Neural Networks, Support Vector Machine and Genetic Expression Programming (Kisi, O. & *al.* 2012); Neuro-fuzzy (NF) multi linear regression (Rajaee, T. & *al.* 2009).

Regarding the input and output variables, a great diversity is also observed: rainfall, turbidity, chlorophylle-a concentration, conductivity, ammonium concentration, dissolved oxygen, pH, water level in rivers, piezometry in pitches, flow rate associated with suspend matter concentration rating curve. Originally, Iglesias, C. & *al.* (2014) used so called "synergetic variables" by coupling variables (e.g. temperature multiplicated with discharge); Nieto, P. G. & *al.* (2014) designed various models depending on wet state of the basin. Considering output variables, one can denote turbidity, chlorophylle-a concentration, suspend matter concentration.

Time steps can be short (15 minute, hours), or larger to take profit on regularization effects (day, month) and lead times evolve consequently (e.g. 12 hours lead time for hourly sampled database).

Unhopefully it appears that in several studies the database is not sufficient to be able to catch the complexity of the studied relation; for example 3 months, or one year of data. Similarly, the quality criteria are not always standard, thus preventing to make a comparison of efficiency. One can nevertheless cite the work of Rajaee, T., & *al.* (2009) who obtained R^2 (Nash criterion) between 0.7 and 0.23 to predict suspended sediment; Nieto, P. G. & *al.* (2014), Iglesias, C. & *al.* (2014) and Kisi, O., & *al.* (2012) obtained better results at a longer time step (day) with R^2 near 0.9. Nevertheless models are not always accurately described, it is thus difficult to evaluate if the model has a real added value (depending if it is feed-forward or recurrent; see section 3.3).

1.4. Organization of the paper

In the presented context, the paper is organized as following: the second part describes the site of interest: the *Yport* Pumping well, its situation and the related database. The third section addresses technics of modeling using neural networks, and the designed models: the recurrent and feed-forward options as well as the two-branches and one branch multilayer perceptron. After the general presentation of all designed models in section 4, results are presented in part 5 and discussion is proposed in part 6. Then the conclusion can consider an operational utilization of the proposed work.

2. Yport pumping well study site

2.1. The Seine Maritime Regional context

The *Yport* basin is localized in *Normandie* in the *Pays de Caux* at the west of the *Seine Maritime* department. The yearly mean temperature of this region is between 10°C and 12°C. The yearly mean rainfall is between 600 mm and 1100 mm and the high rainfall period is between October and May. This region is under the influence of continental and oceanic rainfalls. *Yport* pumping well takes water in the chalk aquifer situated at the West part of the Paris Basin. This chalk formation is composed of five levels (Laignel, B., 1997): the Campanian (30 to 40 m thick), the Santonian (30 m to more than 60 m thick), the Coniacien (20 m to more than 60 m thick), the Senonian (120 m to 160 m thick), the Turonian (70 m, to 90 m thick) and the mean and superior Cenomanian (30 m to 35 m thick). The bottom of this formation is composed of grey clay. The superficial formation is composed of 3 formations: the sand and sandy-clay deposit, generally localized in collapsed rift or karst pocket (Laignel, B. 2003; Hauchard, E. & Laignel, B. 2008), Plio-quaternary flint clay coming from chalk dissolution, quaternary loess and alluvium. These different formations are affected by structural deformation and accidents.

One can note two structural accidents: a major fault whose direction is $N150^{\circ}-160^{\circ}$ and $N120^{\circ}-130^{\circ}$ usually called "Armorican direction" (Fig. 1; faults F_1 , F_5 , F_6 and F_7) and a secondary network of faults with a $N30^{\circ}-N50^{\circ}$ variscan directions (Fig. 1; faults F_3 and F_{10}) (Hauchard, E. & Laignel, B. 2008).



2.2. Yport pumping well

In this context, the *Yport* pumping well is exploited by the *Communauté d'Agglomération Havraise* (CODAH) to supply the *Le Havre* inhabitants. The implementation of the *Yport* water station was motivated by the observation of the *Yport* Fountain localized on the tidal zone of *Yport* beach. With a flow rate estimation of its outlet spring, of 1,5 m³/s (Hole, J. P. & Roux, J. C. 1978) and 2.8 m³/s (Gaillard, T. & *al.* 2012), the alimentation basin of the *Yport* Spring appears as a good candidate for drinking water exploitation. The *Yport* pumping well was thus implanted in 1994 leading directly in a major karst drain. Due to the karst nature of the aquifer, whose area of the topographic part is only 18 km² (Gaillard, T. & *al.* 2012), several estimation basin was extended to 225 km². Complemented by the analysis of numerous tracer tests realized from 1974 to 2012, the area was reestimated to 140 km². Finally, in 2012, the SAFEGE (2012) report delimited the alimentation basin area at 320 km² (Fig. 2).



At *Yport* Plant, turbidity is measured with a nephelometer (which analyses the light scattered at 90° by the suspended particles). The nephelometer is considered as well calibrated, thus the estimation of uncertainty is the one given by the manufacturer: 2% for turbidity between 0 NTU and 40 NTU and 5% for turbidity superior to 40 NTU. Regarding the production process, when turbidity exceeds the threshold of 100 NTU it is necessary to make water longer decanting. This diminishes the output flow of 20% to 30%. Being able to anticipate the 100 NTU threshold thus would allow to anticipate by: (*i*) stock more water, (*ii*) assess the quality of treatment chain. Two goals will be thus addressed in this study: prediction of turbidity peaks, and 100 NTU threshold surpassing. The necessary anticipation time should be at least 12 hours lead-time, or, if possible 24 hours lead-time.

2.3. Database

2.3.1. Database analysis

Thanks to the database provided by the CODAH, the selection of measurement locations was done taking into account the basin boundaries. The database contains the following information:

- Rainfalls measurements recorded by 11 rain gauges since 2009 to 2015 with an hourly time step,
- Turbidity, also with hourly time step, since 2009 to 2015. The turbidity was measured at the input of the drinking water station, not in the karst drain. Thus, when the pumping is not active the measured turbidity is not the actual turbidity in the karst, but corresponds to the decantation of the water contained in the tank. This point is important to be considered in discussion.

After the pre-treatment of these data, the process of rain gauges selection was done using cross correlation (Mangin, A. 1975) using the equations 1.

$$C_{k} = \frac{1}{n} \frac{\sum_{i}^{n-\kappa} (x_{i} - \overline{x}) (y_{i+\kappa} - \overline{y})}{\sigma_{x} \sigma_{y}}$$
(1)

Where k is the offset, C_k the correlation coefficient, x and y the data, respectively \overline{x} and \overline{y} the mean of data, σ_x and σ_y the standard deviation of both datasets and n the number of considered samples.

It must be noticed that the cross correlation calculation only provides information about the linear relation linking the studied variables (a simple product is done between them). Thus because the behavior of the rainfallturbidity relation is highly nonlinear, the information obtained by this way is not fully representative of the actual relation. Nevertheless it was shown that the result of the cross correlation could be useful to initialize the process of neural network design (Kong-A-Siou, L. 2011). Cross correlation calculations between the turbidity and 11 available rain gauge rainfalls were made. Mangin, A. (1981) defined a correlation threshold of 0.2, which indicates, for his study site, the delay for which the studied variables were statistically dependent (cross-correlation higher than 0.2) and for which the variables were statistically independent (variable with a cross-correlation lower than 0.2). Thanks to this threshold, the "memory effect" which indicates the duration for which a variable influences another variable can be estimated. Moreover, the maximum of the correlation is interpreted as the "response time" between the maximum of rainfalls and the maximum of turbidity. Cross correlations were calculated between each rain gauge measurements and the turbidity at Yport. Both response times and memory effects were extracted for each cross-correlogramm and are presented in Table 1. Assuming that the 0.2 threshold can be apply to our case study, 6 rain gauges were selected.

The selected rain gauges for this study are thus; *Froberville*, *Annouville*, *Goderville*, *Anglesqueville*, *Mannevillette* and *Etainhus*, they are localized in Fig. 2. For these 6 rain gauges, the memory effect is near to 37 hours excepted for *Mannevillette*, which has a memory effect about 30h. One can note that the amplitude of the maxima of cross-correlographs are very low, showing a very complicated relation, also difficult to identify. It must be noticed that the *Mannevillette* rain gauge is not localized in the basin. Due to its costal localization, we supposed that it could be useful to represent the costal precipitations.

Rain gauges	Maximum of correlograph	Response time (hours)				
Froberville	0.19	37				
Annouville	0.20	35				
Goderville	0.25	37				
Anglesqueville	0.19	37				
Gonneville	0.15	40				
Mannevillette	0.21	30				
Octeville	0.14	30				
Etainhus	0.19	37				
St Romain	0.17	38				
St Nicolas	0.11	30				
Le Havre	0.14	40				

Table 1: Correlation coefficients for cross correlation of rainfalls and Yport turbidity.

Unhopefully, after the rain gauge selection, and due to lacuna in the rainfall data, it was not possible to design a continuous model. We thus proceeded to the selection of representative events.

2.3.2. Event selection

The selection of events was done thanks to the mean of the 6 rain gauges data. An event was selected if the cumulative mean rainfall reaches more than 3.5 mm in a sliding time-window of 24h. Following this method, 22 events were selected: 12 events without turbidity peak and 10 events with turbidity peak (Table 2). We consider that an event presents a turbidity peak when the maximum of the turbidity reaches 50 NTU. Selecting events thanks to the rainfall amount is better than selection them using only turbidity peaks because in this case, the model can distinguish the two configurations: (*i*) a rainfall event brings a turbidity peak, and (*ii*) a rainfall event doesn't bring any turbidity peak.

			Table 2: Sin	ple statistics of ev	vents in the d	atabase.						
	Events v	without turl	oidity peak		Events with turbidity peak(s)							
Event	Duration	Turbidi	ty (NTU)	Rain (mm)	Event	Duration	Turb (NT	idity TU)	Rain (mm)			
	(hours)	Max	Min	Cumul		(hours)	Max	Min	Cumul			
1	288	7	1	15.9	2	624	302	1.54	41.3			
4	384	10	0.9	14.1	3	1008	135	0	26.7			
5	336	8	1.5	17.5	6	720	245	1.5	42.0			
8	360	27	1	22.5	7	744	85	0.1	19.2			
9	384	9	1.0	20.2	10	576	256	0.9	24.9			
12	456	12	0.8	28.7	11	744	308	0.9	54.8			
13	576	13	0.9	30.8	16	648	405	0.8	53.8			
14	384	14	0.9	23.3	17	744	157	0.5	50.7			
15	600	15	0.9	31.7	18	744	87	2.2	42.8			
19	504	19	1.5	30.3	22	623	53	0.8	44.2			
20	576	20	0.9	40.8								
21	600	48	0.9	48.5								

Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance

2.3.3 Split of the database for Neural Network modelling

When designing a model using a neural network, it is necessary to split the database in several subsets: training set, stop set (usually called validation set) and test set. The role of each set will be explained in section 3.5. Depending on the test set, each model design stage has its own splitting and consequently a different database and a different resulting model.

3. Neural network modelling

Schematically, neural networks used in this study are based on the multilayer perceptron, which is composed of several neurons organized in layers. A neuron is a mathematical operator applying a transformation of its input towards its output. The architecture of the model is defined by (i) the number of intermediate layers (called hidden layers), and (ii) the number of inputs and the number of neurons in each layers. The complexity of the model is usually associated to the number of free parameters of the models; it is thus mathematically deduced from the previous numbers (Kong-A-Siou, L. & al. 2014).

3.1. Multilayer perceptron

The most used neural network in hydrological modelling is the multilayer perceptron. It has two fundamental properties: (*i*) the universal approximation (Hornik, K. & *al.* 1989) and (*ii*) the parsimony (Barron, A.R. 1993). The universal approximation implies that the model is able to identify every non-linear differentiable function, and that the accuracy of the identification depends on the number of hidden neurons (neurons in the hidden layers). The output of the model depending linearly from the parameters, the parsimony property implies that the model needs less parameters than others statistical models, whose output depends linearly from the parameters. The Fig. 3 presents a multilayer perceptron with one hidden layer of *h* hidden neurons and one output neuron. In the present study, the hidden layer is composed of nonlinear neurons using a sigmoidal function (arctg(.)). The output neuron is a linear one, calculating thus a simple weighted sum of his inputs. As this model is very famous, it is no longer detailed herein; for more information on multilayer perceptron, the reader can refer to (Dreyfus, G. 2005), or to (Kong-A-Siou, L., & *al.* 2011). The model represented in Figure is fed by a set of exogenous variables, in this study: rainfalls (**u**_r), evapotranspiration (**u**_e) and observed turbidity (y_o). The model delivers, as output, the estimated variable of interest (y^k), *k* is the discrete time step.



3.2. Two-branches model

Based on the multilayer perceptron, a specific model has been designed previously, aiming to better represent the role of evapotranspiration on the water circulating on the basin or inside the hydrological basin (Johannet, A. & *al.* 2008). In this conceptual hypothesis, evapotranspiration is seen as a removal of an amount of water from the rainfall-runoff transformation. This behaviour is implemented in the model using two branches: the main branch implements the rainfall-runoff relation (inputs are rainfalls and output is the runoff); a second branch is fed by the evapotranspiration and is connected to layer that is supposed implementing the storage of water in the aquifer. On the same way, in the case of rainfall-turbidity relation, the postulated process is divided in two branches: as illustrated in Fig. 4, the rainfall-turbidity relation is modelled by the main (upper) branch; it is a fast process forced by recent rainfall; while the evapotranspiration behaviour is represented in the lower branch: it is a slow process influenced by long-term passed rainfalls. The two-branches model is thus designed in order to differentiate both processes by the architecture itself and to help identifying the processes operating at different time-scale. The acronym UA refers the multilayer perceptron used as Universal Approximator: in place of UA-boxes of the bloc-diagram, a MLP-based UA is used based on Fig. 3 model.





3.3. Feed forward and recurrent neural network

As the behaviour of the rainfall-turbidity relation is dynamic, it is important to take into account the information about the state of the basin. This can usually be done using two kinds of models: feed-forward and recurrent models (Nerrand O. & *al.* 1993, Artigue, G. & *al.* 2012).

3.3.1. Feed-forward neural network

The feed-forward (Fig. 6) model is a multilayer perceptron fed by only exogenous inputs. Specifically, added to exogenous variables (rainfall, temperature, evapotranspiration...), this model receives variables of the measured output, here the turbidity, at previous time steps $(k-1, \ldots k-r)$, if the instant *k* is the current time step. In automatic control, this information can be considered as providing the "state" of the system (position (*k*), speed (*k* and *k*-1), acceleration (*k*, *k*-1 and *k*-2)). The feed-forward model can mathematically be explained as:

$$y^{k}(\mathbf{u}^{k},\mathbf{w}) = g_{NN}(y_{o}^{k-1},...,y_{o}^{k-r},\mathbf{u}^{k},...,\mathbf{u}^{k-m+1},\mathbf{w})$$
(2)

where y^k is the estimated turbidity, g_{NN} is the non-linear function implemented by the neural network, k is the

discrete time step, y_{o}^{k} is the measured (or observed) turbidity, \mathbf{u}^{k} is the vector of exogenous variables (rainfalls (\mathbf{u}_{r}), evapotranspiration (\mathbf{u}_{e}), etc), *r* is the order of the model, *m* is the width of the sliding time window of exogenous variables (usually various *m* are chosen for each variable during the model design), **w** is the matrix of parameters.



3.4. Training and Bias-variance trade-off

Neural networks are statistical models that are designed in relation to a database. Contrarily to other models implementing a behavioural law, whose parameters are adjusted during a calibration stage; neural networks need to adjust simultaneously the "law" to implement and the parameters of this "law". To this end they need: (*i*) more parameters to adjust, and (*ii*) a careful design in order to identify the right behavioural law underlined by/in the data. This is done during the training stage where the parameters are iteratively calculated using, usually, a gradient descent aiming to minimize a mean squared error (the training error). In this study the Levenberg-Marquardt training rule was chosen (Levenberg, K. 1944 and Marquardt, D. W. 1963). Moreover the quality of the obtained model after training is evaluated on a "test set" that ideally represents an archetypal situation. For example if the model is dedicated to flash flooding, it must be evaluated on the most important and rapid flood of the database; for droughts, it must be evaluated on the more intense drought, etc. The model's ability to be efficient on the test set is called *generalization*.

In 1992, (Geman, S. & *al.* 1992) formalized "the bias-variance trade-off" that is the major trap of machine learning. It expresses the idea that the training error is not representative of the generalization error because the efficiency of the training algorithm makes the model specific to the training set. This specialization of the model over the training set is called overtraining. Overtraining is exacerbated by large errors and uncertainties in field measurements; the model then learns the specific realization of the noise in the training set. The bias-variance trade-off is characterized by the promoting of the bias at the expense of the variance and vice versa. The bias and the variance have an inverse variation. The bias decreases with the increase in the number of parameters in the model. This variation can be explained by the better adjustment of the model with higher number of parameter. The variance is influenced by the complexity of the model and higher is the complexity, higher is the sensitivity of the model to the signal particularity and noise. The bias/variance dilemma can be described as the research of equilibrium between bias and variance. This equilibrium is defined as the minimum of the bias and variance sum beyond which the model is considered as overfitted. The trap of bias-variance can be avoided using regularization methods, particularly cross-validation and early stopping.

3.5. Regularization methods

To avoid overfitting and increase robustness of the models, three regularization methods can be used: cross validation, early stopping, and ensemble models.

3.5.1. Early stopping

Early stopping prevents the model from overtraining (Stone, M. 1974). It necessitates that a stop set is set apart from the database. The training procedure is stopped thanks to the stop set (usually improperly called validation set in the literature), and the model quality is measured on the remaining part of the database: the test set, which is separate from the both previous sets. The choice of the stop set is crucial as it influences the model in a very important way. For this reason we proposed in (Toukourou, M. & *al.* 2011) to choose the stop set, for each model, in taking the set having the best score when used in test set. This choice guarantees that there is a strong coherence between the training set and the stop set.

3.5.2. Cross validation

The goal of the cross validation (Stone, M. 1974) is to diminish the variance. To this end, the training set is divided in K subsets. A subset can be a year or a flood event. The model is trained using the K-1 remaining subsets and a quality criterion is calculated on the remaining set which is then considered as a validation set. Each subset becomes a validation set, each times its turn and the training is done on the remaining database. At the end of the procedure, the qualities score (S) is calculated thanks to the error calculated for each set (S_i):

 $S = \frac{1}{\kappa} \sqrt{\sum_{i=1}^{K} (S_i)^2}$

Depending on the goal of the model, different scores can be defined used various quality criteria ($S_i(.)$).

By calculating the cross-validation score for several parameters initializations and model architectures (number of hidden neurons, input window widths), one can select the model with the smallest generalization error

(4)

(Dreyfus, G. & al. 2008, Kong-A-Siou, L. 2011). By this way, the variance (of the bias variance tradeoff) is minimized.

3.5.3. Ensemble models

Darras, T. & *al*, (2014) showed that sometimes the cross-validation was unable to select an efficient model because the random initialization of the model has a predominant effect. In order to overcome this problem Darras, T. & *al*, (2014) proposed to build an ensemble model and to calculate, at each time step, the output as:

$$y_{M}^{k} = Median_{X}(y_{i}^{k})$$

(5)

where y_M^k is the output of the ensemble model at the time-step k, y_i^k is the output of one model inside the ensemble at time-step k, and *Median*_X represents the median calculated on the outputs of a set of X models. The number of models in the ensemble must be chosen depending the application.

In the present work, after a study following the method of (Darras, T. & *al*, 2014), 10 models were generated in the ensemble, differing only by their parameters initialisation; also the considered output of the ensemble, is not the Median₁₀, but the Maximum₁₀:

 $y_{X}^{k} = Maximum_{10}(y_{i}^{k})$

with the same notation as previously.

3.6. Model optimization

The optimization of the model is made thanks to the cross-validation method operating on the output of the ensemble model (see eq. 5). This adjustment is proceed by adjusting hyperparameters using cross-validation:

- The window widths of the different (exogenous) input variables: *m* in equations 2 and 3.
- The order of the model, corresponding to the window width of the estimated (or observed if the model is non-recurrent) targeted variable (output variable), for previous time steps, applied at the input of the model: *r* in equations 2 and 3.
- The number of neurons in the hidden layers: $n_{\rm h}$

Also, the resulting model could depend on several supplementary choices:

- The random initialization of the model: this effect is overcome using the ensemble model
- The number of training iterations: this choice is made automatically thanks to early stopping.
- As an example, the Fig. 8 shows the evolution of the cross validation score (S) versus the number of hidden neurons (n_h) .



3.7. Quality criterion

It is important to evaluate the quality of models using impartial criteria. As two goals of modelling are investigated in this work: peak prediction, and threshold overtaking, 2 kinds of criteria will be used, they are presented hereafter.

3.7.1. Nash criterion

The Nash criterion is often used in hydrology (Nash, J.E. & Sutcliffe, J.V. 1970); it corresponds to the *R*² determination coefficient that expresses the ratio between the simulated variance and the observed variance:

$$R^{2} = 1 - \frac{\sum_{k=1}^{n} (y_{o}^{k} - y^{k})^{2}}{\sum_{k=1}^{n} (y_{o}^{k} - \overline{y_{o}})^{2}}$$

where:

 y_{o}^{k} is the observed value at the time k,

 \mathbf{y}^{k} is the simulated value at the time k

n is the number of samples in the considered set,

 y_{0}^{k} is the average of the observed values in the considered set.

This criterion has to be close to one, which means that the predicted values are equal to observed values. Generally speaking, for flash responses, where an event modelling is requested, a Nash criterion value superior to 0.8 is considered as a good value. Nevertheless, as the quality of the prediction concerns specifically the peak, other criteria more sensitive to the quality of peak prediction were added.

(6)

3.7.2. The peak percentage (P_P)

The peak percentage (P_P) was introduced by Artigue, G. & *al.* (2012), it provides the information about the quality of the peak value estimation. The P_P is estimated thanks to the following expression:

$$P_{P} = \frac{Max(y)}{Max(y_{o})},$$

(7)

Then, if $Max(y_0) \ge Max(y)$, the ratio is ≤ 1 ; in the contrary the ratio is >1. A value lower than one indicate thus that the estimated peak value is lower than the observed one.

3.7.3. The peak delay (P_D)

The lag: this value indicates the time delay between observed and measured peaks. It will be written in time step, that is to say in this study in hours. If this value is positive, the prediction is late. If it is negative, the prediction is in advance (Fig. 9).

$$P_D = k_o^{\max} - k^{\max}$$
(8)



Figure 9: Representation of the P_D and P_p criteria. Dash line for observed peak, and dot line for predicted peak.

3.7.4. Errors in threshold detection

Regarding the second kind of information useful for operational needs: the threshold overtaking, we proposed to detect the number of errors in warning: false positive and false negative. We can thus introduce 4 new criteria:

- False positive: F_p is equal to 1 if the prediction proposes a threshold overtaking while it is absent

- False negative: F_n is equal to 1 if the prediction has no threshold overtaking while this threshold is surpassed

- Delay in rising: Dr represents the delay between the actual threshold overtaking and the predicted one.

- Delay in decreasing: D_d represents the delay between the actual threshold crossing during decreasing and the predicted one.

Of course, when an event of rainfall don't lead to a measured turbidity peak, the previous criteria can't be calculated.

4. Overview of the designed models

As the behaviour to model using neural networks is supposed to be highly nonlinear, various trials of modelling strategies were done. Synthetically one has to choose the time-step of the signals and the representation of rainfall. This was done, step by step, in modifying only one "option" between the following: (*i*) time-step, (*ii*) mean rainfall/spatialized rainfalls. Thanks to operational needs (section 2.2), two prediction horizons were targeted: 12 hours lead-time and 24 hours lead-time. It must be noticed that results of the models designed at 1 hour time-step and 12 h-lead time have been described already in a conference paper (Savary, M. & *al.* 2017); investigations are extended in the present paper to 6 hours time-step and 24 hours lead-time. This added work allows to achieve the design of the models and to strengthen the proposed conclusions. Also, illustrations are proposed on different event: event 11 in the present paper, event 2 in (Savary, M. & *al.* 2017).

Designed architectures using cross validation are synthetized in Table 3. For each model, 2 goals will be presented hereafter: "turbidity prediction" and "100 NTU threshold overtaking". The latter is important to operational needs as seen in section 2.2.

In order to be able to discuss the performances of models regarding the bias-variance trade-off (section 3.4), the complexity of each model is provided by the number of free parameters for each model. Indeed an augmentation of the number of parameters automatically increases the complexity of the model and thus makes the model more sensitive to overfitting. For each condition, the both models: multilayer perceptron (MLP) and two-branches MLP were designed; for each model the two possibilities (recurrent or feed-forward) were also investigated.

All these possibilities lead designing 32 models: 8x2x2=32 models, the following section aims to present them and their properties. This represents a huge quantity of information; nevertheless, as this work is one of the first works addressing straightforwardly the rainfall-turbidity modelling, we estimated that it could be interesting to present them quite exhaustively.

- This presentation will allow deducing rules regarding operational or research goals:
- Is it useful to increase the time-step in order to decrease the complexity of the model?
- Does the rise of time-step allow extending the lead-time?
- Is the non-recurrent model less complex than the equivalent recurrent model?
- Does the spatialized rains provide significant advantage balancing the rise in complexity?

Table 3: Designed models. For each investigated configuration, 4 models were designed: recurrent/feed-forward and simple MLP/ two-branches models

Sampling	12 hou	rs lead-time	24 hours lead-time				
Rainfall data	Mean rainfall	Spatialized rainfalls	Mean rainfall	Spatialized rainfalls			
One hour time-step	Х	Х	Х	Х			
Six hours time-step	Six hours time-step X		Х	Х			

For each model, the evapotranspiration was applied using a very simple expression as suggested by Oudin, L. & *al.* (2005a) and Oudin, L. & *al.* (2005b), who tested 27 different potential evapotranspiration estimations with 4 lumped models on 308 catchments in France, Australia and the United States. Their conclusion was that a very simple evapotranspiration model was as efficient as Penman-Monteith evapotranspiration in rainfall-runoff modelling applications. Moreover lumped rainfall-runoff models can be improved by using simpler evapotranspiration estimations. Contrarily to (Oudin, L. & *al.* 2005a,b) the present work is not addressing rainfall-runoff modelling, nevertheless, as turbidity may come from surface rivers (exogenous origin), it is logical to follow the recommendations of Oudin, L. & *al.* 2005a,b. The chosen evapotranspiration is thus a Gaussian curve centred at the end of June. This kind of evapotranspiration was found efficient to model rainfall-runoff on a karst aquifer by (Kong-A-Siou, & *al.* 2014).

Cross-validation was used to define the window widths of input variables (e.g. m in eq. 1); regarding rainfall: displayed window widths in section 5 include the future rainfalls over a window equal to the lead-time. For example, a future window width equal to 12 (12 h lead-time) and a past window width of 10 h (rain past history of 10h) lead to a total window width of 22 time-steps.

5. Results

The database contains 22 rainfalls events whose only 10 events triggering a turbidity peak (turbidity>40 NTU), a kind of *cross-test* was performed in order to assess the quality of prediction on each event. During the simulations, it appeared that there was never a false turbidity peak, predicted for an event without measured turbidity peak (listed in Table 2). In the following, the test results are thus presented only on all events of turbidity of the database.

Let us recall that in the following, for each model, an ensemble is built as explained in section 3.5.1. The considered output is the Maximum₁₀ at each time step, in order to not miss any turbidity event, and because usually the output underestimated the measured value. Thus Nash criteria, P_p and P_d , are calculated thanks to the Maximum₁₀ (see eq. 5) of the ensemble model.

Uncertainty coming from the models themselves is not skilfully defined, but simply visualized in the hydrographs as proposed by (Kong-A-Siou, L & *al*, 2015). It is shown in addition to the turbidigraph by a grey band defined between the second highest value of the ensemble and the second lowest value. This uncertainty corresponds to the variability induced by the initialization of the parameters, except the outliers.

On the following figures, readers will see some information: (*i*) the mean rainfall, (*ii*) the uncertainty as defined previously, (*iii*) the Median₁₀ of the ensemble. To not add confusion, the output is not represented but is very close to the maximum of the grey band of uncertainty. Figures allow estimating the dynamics of the turbidigraph for one event while Table 4 to Table 27 will present the quality criteria calculated for each event and synthetized on the database. The event 2 was chosen because it is both intense and with double peak. Predicting this kind of event is crucial and difficult.

5.1. One-hour time-step and 12-hours lead-time models

5.1.1. Mean rainfall

Four architectures were designed using signals sampled at one-hour time-step, 12-hours lead-time and mean rainfall. They are summarized in the Table 4 and an illustration is provided in Fig 10.

,	A	ML	P	Two branches			
F	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent 15 1 15		
	Rainfall layer	Х	Х	10	15		
Hidden layers	Evapotranspiration layer	Х	Х	1	1		
	Output layer	5	5	10	15		
	Rainfall	50	30	50	50		
Input windows	Evapotranspiration	3	3	3	3		
_	Order	1	1	10	1		
Numb	per of parameters	281	181	745	1265		

Table 4: Synthesis of the selected architectures for one-hour time-step, 12-hours lead-time and mean rainfalls

It appears first that the complexity of models is very different from one model to another. The perceptron models are always more parsimonious than the two-branches models; this is inherent to the architecture, as the MLP architecture is simpler. Also two-branches models implement a more sophisticated relation for the role of evapotranspiration. The two-branches feed-forward model has a higher order than the three other models; this suggests that the dynamics of the process is better represented using this architecture. The multilayer perceptron recurrent model is the model that needs the smallest rainfall window width. Maybe this should be an indicator that this model doesn't manage very well the complexity of the rainfall-turbidity relation. These suggestions are coherent with the performance criteria presented in Table 5: the two-branches feed-forward model has the slower delay and the multilayer perceptron recurrent model has the worse peak prediction (regarding the median).

Turbidity peak prediction

Regarding the peak of turbidity prediction (more simply called hereafter "peak prediction"), the quality of the predicted peaks is estimated using the P_P and P_D (presented in section 3.7). Table 5 presents the obtained quality criteria for each event in cross-test situation. Each line presents the scores obtained for a specific event in cross test. The two last lines sum up the performances by calculating the median and the mean for all the events. Looking at the difference between the median and the mean, it appears first that, satisfactorily, the forecasting behaviour is quite stable on the whole test database. Moreover, the median score reveals that the two branches feed-forward model seems to be the best model: its median for the peak percentage presents an overestimation of only 6% while the median delay is only 4h.

Let us recall that the stop set is not always the same for each event in test. This is because the stop set is chosen as being the best event in prediction when performing the cross validation (see section 3.5.1).

		I	MLP			Two-	branches	
Б. (Feed-	forward	Re	current	Feed-	forward	Re	current
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	$P_d(\mathbf{h})$
T 2	<u>81</u>	19	40	21	80	<u>4</u>	38	18
T 3	<u>109</u>	4	71	<u>3</u>	113	4	61	15
T 6	<u>93</u>	14	70 10		217 <u>8</u>		80	10
T 7	<u>83</u>	14	75 3		Ste	op set	75	<u>2</u>
T ₁₀	50	14	Stop set		<u>86</u>	7	41	<u>4</u>
T11	59	8	26 <u>3</u>		<u>80</u>	15	28	10
T ₁₆	66	14	44	10	<u>106</u>	<u>2</u>	S	top set
T ₁₇	Ste	op set	50	9	<u>106</u>	<u>5</u>	88	101
T18	90	1	80	<u>0</u>	<u>98</u>	2	153	3
T ₂₂	169	4	149	6	<u>144</u>	<u>3</u>	152	14
Median	83	14	70	6	<u>106</u>	<u>4</u>	75	10
Mean	89	10	67	7	<u>114</u>	<u>6</u>	80	20

 Table 5: Quality criteria for one-hour time-step, 12-hours lead-time and mean rainfalls. For each event, the best criterion is underlined. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

Threshold overtaking

Regarding the application of the previously presented models to threshold overtaking, the 100 NTU threshold was chosen for operational needs (see section 2.2). Table 6 presents the quality criteria obtained with the previously presented models. These criteria are different from the criteria used in the Table 5, they were presented in section 3.7.4. Following these quality criteria, the feed-forward multilayer perceptron is now the best model. Nevertheless it can be noticed that no false negative is recorded. This is a very useful property of the models.

			(see Ta	ble 2) ar	e not pr	esented	becaus	e there is	s no fal	se posit	ive in th	eir pred	ictions.			
				М	LP				Two branches							
Event]	Feed-	forward	l		Recu	urrent		Feed-forward					Recurrent		
	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)
T2	0	0	6	8	0	0	6	-53	0	0	6	11	0	0	6	-76
T3	1	0	7	4	0	0	0	-22	1	0	0	3	1	0	-11	-4
T ₆	0	0	-4	9	0	0	-4	4	0	0	-4	10	0	0	-5	6
T 7	0	0	Х	Х	0	0	Х	Х		Sto	p set		0	0	Х	Х
T ₁₀	0	0	-9	11		Sto	p set		0	1	-2	10	0	0	-1	9
T ₁₁	0	0	-1,5	-3 / 4	0	2	Х	Х	0	1	3	6	0	1	-17	-31
T16	0	0	1	12	0	0	6	-19	0	0	14	18		Stop	o set	
T ₁₇		Sto	op set		0	3	Х	Х	1	3	Х	Х	0	3	Х	Х
T ₁₈	0	0	Х	Х	1	0	Х	Х	0	0	Х	Х	0	0	Х	Х
T22	0	0	Х	Х	1	0	Х	Х	1	0	Х	Х	0	0	Х	Х
Median (+ -)	11%	0%	6 -4	9 X	22%	55%	6 0	4 -22	33%	55%	5 -3	10 X	11%	44%	6 0	8 -31
Mean (+ -)			5 -5	8 X			4 0	4 -31			6 -3	10 X			6 0	8 -37



Figure 10: Turbidity forecasting with a lead-time of 12h, using one-hour time-step and mean rainfall. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

5.1.2. Spatialized rainfall

In this section, rainfalls are spatialized to take into account the spatial heterogeneity of rainfalls and the possible higher intensities of rain. The information about rainfall is thus inputted to the models from 6 rain gauges (*Froberville, Annouville, Goderville, Anglesqueville, Mannevillette* and *Etainhus*), as presented in section 2.3.1. Raingauges are localized in Fig. 2. Thanks to the procedure presented in section 3.6, four architectures were designed; they are summarized in Table 7 and an illustration is provided in Fig. 11.

	A	ML	Р	Two branches			
1	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent		
	Rainfall layer	Х	Х	10	15		
Hidden Layers	Evapotranspiration layer	Х	Х	1	1		
	Output layer	5	5	10	15		
	Past Rainfall (*6)	38	18	38	38		
T	Futur Rainfall	12	12	12	12		
Input window	Evapotranspiration	3	3	3	3		
	Order	1	1	10	1		
Numb	per of parameters	1231	631	2645	3905		

Table 7: Synthesis of the selected architectures for spatialized rainfalls, one-hour time-step and 12-hours lead-time models.

It appears that the complexity of the models have greatly increased due to the window width of the 6 raingauges. Actually models are very similar to the previous models using mean rainfalls, only rainfalls window widths are different between both kinds of models. We will see in the next sections if the quality increased also.

Turbidity peak prediction

Regarding the peak prediction, Table 8, similarly to Table 5, presents the obtained quality criteria for each event in cross-test situation.

Table 8: Quality criteria for one-hour time step, 12-hours lead-time models and spatialized rainfalls. For each event, the best criterion is underlined. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

		М	LP		Two-branches						
	Feed-f	forward	Rec	urrent	Feed-f	forward	Recu	urrent			
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)			
T2	<u>90</u>	<u>16</u>	32	20	128	17	36	27			
T 3	<u>105</u>	8	77	<u>3</u>	102	4	80	84			
T 6	<u>117</u>	11	46	12	174	<u>11</u>	56	13			
T ₇	<u>80.5</u>	14	82 4		Sto	p set	62	<u>8</u>			
T10	73	14	Stop set		<u>85</u>	9	64	<u>3</u>			
T ₁₁	68	14	29	<u>3</u>	<u>66</u>	11	31	11			
T ₁₆	62	15	40	40 12 <u>22</u>		<u>9</u>	Sto	p set			
T17	Sto	p set	49	11	<u>95</u>	<u>7</u>	61	8			
T ₁₈	115	14	83	<u>69</u>	<u>135</u>	10	93	13			
T22	129	10	172	7	<u>230</u>	<u>10</u>	182	8			
Median	95	<u>13</u>	68	15	<u>137</u>	20	74	19			
Mean	95	12,9	68	15,7	137	9,8	74	19,4			

Looking at Table 8 and Table 5, it is easy to note that the peak prediction is not improved by the application of spatialized rainfalls. Actually the quality criteria are deteriorated. The best model seems to be the Feed-forward MLP; its peak is good but the delay very bad.

Threshold overtaking

Regarding the application of the previously presented models to threshold overtaking, quality criteria are presented in Table 9. Contrarily to what happened previously, criteria are sometimes better and sometimes poorer. Two models appeared as equivalent in quality: two-branches recurrent and feed-forward multilayer Perceptron.

 Table 9: Quality criteria for spatialized rainfalls, one-hour time step and 12-hours lead-time models. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

				MI	LP						r	Гwo-br	anche	s		
		Feed	-forwar	d		Rec	urrent		Feed-forward				Recurrent			
Event	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)
T_2	0	0	-5	8	0	0	0	-88	0	0	-3	-8	0	0	-9	-32
T 3	1	0	4	-24	1	0	4	-24	0	0	6	-3	1	1	Х	Х
T ₆	1	0	-1	-15	1	0	1	-15	0	0	4	-10	1	0	4	-9
T ₇	0	0	Х	Х	0	0	Х	Х		St	top set		0	0	Х	Х
T10	0	0	3	17		Sto	op set		0	0	-4	-11	0	0	-3	17
T11	0	2	Х	Х	0	2	Х	Х	0	0	07-sept	1	0	1	X/-25	X/-20
T ₁₆	0	0	-6	-19	0	0	-6	-19	0	0	-6	-13		St	op set	
T17		St	op set		0	3	Х	Х	0	2	13	-13	0	2	10	-8
T ₁₈	0	0	Х	Х	0	0	Х	Х	1	0	Х	Х	0	0	Х	Х
T ₂₂	1	0	Х	Х	1	0	Х	Х	2	0	Х	Х	1	0	Х	Х
Median																
(+ -)	33%	22%	4 -5	13 -19	33%	55%	1 0	X -22	33%	22%	7 -4	1 -11	33%	44%	7 0	17 -15
Mean	5570	2270	4 -4	13 -19	5570	5570	2 0	X -37	5570	2270	8 -4	1 -10	5570	-77/0	7 0	17 -17
(+ -)																



Figure 11: Turbidity forecasts with a lead-time of 12h using one-hour time-step and spatialized rainfall. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

5.2. Six-hours time-step and 12-hours lead-time models

5.2.1. Mean rainfall

In this section the sampling time step is now equal to 6 hours in order to diminish the number of parameters, i.e. the complexity of the model, to improve the generalization. The four selected architectures are summarized in the Table 10 and an illustration is provided in Fig 12.

	. 1.	ML	P.	Two bra	anches
F	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent
	Rainfall layer	Х	Х	11	11
Hidden layers	Evapotranspiration layer	Х	Х	1	1
	Output layer	5	13	10	13
	Rainfall	10	5	10	10
Input windows	Evapotranspiration	3	3	3	3
	Order	4	8	3	1
Numb	er of parameters	96	235	366	477

Table 10: Synthesis of the selected architectures for six-hours time-step, 12-hours lead-time and mean rainfalls.

First, compared to Table 4, one can note that the mean number of parameters has decreased of a factor 2.4. The lengths of rainfall window widths are quite similar (in hours). For example, in Table 4 the length of input window is 30 hours for the recurrent multilayer Perceptron and is 5*6=30 hours for the same model in Table 10, this is not the case for the other inputs: order or ETP. Moreover, strangely, comparing Table 4 and Table 10, it appears that the number of neurons is sometimes greater using 6-hours time-step, sometimes lower.

Peak prediction

Regarding the application of the previously presented models to peak prediction, Table 11 presents the quality criteria. Comparing with Table 5, which presents same criteria obtained with the one-hour time-step models, it appears that the anticipation of the models is worse. Indeed looking at the line presenting the median, the delays are equal to 3 time-steps, i.e. 18 h versus 3 hours in case of one-hour time step.

For this application, resampling to 6-hours diminishes significantly the complexity, but decreases dramatically the ability of anticipation as the median of delay on the peak prediction is 18h.

Table 11: Obtained quality criteria for six-hours time-step, 12-hours lead-time and mean rainfalls. For each event, the best criterion is underlined. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

		М	LP			Two-b	ranches		
-	Feed-	forward	Rec	urrent	Feed-t	forward	Recurrent		
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	
T2	<u>68</u>	18	41	24	148	<u>24</u>	39	18	
T ₃	<u>96</u>	18	78 <u>12</u>		111	111 12		18	
T ₆	<u>117</u>	18	71	12	243	<u>18</u>	81	24	
T ₇	<u>81</u>	24	67 6		Sto	p set	75	<u>12</u>	
T10	60	24	Stop set		<u>75</u>	18	48	<u>12</u>	
T11	66	24	34	<u>12</u>	<u>63</u>	18	28	12	
T16	63	24	35	24	<u>132</u>	<u>18</u>	Sto	p set	
T17	Sto	p set	44	12	<u>122</u>	<u>12</u>	54	12	
T ₁₈	91	24	83	<u>42</u>	<u>92</u>	12	75	30	
T ₂₂	149	12	167	18	<u>187</u>	<u>18</u>	162	12	
Median	88	18	69	18	<u>130</u>	<u>18</u>	70	18	
Mean	88	20,4	69	18	130	16,8	70	16,8	

Threshold overtaking

Regarding the application to threshold overtaking, Table 12 presents the quality criteria. In this case, the comparison between Table 6 and Table 12 shows that the behaviours are quite similar. No advantage is thus taken by resampling the data.

	MLP								Two branches								
-		Feed	-forwar	d		Re	current			Feed-forward				Recurrent			
Event	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	F_{p}	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	F_{p}	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	
T ₂	0	0	0	12	0	0	0	-36	0	0	0	18	0	0	0	-60	
T 3	1	0	0	0	0	0	0	-24	1	0	0	0	1	0	6	-12	
T ₆	0	0	12	6	0	0	6	-12	0	0	6	6	0	0	12	6	
T 7	0	0	Х	Х	0	0	Х	Х		S	top set		0	0	Х	Х	
T ₁₀	0	0	6	18		St	op set		0	0	18	18	0	0	1	2	
T ₁₁	0	0	18/18	0/6	0	1	X/18	X/-42	0	0	12-juin	06-juin	0	2	Х	Х	
T ₁₆	0	0	6	18	0	0	0	24	0	0	-6	18		Stop	set		
T ₁₇		St	op set		0	3	Х	Х	0	2	18	18	0	2	30	6	
T ₁₈	0	0	Х	Х	0	0	Х	Х	0	0	Х	Х	0	0	Х	Х	
T ₂₂	0	0	Х	Х	0	0	Х	Х	1	0	Х	Х	0	0	Х	Х	
Median (+ -) Mean (+ -)	11%	0%	6 X 9 X	6 X 9 X	0%	44%	0 0 2 0	24 -18 24 -20	22%	22%	6 -6 9 -6	12 X 11 X	11%	44%	6 0 10 0	6 -36 5 -36	

 Table 12: Quality criteria for mean rainfalls, six-hours time-step and 12-hours lead-time models. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.



Figure 12: Turbidity forecasts with a lead-time of 12h using one-hour time-step and mean rainfall. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

5.2.2. Spatialized rainfall

Using signals re-sampled to 6-hours time-steps, the four selected architectures are summarized in the Table 13. Comparison with the Table 10 shows that there is more neurons for the models fed by spatialized rainfalls; the complexity (number of parameters) is thus greater, helped by the spatialization of rains that increased the number of input variables. An illustration is provided in Fig 13.

	A 1 4	ML	.P	Two bra	unches
F	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent
	Rainfall layer	Х	Х	5	7
Hidden layers	Evapotranspiration layer	Х	Х	1	1
	Output layer	14	7	15	15
	Past Rainfall (*6)	8	8	8	8
T / 1	Futur Rainfall	2	2	2	2
Input windows	Evapotranspiration	10	3	3	3
	Order	4	8	3	1
Numb	er of parameters	925	442	425	527

Table 13: Synthesis of the selected architectures for six-hour time-step, 12-hours lead-time and spatialized rainfalls.

Peak prediction

Regarding the application of the previously presented models to peak prediction, Table 14, similarly to Table 11, presents the obtained quality criteria for each event in cross-test situation. The evidence is that the spatialization of rainfalls has worsened the results. Comparing Table 8 and Table 14, it is easy to note that the peak prediction is not improved by the re-sampling of data. Actually, as previously, the quality criteria are deteriorated by re-sampling at 6-hours time-step.

		М	LP			Two-bi	ranches	
	Feed-	forward	Rec	urrent	Feed-f	forward	Recu	ırrent
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)
T ₂	<u>67</u>	54	41	18	186	<u>36</u>	38	36
T ₃	<u>94</u>	18	64	<u>16</u>	115	12	41	18
T 6	<u>85</u>	18	60	18	197	<u>24</u>	61	18
T 7	<u>63</u>	30	52	36	Sto	p set	73	<u>12</u>
T10	65	24	Sto	p set	<u>116</u>	18	57	<u>12</u>
T11	68	24	31	<u>18</u>	<u>66</u>	18	34	18
T ₁₆	60	24	40	18	<u>250</u>	<u>12</u>	Sto	p set
T ₁₇	Sto	p set	49	6	<u>120</u>	<u>12</u>	56	18
T ₁₈	105	102	94	<u>35</u>	<u>89</u>	30	113	24
T ₂₂	89	18	115	12	<u>293</u>	<u>18</u>	193	12
Median	77	36	61	18	159	18	74	18
Mean	77	35	61	18	159	20	74	19

 Table 14. Quality criteria for spatialized rainfalls, six-hours time-step and 12-hours lead-time models. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

Threshold overtaking

For threshold overtaking, quality criteria are presented in Table 15. Comparing Table 15 and Table 12 (mean rainfalls, same sampling), spatialized rains increase the delays. The same applies to the comparison between Table 15 and Table 9 (one-hour time-step). Resampling increases the delays. Satisfactorily there is no false negative event.

				ML	Р							Two b	oranche	s		
		Fe	ed-forward			Rec	urrent			Feed-	forward	1		Rec	urrent	
Event	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	<i>d</i> _d (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	F_n	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)
T ₂	0	0	12	12	0	0	0	-72	0	0	-18	12	0	0	18	-18
T ₃	0	0	6	0	0	1	Х	Х	0	0	0	0	0	0	6	-12
T ₆	0	0	18	-6	0	0	12	-12	1	0	-6	0	1	0	6	-6
T 7	0	0	Х	Х	0	0	Х	Х		Sto	op set		0	0	Х	Х
T 10	0	0	18	18		Sto	op set		0	0	-6	18	0	0	0	24
T11	0	0	-12/-18	-6	0	1	-6/X	-6/X	1	0	-0,5	-6	0	0	0/18	1,5
T ₁₆	0	0	0	6	0	0	6	6	0	0	-6	12		Sto	p set	
T 17			Stop set		1	3	Х	Х	1	2	12	18	0	3	Х	Х
T ₁₈	0	0	Х	Х	0	0	Х	Х	0	0	Х	Х	2	0	Х	Х
T ₂₂	0	0	Х	Х	0	0	Х	Х	2	0	Х	Х	2	0	Х	Х
Median																
(+ -)	0%	0%	12 X	9 -6	11%	55%	6 0	6 -12	55%	22%	12 -6	12 -6	55%	33%	6 0	24 -15
Mean (+ -)	070	070	11 X	9 -6	11/0	2070	6 0	6 -30	2070	/0	8 -7	10 -6	2370	2370	8 0	24 -14

 Table 15. Quality criteria for spatialized rainfalls, six-hours time-step and 12-hours lead-time models. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.



Figure 13: Turbidity forecasts with a lead-time of 12h using one-hour time-step and spatialized rainfalls. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

5.3. One-hour time-step and 24-hours lead-time models

5.3.1. Mean rainfall

Aiming to predict at 24-hours lead-time thanks to mean rainfall, four architectures were designed. Comparing the Table 16 to Table 4 (1h time-step, 12h lead-time), one can note that the numbers of neurons are the same. Increasing the lead-time seem thus to not necessitate more complexity. Regarding the temporal window widths, they are generally longer for the models aiming to predict 12h ahead. It appears thus that the 24h lead-time architectures couldn't be more able to deal with complicated hydrological situations. An illustration is provided in Fig. 14.

,	\	ML	P	Two bra	anches
F	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent
	Rainfall layer	Х	Х	10	15
Hidden layers	Evapotranspiration layer	Х	Х	1	1
	Output layer	5	5	10	15
	Rainfall	45	30	30	35
Input windows	Evapotranspiration	3	3	3	3
	Order	1	1	10	1
Numb	er of parameters	256	181	545	1040

Table 16. Synthesis of the selected architectures for one-hour time-step, 24-hours lead-time mean rainfalls.

Peak prediction

Regarding the application of the previously presented models to peak prediction, Table 17 presents the quality criteria. One can note that no model can be efficient for both criteria: peak prediction and delay. Focusing on the time delay, it appears that the best model is the recurrent two-branches model Comparing Table 17 to Table 5 (12h lead-time with 1-hour time step), it is noticeable that the extension of the lead-time has decreased or no improved the results for the four architectures.

Table 17: Quality criteria for mean rainfalls, one-hours time-step and 24-hours lead-time models. Results of events without turbidity
peaks (see Table 2) are not presented because there is no false positive in their predictions.

		М	LP			Two-b	ranches	
	Feed-	forward	Rec	urrent	Feed-f	forward	Recu	urrent
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)
T2	58	16	<u>71</u>	<u>15</u>	42	<u>15</u>	30	18
T 3	<u>93</u>	6	65	<u>3</u>	29	4	63	4
T ₆	141	<u>9</u>	<u>69</u>	<u>9</u>	136	<u>9</u>	58	10
T ₇	60	<u>13</u>	60	6	Sto	p set	<u>69</u>	<u>4</u>
T10	<u>38</u>	<u>2</u>	Sto	p set	28	2	36	3
T11	<u>48</u>	10	27	<u>1</u>	27	28	25	8
T16	<u>59</u>	16	48	<u>9</u>	57	22	Sto	p set
T17	Sto	p set	<u>53</u>	9	4	21	44	<u>3</u>
T ₁₈	<u>106</u>	51	82	70	56	<u>17</u>	68	62
T ₂₂	157	6	<u>123</u>	<u>6</u>	11	63	141	6
Median	60	10	<u>65</u>	9	27	22	58	<u>6</u>
Mean	<u>84</u>	15	66	15	38	26	59	<u>13</u>

Threshold overtaking

Regarding the application of the previously presented models to threshold overtaking, Table 18 presents the quality criteria. In this case, the comparison between Table 6 and Table 18 shows that the conclusions are quite similar to the previous ones (peak prediction). Extending the lead-time to 24h prediction results in degraded models.

				MI	_P						1	Two br	anches			
		Feed	-forwar	d		Re	current			Feed-f	orward			Recu	irrent	
Event	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	F_{p}	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	F_{p}	Fn	$d_{\rm r}$ (h)	<i>d</i> _d (h)
T2	0	0	-6	-9	0	0	-3	-91	0	0	-4	-8	0	1	Х	Х
T ₃	1	0	6	7	0	0	4	-23	0	1	Х	Х	0	1	Х	Х
T ₆	0	0	4	9	0	0	4	-8	0	0	3	10	0	0	10	-2
T7	0	0	Х	Х	0	0	Х	Х		Stop	o set		0	0	Х	Х
T10	1	1	Х	Х		S	top set		0	1	Х	Х	0	1	Х	Х
T11	0	1	X/9	X/-12	0	1	X/18	X/-38	0	1	X/21	X/5	0	2	Х	Х
T ₁₆	0	0	-10	21	0	0	-6	17	0	0	-9	17		Stop	o set	
T17		St	op set		0	3	Х	Х	0	3	Х	Х	0	3	Х	Х
T ₁₈	1	0	Х	Х	0	0	Х	Х	0	0	Х	Х	0	0	Х	Х
T ₂₂	0	0	Х	Х	0	0	Х	Х	0	0	Х	Х	0	0	Х	Х
Median (+ -) Mean (+ -)	33%	22%	5 -8 6 -8	9 -11 12 -11	0%	44%	4 -5 9 -5	17 -31 17 -40	0%	66%	12 -7 12 -7	10 -8 11 -8	0%	88%	10 X 10 X	X -2 X -2

 Table 18: Quality criteria for mean rainfalls, one-hours time-step and 24-hours lead-time models. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.



Figure 14: Turbidity forecasts with a lead-time of 24h using one-hour time-step and mean rainfall. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

5.3.2. Spatialized rainfall

Comparing designed models with those obtained using mean rainfalls (Table 16), one can note that the numbers of neurons in different layers are the same (Table 19). Regarding rainfalls, the window widths are larger, increasing thus the complexity of the models as shown by the increasing number of parameters. For ETP and the order, windows are quite similar. An illustration is provided in Fig 15.

		ML	Р	Two bra	anches
F	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent
	Rainfall layer	Х	Х	10	15
Hidden layers	Evapotranspiration layer	Х	Х	1	1
	Output layer	5	5	10	15
	Past Rainfall (*6)	42	13	21	11
T / ' 1	Futur Rainfall	24	24	24	24
Input windows	Evapotranspiration	3	3	3	3
	Order	1	3	10	1
Numb	er of parameters	1411	551	1745	1655

Table 19: Synthesis of the selected architectures for spatialized rainfalls, one-hours time-step and 24-hours lead-time models.

Peak prediction

Looking at the quality of models designed using spatialized rainfalls, one can compare Table 20 to Table 17 (mean rainfalls). It appears that in this configuration the spatialized information allows to reach very good maxima of turbidity, better than using mean rainfalls. Delays are also significantly diminished but they remain high diminishing thus the anticipation.

Table 20: Quality criteria for spatialized rainfalls, one-hours time-step and 24-hours lead-time models. For each event, the best criterion is underlined. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions

				men premenon	54			
		М	LP			Two-b	ranches	
	Feed-	forward	Rec	urrent	Feed-	forward	Recu	urrent
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)
T ₂	<u>87</u>	<u>4</u>	51	22,5	72	22,5	70	9,5
Т3	76	8,5	118	<u>1,5</u>	<u>93</u>	2	134	<u>1,5</u>
T ₆	<u>110</u>	9	75	<u>1</u>	182	8		
T ₇	57	13	<u>127</u>	<u>2</u>	Sto	p set	188	4,5
T 10	43	<u>1</u>	Sto	op set	<u>61</u>	3,5	54	7,5
T ₁₁	52	<u>0</u>	47	2	<u>59</u>	1	41	1
T ₁₆	54	1,5	38	11	199	<u>1</u>	Sto	p set
T 17	Sto	p set	<u>99</u>	<u>1</u>	105	8	<u>99</u>	2
T ₁₈	<u>101</u>	<u>0</u>	141	5	122	<u>0</u>	179	<u>0</u>
T ₂₂	<u>164</u>	<u>1</u>	253	<u>1</u>	269	<u>1</u>	306	6,5
Median	76	<u>1,5</u>	99	2	105	2	<u>99</u>	2
Mean	83	4,2	<u>105</u>	5,2	129	5,2	119	<u>3,6</u>

Threshold overtaking

Regarding threshold overtaking, it appears by comparing Table 21 to Table 18 (mean rainfalls) that there is more errors in threshold detection, especially at the beginning of the event (Fig. 15).

Table 21: Quality criteria for spatialized rainfalls, one-hours time-step and 24-hours lead-time models. Results of events without
turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

_				M	LP						-	Гwo bra	anches			
		Feed-	forward	1		Rec	urrent			Feed	-forward			Recu	urrent	
Event	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{ m p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)
T ₂	0	0	8	8	1	0	15	26	0	0	9	-4	1	0	22	18
T3	1	0	-4	7	1	1	Х	Х	0	0	-6	13	1	0	5	0
T ₆	1	0	-5	-7	0	1	Х	Х	0	0	-4	-12	1	0	-4	9
T 7	0	0	Х	Х	1	0	Х	Х		St	op set		2	0	Х	Х
T10	0	0	-9	5		Sto	op set		0	0	-7	4	0	0	12	-4
T ₁₁	0	1	-25	-7	3	0	23/3	0,25	0	0	-4/-13	8/-6	0	2	Х	Х
T ₁₆	0	0	-1	-16	0	0	9	-8	0	0	3	-16		Sto	p set	
T ₁₇		Sto	op set		0	3	Х	Х	1	2	-8	-10	3	1	5	-10
T ₁₈	1	0	Х	Х	2	0	Х	Х	1	0	Х	Х	6	0	Х	Х
T ₂₂	2	0	Х	Х	3	0	Х	Х	2	0	Х	Х	4	0	Х	Х
Median (+ -) Mean (+ -)	55%	11%	6 -5 6 -9	7 -7 7 -10	122%	5%	12 X 13 X	26 -4 26 -4	44%	22%	6 -7 6 -7	8 -10 8 -10	200%	33%	9 -4 11 -4	9 -7 9 -7



Figure 15: Turbidity forecasts with a lead-time of 24h using one-hour time-step and spatialized rainfall. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

5.4. Six-hours time-step and 24-hours lead-time models

5.4.1. Mean rainfall

Synthesis or selected architectures are presented in Table 22. Comparing Table 22 with Table 16 (1-hour time step and 24h lead-time), it is possible to note that only the MLP model is more complex (more neurons) others are less complex. Globally the number of parameters is lower with 6 hours time-steps, that was the goal: diminish the complexity to improve the generalization. Illustration is provided in Fig. 16.

	A	ML	P	Two bra	anches
I	Architecture	Feed-forward	Recurrent	Feed-forward	Recurrent
	Rainfall layer	Х	Х	11	8
Hidden layers	Evapotranspiration layer	Х	Х	1	1
	Output layer	22	9	6	11
	Rainfall	8	5	6	5
Input windows	Evapotranspiration	3	3	3	3
	Order	4	8	3	1
Numb	per of parameters	375	163	202	295

Table 22. Synthesis of the selected architectures for six-hours time-step, 24-hours lead-time models and mean rainfalls.

Peak prediction

Table 23 presents the quality criteria. Comparing Table 23 to Table 5 (1h time-step, 12h lead-time), it appears that the results are significantly worse in Table 23: diminishing the complexity of the model didn't balance the extension of the lead-time. Comparing Table 23 to Table 11 (6h time-step, 12h lead-time), one can note that the results are quite similar, it is noticeable that the extension of the lead-time has a little decreased the results. Comparing the performance between Table 23 and Table 17 (1-hour time step, 24h lead-time), it appears that both P_p and P_d are significantly better for the two-branches models. Regarding Fig. 16c, it appears that prediction of the FF-MLP is significantly better than the equivalent prediction of Fig. 14c. by re-sampling at 6h-time step.

		M	LP		Two-branches						
	Feed-f	forward	Rec	urrent	Feed-f	orward	Recurrent				
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)			
T ₂	61	18	39	24	<u>71</u>	<u>6</u>	40	18			
T ₃	<u>91</u>	<u>0</u>	78	<u>0</u>	128	18	79	<u>0</u>			
T ₆	152	<u>18</u>	85	<u>18</u>	213	<u>18</u>	<u>93</u>	<u>18</u>			
T 7	65	<u>6</u>	52	36	Sto	p set	<u>73</u>	<u>6</u>			
T10	42	18	Sto	op set	<u>59</u>	<u>12</u>	47	<u>12</u>			
T11	44	36	32	<u>6</u>	<u>50</u>	<u>6</u>	32	12			
T ₁₆	<u>52</u>	36	47 12		184 <u>6</u>		Stop set				
T ₁₇	Sto	p set	52	<u>18</u>	<u>83</u>	24	51	<u>18</u>			
T ₁₈	<u>79</u>	<u>18</u>	72	30	<u>79</u>	24	71	30			
T ₂₂	154	18	156	6	168	10	<u>156</u>	<u>0</u>			
Median	65	18	52	18	<u>83</u>	<u>12</u>	71	12			
Mean	82	18	68	17	<u>115</u>	14	71	13			

 Table 23. Synthesis of the selected architectures for six-hours time-step, 24-hours lead-time and mean rainfalls. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.

Threshold overtaking

Table 24 presents the quality criteria. Comparing Table 24 with Table 6 (1-hour time-step and 12-hours leadtime), we can note that the results are similar, meaning that the prediction 24h-ahead is quite equivalent to the prediction 12h again. This result could be very useful for operational applications. Comparing Table 24 with Table 12 (6h time-step, 12h lead-time), results are similar: extending the lead time has no clear influence on threshold overtaking. Finally, comparing Table 24 to Table 18 (1-hour time step, 24h lead-time) it appeared that delays are increased but there is less false negative. Also for threshold overtaking, increasing the time-step improves the generalization.

	MLP								Two branches							
	Feed-forward			Recurrent			Feed-forward				Recurrent					
Event	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)
T ₂	0	0	-6	0	0	0	0	-54	0	0	-12	-6	0	0	0	24
T 3	1	0	6	-12	1	0	6	-12	0	1	6	12	0	0	6	-12
T ₆	0	0	12	12	0	0	6	-12	0	0	0	6	1	0	6	-12
T 7	0	0	Х	Х	0	0	Х	Х		Sto	op set		0	0	Х	Х
T ₁₀	1	0	6	-6	Stop set			1	0	6	0	0	0	0	24	
T ₁₁	0	1	18/6	6	0	1	18	-24	0	0	6/12	6/-6	0	1	24	-18
T 16	0	0	0	18	0	0	0	18	0	0	0	18		Ste	op set	
T17		St	op set		0	3	Х	Х	0	2	18	18	0	2	18	6
T ₁₈	0	0	Х	Х	0	0	Х	Х	1	0	Х	Х	0	0	х	Х
T ₂₂	2	0	Х	Х	0	0	Х	Х	2	0	Х	Х	0	0	Х	Х
Median (+ -) Mean (+ -)	44%	11%	6 -6 8 -6	9 -9 9 -9	11%	44%	6 X 6 X	18 -18 18 -26	44%	33%	6 -12 7 -12	9 -6 10 -6	11%	33%	6 X 9 X	24 -12 18 -14

Table 24. Quality criteria for mean rainfalls, six-hours time-step and 24-hours lead-time models. Results of events without turbidity
peaks (see Table 2) are not presented because there is no false positive in their predictions.



143
5.4.2. Spatialized rainfall

Four architectures were designed and summarized in Table 25. Comparing Table 25 to Table 22 (mean rainfalls), it appears that the lengths of input windows are increased in Table 25 for recurrent models. Comparing with Table 19: re-sampling the number of parameters diminish them using 6 hours time-steps. An illustration is provided in Fig. 17.

	_	Ν	/ILP	Two ł	oranches
	Architecture	Feed- forward	Recurrent	Feed- forward	Recurrent
	Rainfall layer	Х	Х	11	7
Hidden layers	Evapotranspiration layer	Х	Х	1	1
	Output layer	9	12	7	15
	Past Rainfall (*6)	1	6	5	6
Input	Futur Rainfall	4	4	4	4
windows	Evapotranspiration	10	3	3	3
	Order	4	8	3	1
Num	ber of parameters	235	637	509	457

Table 25. Synthesis of the selected architectures for spatialized rainfalls, six-hour time-step and 24-hours lead-time models.

Peak prediction

Table 26, similarly to Table 23 (mean rains), presents the obtained quality criteria for each event in cross-test situation. Except for the feed-forward two-branches model whose median have a perfect prediction of the peak, all other results are worse, especially the time-delays that are very bad.

Table 26: Quality criteria for spatialized rainfalls, six-hours time-step and 24-hours lead-time models. For each event, the best criterion is underlined.

		I	MLP		Two-branches					
	Feed-f	forward	Rec	urrent	Feed-t	forward	Recurrent			
Event	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)	P_P	P_d (h)		
T2	56	24	41	12	<u>107</u>	<u>36</u>	38	36		
T ₃	<u>93</u>	<u>12</u>	65	<u>6</u>	68	18	69	<u>18</u>		
T ₆	59	<u>0</u>	73	<u>6</u>	124	<u>81</u>	<u>70</u>	<u>12</u>		
T 7	58	<u>24</u>	52	36	Sto	p set	<u>35</u>	<u>12</u>		
T10	48	18	Sto	p set	<u>65</u>	<u>24</u>	60	<u>18</u>		
T11	46	24	32	<u>6</u>	<u>60</u>	<u>19</u>	33	18		
T ₁₆	<u>54</u>	24	39	18	181	<u>12</u>	Sto	p set		
T17	Sto	p set	54	<u>6</u>	<u>100</u>	18	60	<u>12</u>		
T ₁₈	<u>84</u>	<u>24</u>	86	66	<u>77</u>	30	60	12		
T ₂₂	213	6	120	12	358	12	<u>136</u>	<u>18</u>		
Median	58	24	54	12	100	<u>18</u>	60	18		
Mean	79	17	62	19	127	<u>20</u>	62	18		

Threshold overtaking

Quality criteria are presented in Table 27. Comparing Table 27 and Table 24 (mean rainfalls), it appears that behaviours are similar. Comparing Table 27 and Table 15 (six-hours time-step and 12-hours lead-time) quality is also similar. This is very interesting for operational needs.

									T							
				М	LP							Two b	ranche	s		
-		Feed	-forward	1		R	ecurrent			Feed-	forward	1		Re	current	
Event	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	Fp	Fn	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)	$F_{\rm p}$	F_{n}	$d_{\rm r}$ (h)	$d_{\rm d}$ (h)
T ₂	0	0	-6	0	0	0	6	-24	0	0	-12	-6	0	0	0	-24
T 3	1	0	6	-12	0	1	Х	Х	0	0	6	-12	0	0	5	-12
T ₆	0	0	12	12	0	0	6	0	0	0	0	6	1	0	5	-12
T 7	0	0	Х	Х	0	0	X	Х		Sto	op set		0	0	Х	Х
T ₁₀	0	0	6	-6		S	Stop set		0	0	6	6	0	0	0	24
T ₁₁	0	1	18	6	0	0	0/18	6/30	0	0	-0,5	1	0	1	24	-18
T ₁₆	0	0	0	18	0	0	6	-6	0	0	0	18		St	op set	_
T17		St	op set		0	2	-6	6	0	2	18	18	0	2	18	6
T18	0	0	Х	Х	0	0	Х	Х	1	0	Х	Х	0	0	Х	Х
T ₂₂	2	0	Х	Х	0	0	Х	Х	2	0	Х	Х	0	0	Х	Х
Median (+ -) Mean (+ -)	33%	11%	6 -6 8 -6	9 -9 9 -9	0%	33%	6 -6 7 -6	6 -15 11 -15	33%	22%	6 -9 7 -9	12 -6 12 -8	11%	33%	5 X 9 X	15 -15 15 -17

 Table 27: Quality criteria for spatialized rainfalls, six-hours time-step and 24-hours lead-time models. Results of events without turbidity peaks (see Table 2) are not presented because there is no false positive in their predictions.



Figure 17: Turbidity forecasts with a lead-time of 24h using six-hour time-step and spatialized rainfall. Turbidity (left scale), and rainfalls (right scale). Measured turbidity (solid line), Maximum₁₀ turbidity (doted line) and Median turbidity (dashed line). Test on the event 2. Uncertainty is shown in grey area. a) Feed-forward multilayer perceptron model, b) Recurrent multilayer perceptron model, c) Feed-forward two-branches model and d) Recurrent two-branches model.

6. Discussion

In order to synthetize the results extensively proposed in the previous section, we propose to first discuss the quality of the proposed predictions in regarding to 5 subjects: uncertainty, dynamics, best lead-time, best synchronous peak estimation and best threshold overtaking based on the synthesis of Table 28. Second, we will propose a more basic discussion regarding the general behaviour of the designed models.

Based on Fig.10 to Fig.17, visual estimation of these quality criteria is done and synthetized in Table 28.

Table	Table 28: Synthetic appreciation of the quality of models. U, D, P _d , P _p and TO mean respectively uncertainty, dynamics, peak delay, Percentage of peak, threshold overtaking. 2-BR and MLP means respectively: 2-branches or multilayer perceptron, <i>n</i> means neutral (neither positive nor negative)																															
		1h sa	mplin	g rate:	; 12h]	lead-	time			6h sa	amplir	ng rat	e; 12h	lead-	time			1h s	sampli	ing rat	e; 24h	lead-	time			6h sampling rate; 24h lead-time						
	М	lean R Fig.	ainfal	ls;		Spati Rain Fig	alized falls;		Mean Rainfalls; Fig. 12.			Spati	Spatialized Rainfalls;		M	Mean Rainfalls;		Spatialized Rainfalls; Fig. 15.		15.	М	Mean Rainfalls;		ls;	Spatialized Rainfalls; Fig. 17.		ılized falls; 17.					
	М	LP	2-]	BR	MI	LP	2-H	BR	М	ILP	2-F	BR	М	LP	2-]	BR	М	LP	2-	BR	M	LP	2-]	BR	M	LP	2-J	BR	M	LP	2 I	3R
	-a- FF	-b- R	-c- FF	-d- R	-a- FF	- b- R	-c- FF	- d- R	-a- FF	-b- R	-c- FF	-d- R	-a- FF	-b- R	-c- FF	-d- R	-a- FF	-b- R	-c- FF	-d- R	-a- FF	-b- R	-c- FF	-d- R	-a- FF	-b- R	-c- FF	-d- R	-a- FF	-b- R	-c- FF	-d- R
U	+	+	-	+		n	-	+	n	+++		+	++	++		n	+	+	-	+	n		-		n	++		+	n	++ +		++
D	-	+	-		n	-	-	-	+		n	•	+	n	-	n	+	-	+	-	-		-		-		+		-	n	-	-
P _d (h) (best)			4		13				18							18				6	2							12		12		
P _p (best)			106		95				88							74		65				99					83				100	
TO (Best)	11 %	77 %	88 %	55 %	55 %	88 %	55 %	77 %	11 %	44%	44 %	55 %	0%	66%	77 %	88 %	55 %	44 %	66 %	88%	66 %	127 %	66 %	233 %	55 %	55 %	77 %	44 %	44 %	33 %	55 %	44 %

6.1. Quality of predictions

6.1.1. Uncertainty

Uncertainty was previously defined (beginning of section 5); it is estimated in Table 28 by qualitative measurements from "- -"to "++". In a first approach one can consider that the best uncertainty is the lower. Following the Table 28, the lower uncertainty is found in Fig. 12-b and Fig. 17-b, both related to the recurrent MLP, and to the 6-hours time-step. Unhopefully it appears also that this property is more frequently associated to a very weak prediction of the peak of turbidity. Our interpretation is that the uncertainty is diminished by the 6-h sampling rate that regularizes the precipitations. Contrarily it seems that the worse uncertainty (the wider) could frequently be associated to the best performances measured by the peak-percentage (Fig. 10-c, Fig. 11-a, Fig. 15-b, Fig. 16-c, Fig. 17-c). This could be explained because the predicted output is the Maximum₁₀ of the ensemble: more dispersion stimulates the existence of high values, giving a best prediction of the maxima.

The worse uncertainty is also frequently correlated to the models -c-: the feed-forward two-branches models, when the best uncertainty is associated to the models -b: recurrent multilayer perceptron: the "opposite" models provide thus "opposite" uncertainty. This observation is an indication of self-coherence of results.

6.1.2. Dynamics

The dynamics is the ability to well follow the variations of the signal. It is estimated in the Table 28 by qualitative measurements from "- -"to "+". Actually, there is not output with a very good dynamic. The best ones are provided by Fig. 10-b, Fig. 12-a, Fig. 13-a, Fig. 14-a, Fig. 14-c and Fig 16-c. Except in Fig. 10-b, all of these models are feed-forward. This is a surprising observation as usually, in rainfall-runoff modeling, recurrent models have a good dynamic. The principal fault is that the rising predicted turbidity is mot synchronous with the observed one: it begins with the beginning of rains when actual turbidity begins 15 hours later for the event 2. This point suggests that rainfalls should be introduced in a different way in the model.

The worse dynamics ("--") are observed in Fig. 10-d, Fig. 12-b, Fig. 15-b, Fig. 15-d, Fig. 16-b and Fig 16-d. Models -b and -d correspond also to recurrent models, reinforcing the previous surprise. Only the model represented in Fig. 16-d provides a good operational property as it provides the best Threshold Overtaking, as shown in the last line of Table 28. Except this model, no other model provides good operational properties. This underlines the crucial role of the dynamics. In contrast to the previous point, models with good dynamics don't necessarily give good operational properties.

One can synthetize these observations saying that it is necessary to not miss the dynamics to have good operational features.

6.1.3. Peak delay

Going into the Table 28 by operational needs: 3 last lines, we begin by the peak-delay. One can note that best peak-delays are not provided by a specific architecture: neither spatialized, neither mean rainfalls. The principal point is the sampling period: the greater it is (Fig. 12, 13, 16 and 17), the longer are the delays. Considering this criterion, it is thus necessary to choose the lower sampling rate.

6.1.4. Percentage of peak

Regarding the percentage of peak, and also looking at the Table 28, the P_p is the best for architectures: (10)c, (11)a, (12)a, (13)d, (14)b, (15)b, (16)c, (17)c. The kind of architecture promoting the quality of peak (a, b, c or d) thus is not clear. Nevertheless, one can note that the prediction of the maximum value is better for mean rainfalls except for 24-h lead-time. If we consider both: maximum value and delay, the best couples happen together for 12h lead-time, and separately for 24h lead-time, meaning that both are more difficult to reach in this case. No influence of sampling rate can be identified.

6.1.5. Threshold overtaking

Considering the second kind of operational need: threshold overtaking, it is reported in the table based on the addition of both false positive and false negative predictions. Percentages can be superior to 100% because several events have more than one peak; it is also possible to make several false positive for one actual peak. The best model is identified in the last line of Table 28 by a white background. Feed-forward MLP dominates clearly for 12h lead-time, when no kind or architecture clearly dominates for 24h lead-time. Calculating average on all the models, the feed-forward MLP is the best with 37% while architectures b, c, d reach respectively 67%, 66% and 86%.

6.2. Basic analysis

6.2.1. Sampling

Choose the best sampling period is a crucial question because it determines the quality of the results. Sampling too small lead to increase the number of parameters then inducing overfitting. Sampling too large prevent catching the good dynamics of phenomena. This fundamental question is rarely addressed in the literature in hydrology. Famous tools exist to mathematically address this question: the Fourier transform and the Shannon-Nyquist theorem, nevertheless the impact on bias-variance trade-off has no mathematical solution. For this reason we tried in this study to see if the augmentation of the sampling period, allowing diminishing the number of parameters (the complexity), allows to increase the performance. The response is not trivial: for 1h time-step, the ratio: complexity of phenomena/number of parameters is good for 12h lead-time, but not optimal for 24h lead-time. For 6h time steps, catch phenomena are not the same and the ratio: complexity of phenomena/number of parameters is better using 24h lead-time.

We can thus remember the hypothesis saying that having a longer time-step should allow predicting with a longer lead-time.

6.2.1. Anticipation

Finally, regarding the anticipation, one can distinguish 2 goals: the best prediction of level (max of turbidity) at the right time, and the best threshold overtaking. At short lead-time the feed-forward models seem better, as the measured turbidity adds great information. At long lead-time (24h) after resampling to 6h time step, the feed-forward 2-branches models (16-c and 17-c) work better than the others, and the resampling allows to take profit of the spatialized rainfalls. This is noticeable comparing Fig. 17-c and Fig. 10-c: both are quite similar but the 17-c predicts at 24h-lead time. Moreover, calculating the average of false positive and false negative reported in Table 28, the value for models of Fig. 10 and models of Fig. 17 are respectively 58% and 44% with, in the order of Figures numbering, respectively (69%, 39%, 58%, 63%, 123%, 58%) for the other models.

It appears then profitable to extend the lead-time thanks the resampling to 6h period for both goals.

7. Conclusion

Turbidity in karst waters results from complex phenomena that are difficult to model using physical assumptions. To the best of our knowledge, few works investigated the possibility to predict the turbidity up to several time-steps, and no work examine the prediction of rainfall-turbidity relation at small temporal scales, which necessitates to take into account nonlinear processes from different origins: exogenous and endogen turbidity. In addition to a work previously done at 12h lead-time, where we showed the ability of neural networks to predict the turbidity and the threshold overtaking, we wanted to investigate in the present paper the ability to reliably predict the turbidity and the threshold overtaking up to 24h lead-time by resampling signals and possibly by using spatialized rainfalls.

After having presented the state of the art about turbidity modelling and forecasting, and having presented the case study and data of the Yport pumping well for tap water production, we have presented neural network modelling and the fundamental trap of machine learning: bias-variance trade-off. Then, we designed a wide range of models thanks to rigorous variable and complexity selection using cross validation. 32 models were thus trained and their properties were described and appraised in regards to both goals: maximum of turbidity and 100 NTU

threshold overtaking predictions. Moreover a more basic analysis is provided to appreciate the gain provided by resampling and by introducing spatialized rainfalls compared to the initial work. Our interpretation is that the complexity of the model (the number of free parameters) plays the role of a control parameter: if it is too high the model is unable to generalize; if it is convenient, we can use this "volume" of parameters to predict with a short time-step at a moderate lead-time, or to predict using a longer time-step at a longer lead-time. Finally using the same "volume" of parameters, it is also possible to predict using a longer time-step thanks to spatialized rainfalls.

All these considerations allow designing models able, with limited uncertainty, to predict the threshold overtaking and maximum of turbidity peak up to 24h ahead. Moreover several trails were proposed as, for example, improve the representation of rainfalls.

Acknowlegments

The authors would like to thank the CODAH for providing rainfall and turbidity data. The Normandie Region and Seine-Normandie Water Agency are thanked for the co-funding of the study. We are also very grateful to S. Lemarie and J. Ratiarson of for the very helpful discussions they helped organize. Our thanks are extended to D. Bertin for his extremely fruitful collaboration in the design and implementation of the Neural Network simulation tool: RnfPro.

BIBLIOGRAPHY

Artigue, G., Johannet, A., Borrell, V., & Pistre, S. (2012). Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin (southern France). *Natural Hazards and Earth System Sciences*, 12(11), 3307-3324.

Atteia, O., & Kozel, R. (1997). Particle size distributions in waters from a karstic aquifer: from particles to colloids. *Journal of Hydrology*, 201(1-4), 102-119.

Auckenthaler, A., Raso, G., & Huggenberger, P. (2002). Particle transport in a karst aquifer: natural and artificial tracer experiments with bacteria, bacteriophages and microspheres. *Water Science and Technology*, 46(3), 131-138.

Beaudeau, P., Leboulanger, T., Lacroix, M., Hanneton, S., & Wang, H. Q. (2001). Forecasting of turbid floods in a coastal, chalk karstic drain using an artificial neural network. *Ground Water*, 39(1), 109-118.

Bin Omar, A. F., & Bin MatJafri, M. Z. (2009). Turbidimeter design and analysis: a review on optical fiber sensors for the measurement of water turbidity. *Sensors*, 9(10), 8311-8335.

Dreyfus, G. (2005). Neural networks: methodology and applications. *Springer Science & Business Media*, 497.

Dreyfus, G., Samuelides, M., Martinez, J.-M., Gordon, M.-B., Badran, F., Thiria, S., Hérault, L. (2008). Apprentissage statistique. *Eyrolles*. 450

Dussart-Baptista, L., Massei, N., Dupont, J. P., & Jouenne, T. (2003). Transfer of bacteria-contaminated particles in a karst aquifer: evolution of contaminated materials from a sinkhole to a spring. *Journal of Hydrology*, 284(1-4), 285-295.

Darras, T., Johannet, A., Vayssade, B., Kong-A-Siou, L., & Pistre, S. (2014). Influence of the initialization of multilayer perceptron for flash flood forecasting: design of a robust model. *Proceedings ITISE*, 1068-1080.

Darras, T., Estupina, V. B., Kong-A-Siou, L., Vayssade, B., Johannet, A., & Pistre, S. (2015). Identification of spatial and temporal contributions of rainfalls to flash floods using neural network modelling: case study on the Lez basin (southern France). *Hydrology and Earth System Sciences*, 19(10), 4397.

Gaillard, T., Hauchard, E., & Roux, J. C. (2012) Les Fontaines d'Yport (Seine-Maritime), émergences majeures du littoral normand crayeux: Exploitation et vulnérabilité de la ressource en eau. *Ressources et gestion des aquifères littoraux. Cassis 2012.* 235-244.

Geman, S., Bienenstock, E., Doursat, R., 1992. Neural networks and the bias/variance dilemma. *Neural computation* 4, 1–58.

Hardenbergh, W. A. (1938). Water Supply and Purification, Internat. Textbook Co., Scranton, Pa, 280.

Hauchard, E., & Laignel, B. (2008). Morphotectonic evolution of the north-western margin of the Paris Basin. *Zeitschrift für Geomorphologie*, 52(4), 463-488.

Hauchard, E., Laignel, B., & Delahaye, D. (2002). Proposition d'un nouveau schéma structural du Nord-Ouest du bassin de Paris reposant sur l'analyse fractale des réseaux de thalwegs et les données récentes de la géologie régionale. *Comptes Rendus Geoscience*, 334(4), 295-302.

Hole, J. P., Roux. J. C., 1978. Vulnérabilité aux pollutions du bassin hydrogéologique des sources d'Yport (Seine-Maritime), *Rapport BRGM*. 81.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5), 359-366.

Houria, B., Mahdi, K., & Zohra, T. F. (2014). PSO-ANNs based suspended sediment concentration in Ksob basin, Algeria. *Journal of Engineering and Technology Research*, 6(8), 129-136.

Iglesias, C., Torres, J. M., Nieto, P. G., Fernández, J. A., Muñiz, C. D., Piñeiro, J. I., & Taboada, J. (2014). Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain. *Water resources management*, 28(2), 319-331.

Johannet, A., Vayssade, B., & Bertin, D. (2008). Neural networks: from black box towards transparent box. Application to evapotranspiration modeling. *International Journal of Computational Intelligence*, 4(3), 163-170.

Kisi, O., Dailr, A. H., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology*, 450, 48-58.

Kneller, B., & Buckee, C. (2000). The structure and fluid mechanics of turbidity currents: a review of some recent studies and their geological implications. *Sedimentology*, 47(s1), 62-94.

Kong-A-Siou, L., Johannet, A., Estupina, V. B., & Pistre, S. (2015). Neural networks for karst groundwater management: case of the Lez spring (Southern France). *Environmental Earth Sciences*, 74(12), 7617-7632.

Kong-A-Siou, L., Fleury, P., Johannet, A., Estupina, V. B., Pistre, S., & Dörfliger, N. (2014). Performance and complementarity of two systemic models (reservoir and neural networks) used to simulate spring discharge and piezometry for a karst aquifer. *Journal of Hydrology*, *519*, 3178-3192.

Kong-A-Siou, L., Johannet, A., Borrell V., Pistre, S., 2012. Optimization of the generalization capability for rainfall–runoff modeling by neural networks: the case of the Lez aquifer (southern France). *Environ Earth Sci 65*, 2365–2375.

Kong-A-Siou, L., Johannet, A., Borrell, V., Pistre, S., 2011. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *Journal of Hydrology 403*, 367–380.

Kothyari, U. C., & Jain, S. K. (1997). Sediment yield estimation using GIS. *Hydrological sciences journal*, 42(6), 833-843.

Laignel B., (1997). Les altérites à silex de l'Ouest du Bassin de Paris : caractérisation lithologique, genèse et utilisation potentielle comme granulats. *Thèse de Doctorat*. Université de Rouen. 219p.

Laignel B., (2003). Caractérisation et dynamique érosive de systèmes géomorphologiques continentaux et crayeux. Exemple de l'Ouest du Bassin de Paris dans le contexte nord-ouesteuropéen. *Mémoire de HDR*. Université de Rouen. 143p.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2), 164-168.

Mahler, B. J., Personné, J. C., Lods, G. F., & Drogue, C. (2000). Transport of free and particulate-associated bacteria in karst. *Journal of Hydrology*, 238(3-4), 179-193.

Mangin, A. (1975). Contribution à l'étude hydrodynamique des aquifères karstiques (*Doctoral dissertation*, Université de Dijon).

Mangin, A. (1981). Apport des analyses correlátoires et spectrales croisées dans la connaissance des systemes hydrologiques. *CR Acad.* Sci. Paris, 293(2), 1011-1014.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2), 431-441.

Massei, N., Wang, H. Q., Dupont, J. P., Rodet, J., & Laignel, B. (2003). Assessment of direct transfer and resuspension of particles during turbid floods at a karstic spring. *Journal of hydrology*, 275(1-2), 109-121.

Mulia, I. E., Tay, H., Roopsekhar, K., & Tkalich, P. (2013). Hybrid ANN–GA model for predicting turbidity and chlorophyll-a concentrations. *Journal of Hydro-environment Research*, 7(4), 279-299.

Nash, J.E. and Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I – A discussion of principles, *Journal of hydrology*, 10(3), 282-290

Nebbache, S. L. M. A., Loquet, M., Vinceslas-Akpa, M., & Feeny, V. (1997). Turbidité et micro-organismes dans une eau de source en milieu karstique. *European journal of soil biology*, 33(2), 89-103.

Nerrand, O., Roussel-Ragot, P., Personnaz, L., Dreyfus, G., Marcos, S., 1993. Neural networks and nonlinear adaptive filtering: Unifying concepts and new algorithms. *Neural Computation* 5, 165–199.

Nieto, P. G., García-Gonzalo, E., Fernández, J. A., & Muñiz, C. D. (2014). Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: A case study in Northern Spain. *Ecological Engineering*, 73, 192-200.

Oudin, L., Michel, C., & Anctil, F. (2005a). Which potential evapotranspiration input for a lumped rainfall-runoff model? : Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? *Journal of Hydrology*, 303(1-4), 275-289.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005b). Which potential evapotranspiration input for a lumped rainfall–runoff model? : Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, 303(1-4), 290-306.

Pronk, M., Goldscheider, N., & Zopfi, J. (2007). Particle-size distribution as indicator for fecal bacteria contamination of drinking water from karst springs. *Environmental Science & Technology*, 41(24), 8400-8405.

Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M., & Nourani, V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the total environment*, 407(17), 4916-4927.

Ryan, M., & Meiman, J. (1996). An Examination of Short- term Variations in Water Quality at a Karst Spring in Kentucky. *Groundwater*, 34(1), 23-30.

SAFEGE (2012) Étude du Bassin d'Alimentation du forage d'Yport "Phase 1 :Délimitation du bassin d'alimentation et analyse de la vulnérabilité intrinsèque du basin". 112.

Savary, M., Johannet, A., Massei, N., Dupont, J. P., & Hauchard, E. (2016, June). Operational Turbidity Forecast Using Both Recurrent and Feed-Forward Based Multilayer Perceptrons. In *International Work-Conference on Time Series Analysis* (pp. 243-256). Springer, Cham.

Shevenell, L., & McCarthy, J. F. (2002). Effects of precipitation events on colloids in a karst aquifer. *Journal of Hydrology*, 255(1-4), 50-68.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society*. Series B (Methodological), 111-147.

Toukourou, M., Johannet, A., Dreyfus, G., & Ayral, P. A. (2011). Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of "Cévenol flash floods". *Applied Intelligence*, 35(2), 178-189.

Towler, E., Rajagopalan, B., Summers, R. S., & Yates, D. (2010). An approach for probabilistic forecasting of seasonal turbidity threshold exceedance. *Water Resources Research*, 46(6).

Tramblay, Y., Saint-Hilaire, A., & Ouarda, T.B. (2007). Modelling extreme suspended sediment concentrations in North America: frequency analysis and correlations with watershed characteristics. *Water Quality and Sediment Behaviour of the Future: Predictions for the 21st Century*, (314), 20.

Wicks, J. M., & Bathurst, J. C. (1996). SHESED: a physically based, distributed erosion and sediment yield component for the SHE hydrological modelling system. *Journal of Hydrology*, 175(1-4), 213-238.

5.3 COMPLEMENTS SUR L'ARTICLE

5.3.1 FONCTION DE COUT ET MINIMUMS LOCAUX

INITIALISATION ALEATOIRE DES PARAMETRES

Le but de l'apprentissage des réseaux de neurones est le calcul des paramètres du modèle dans le but d'atteindre le minimum de la fonction de coût. Cependant, comme illustré sur la Figure 54, la fonction de coût peut présenter plusieurs minima locaux dans lesquels l'apprentissage peut se bloquer et ainsi empêcher le modèle d'atteindre le minimum global qui correspond à la meilleure optimisation des paramètres. Ainsi, comme illustré sur la Figure 54, le modèle, selon son initialisation, peut ne pas atteindre le minimum global de la fonction de coût.



Figure 54: Illustration de la fonction de coût et des minimums locaux

Ainsi il apparaît que les réseaux de neurones, en fonction de la complexité de l'hypersurface représentant la fonction de coût dans l'hyperespace des paramètres du modèle (de dimension souvent supérieure à 100), peuvent être extrêmement sensibles à l'initialisation du modèle. Pour diminuer les différences dues à cette initialisation, Darras, T. (2015), a proposé de réaliser un modèle d'ensemble rendant la prévision bien plus robuste à l'initialisation.

MODELE D'ENSEMBLE

Dans ses travaux de thèse, Darras, T. (2015) a proposé une méthode permettant de déterminer le nombre de membres dans le modèle d'ensemble, nécessaires pour obtenir une modélisation dont le résultat ne dépend que très peu de l'initialisation.

Nous avons appliqué la méthode proposée par Darras, T. & *al*, 2016 sur le perceptron multicouche dirigé effectuant la prédiction à 12h en utilisant des données horaires et des pluies moyennes.

Nous avons ainsi effectué 200 expériences (200 modèles différents par leur initialisation aléatoire). Sur ces 200 expériences, nous avons ensuite réalisé 1000 tirages aléatoires avec remise de *n* expériences afin de définir quel nombre n permet d'obtenir une prévision ne dépendant que très peu de l'initialisation. Nous avons calculé l'écart type des critères de Nash correspondant aux *n* résultats.

La Figure 55 présente les écart-types obtenus pour la modélisation des événements 16, 11 et 3 dans le cas où nous effectuons : 1, 5, 10, 20, 30, 40, 50 tirages aléatoires avec remise. On observe ainsi qu'il est nécessaire d'effectuer 10 expériences au minimum afin de diminuer significativement l'écart type des sorties de l'ensemble et ainsi limiter l'impact des minimums locaux de la fonction de coût.



Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance

Figure 55: Illustration de l'impact du nombre de modélisation sur l'écart type des résultats de modélisation.

Au-delà de 10 expériences, le ratio entre le nombre d'expériences (et donc le temps de calcul) et le gain de qualité (diminution de l'écart type) n'est pas intéressant.

C'est donc sur ces observations que nous avons fixé le nombre d'expériences effectuées à dix pour obtenir nos résultats de prévision.

5.4 CONCLUSION CHAPITRE 5

La turbidité dans les eaux karstiques résulte de phénomènes complexes qui sont difficiles à modéliser en utilisant une approximation des processus physiques. Au mieux de notre connaissance, peu de travaux se sont attachés à étudier la possibilité de prévoir la turbidité jusqu'à plusieurs pas de temps et aucun travail n'a effectué la prévision de la relation pluie/turbidité pour de petites échelles temporelles. Une telle prévision nécessite de prendre en compte des processus non-linéaires d'origines différentes : turbidité exogène et endogène. En plus d'un travail précédemment fait pour prévoir la turbidité à 12h où nous avons montré la capacité de réseaux neuraux pour prévoir un dépassement de seuil de 100 NTU, nous avons voulu examiner dans cette partie la capacité des modèles à prévoir la turbidité et le dépassement d'un seuil de 100 NTU jusqu'au 24h dans le futur en ré-échantillonnant les signaux mais également en utilisant des pluies spatialisés.

Nous avons ainsi, dans ce chapitre, conçu une vaste gamme de modèles grâce à un choix rigoureux des variables et de la complexité des modèles en utilisant la validation croisée. Ainsi, 32 modèles ont été formés et leurs propriétés ont été décrites et évaluées dans l'objectif de remplir deux but: (*i*) prédire le maximum de turbidité et (*ii*) prédire le dépassement d'un seuil de 100 NTU. Grâce à ces différents modèles et les nombreux critères dévaluations, nous avons montré que la complexité du modèle (le nombre de paramètres libres) peut être assimilé à un "paramètre de contrôle": (*I*) si la complexité est trop haute, le modèle ne peut pas généraliser, (*ii*) s'il n'est pas trop élevé, il est possible d'utiliser ce nombre de paramètres pour prévoir avec un pas de temps court pour un horizon de prévision modéré, ou prévoir l'utilisation d'un pas de temps plus long pour un horizon de prévision plus long. Pour finir, en utilisant le même nombre de paramètres, il est aussi possible de prévoir la turbidité avec un pas de temps plus long en utilisant des données de pluies spatialisées.

En prenant en compte toutes ces informations, il est donc possible de concevoir des modèles capables, avec une incertitude limitée, de prévoir le dépassement d'un seuil et le maximum d'un pic de turbidité jusqu'à 24h en avance. Afin d'améliorer plus encore les résultats des modèles, nous proposons la création d'un assemblage de modèles pour la prévision de la turbidité. Les 32 modèles que nous avons créés sont efficaces pour différents objectifs (synchronisation du pic, amplitude du pic, dépassement d'un seuil,...). En effectuant un modèle

d'assemblage, nous espérons améliorer la généralisation de la prévision. Dans cette voie d'amélioration, nous émettons également l'envie de procéder à la création de différents modèles spécifiques à l'amplitude des événements et d'effectuer un assemblage de ces derniers pour sélectionner le plus adapté à l'événement en cours de prévision. Dans un autre temps, nous avons proposés plusieurs voies d'améliorations telles que la meilleure représentation et prise en compte des pluies. En prenant en compte le besoin de prévision d'un dépassement de seuil de 100 NTU, nous proposons l'utilisation d'une variable à prédire simplifiée. En effet, il serait intéressant d'étudier le comportement et la qualité d'un modèle à réseaux de neurones pour la prévision d'une donnée de turbidité binaire. Nous entendons par binaire, un signal reconstitué pour lequel la valeur serait fixée à o lorsque la turbidité est strictement inférieure au seuil préalablement fixé puis une valeur de 1 lorsque le la turbidité est supérieure ou égale au seuil. Chapitre 5 : Prévision de la turbidité par réseaux de neurones 12h et 24h à l'avance

CHAPITRE 6 : COUPLAGE DE L'ANALYSE MULTIRESOLUTION ET DES RESEAUX DE NEURONES.

6.1 INTRODUCTION

La modélisation par apprentissage statistique à l'aide des réseaux de neurones artificiels est basée sur le contenu informatif des données. L'intérêt majeur de cette approche réside dans la représentation des relations non-linéaires des données liant les entrées aux sorties pour un système donné. De nombreuses techniques issues de la théorie de l'information sont disponibles pour déterminer ce contenu informatif par analyse du signal et traitement du signal. L'idée de coupler ces 2 types d'approches, traitement du signal et modélisation statistique, émerge ainsi logiquement. La question posée est la suivante : est-il possible, grâce à des prétraitements adaptés de simplifier l'analyse d'information faite par le modèle neuronal afin d'améliorer ses performances de prévision ?

Les débuts de l'utilisation de l'analyse multirésolution dans le domaine des géosciences peuvent être illustrés par les travaux de Foufoula-Georgiou, E. & al. (1995) visant à décrire l'application de ces méthodes en géophysique. Par la suite, des travaux d'application de ces méthodes en hydrologie ont de nombreuses fois montrés leur utilité. Ainsi nous citerons les travaux de Labat, D. sur l'application de l'analyse multirésolution à l'étude des débits de sources karstiques (Labat, D. & al., 1999a, Labat, D. & al., 1999b) et de rivières (Labat, D. 2010, Labat, D., 2008, Labat, D. & al., 2004) ou encore sur la relation pluie/ruissèlement (Labat, D. & al., 2000a, Labat, D. & al., 2000b, Labat, D. & al., 2001). Nous noterons également les travaux menés par Massei, N. qui utilise l'analyse multirésolution afin d'interpréter le fonctionnement des hydrosystèmes (Massei, N. & al. 2006). On peut également noter l'application de ces méthodes dans l'analyse de la vulnérabilité de l'eau face aux pollutions agricoles (Kang, S. & Lin, H. 2007). Des travaux portant sur l'analyse de séries temporelles annuelles montrent l'intérêt de l'utilisation de l'analyse multirésolution afin d'étudier l'impact du changement climatique sur les variables hydrologiques (Coulibaly, P. & Burn, D. H. 2004, Anctil, F. & Coulibaly, P. 2004, Coulibaly, P. & Burn, D. H. 2005, Kallache, M. & al., 2005, Labat, D. 2005, Nalley, D. & al. 2012, Massei, N. & al. 2017). Le lecteur intéressé pourra se tourner vers les publications de Torrence, C. & Compo, G. P. (1998) et de Sang, Y. F. (2013) qui effectuent une synthèse de l'application de la transformée en ondelettes pour l'analyse temporelle des signaux hydrologiques.

De nombreux travaux se sont attachés à intégrer l'analyse multirésolution dans les modèles hydrologiques. Nous pourrons ainsi retrouver des travaux visant à modéliser les débits d'une rivière à différents pas de temps : au pas de temps journalier (Rao, Y. S. & *al.* 2014), ou mensuel (Wang, W. & Ding, J. 2003, Kisi, O. & Cimen, M. 2011, Rathinasamy, M. & Khose, R. 2012) ou infra-journalier (Chou, C. 2007, Johannet, A. & *al.* 2012, Makwana, J. J. & Tiwari, M. K. 2014, Shiri, J. & Kisi, O. 2010). Cependant il ressort, outre l'intérêt de prétraiter les signaux en utilisant l'analyse multirésolution avec les modèles hydrologiques, que peu d'informations sur les limites de cette approche sont mentionnées.

L'approche que nous proposons ici s'inscrit dans la continuité des précédentes et vise 2 objectifs : (*i*) évaluer si la décomposition des signaux d'entrée et/ou de sortie en différentes échelles temporelles apporte un gain à la prévision de la turbidité, (*ii*) évaluer les limites de l'approche.

6.2 ANALYSE DU RESIDU DE PREDICTION

Le résidu de modélisation est un signal apportant de nombreuses informations sur les performances du modèle. Dans le cas d'une modélisation/prédiction parfaite, le résidu ne doit contenir que du bruit blanc. Ce bruit correspond au bruit présent dans les données d'entrée et de sortie du modèle. Ainsi, en analysant un résidu contenant des informations en sus du bruit, il est possible d'identifier les variations que le modèle n'arrive pas à prendre en compte. Sur la base de ces observations, le modélisateur peut modifier son modèle afin de l'orienter vers une meilleure prise en compte de l'information contenue dans le résidu.

Parmi l'ensemble des méthodes disponibles pour la caractérisation, l'extraction et la reconstruction du contenu temporel ou spectral d'un signal, notre choix s'est porté sur l'analyse multiresolution, qui est une technique de décomposition par transformée en ondelettes discrète. Nous utiliserons également la transformée

en ondelettes continue comme un préalable pour la visualisation du contenu spectral des signaux de turbidité à traiter. L'analyse multiresolution effectue la décomposition d'un signal en une succession de composantes orthogonales, appelés détails d'ondelette, comportant ainsi chacune une partie de l'information à des échelles de temps différentes. L'intérêt principal que nous voyons à cette technique est qu'elle permet d'effectuer la transformée inverse (retrouver le signal complet), au moyen d'une simple addition, sans erreur. Cet avantage a été mis à profit dans un assez grand nombre de travaux sur le sujet (Torrence, C. & Compo, G. P. 1998 ; Anctil, F. & Coulibaly, P. 2004 ; Labat, D. 2005; Massei, N. & *al.* 2017 ;etc...).

6.2.1 TRANSFORMEE EN ONDELETTES CONTINUE

Grâce aux résultats obtenus dans le chapitre 5 (modèles prédictifs de la turbidité à 12h et 24h), nous cherchons à déterminer, au moyen de l'analyse par ondelettes (continues et multirésolution), quel type d'information est contenu dans le résidu de modélisation. Un exemple de résultat obtenu est présenté sur la Figure 56. Les graphiques représentent les transformées en ondelettes du signal de turbidité (graphique de qauche) et du résidu de prévision (graphique de droite). L'échelle de couleur permet d'observer la puissance du signal en fonction de l'échelle et de la localisation temporelle. Ainsi, la couleur blanche indique la puissance la plus forte et la couleur rouge, la puissance la plus faible. On observe pour la décomposition de la turbidité observée (graphique de gauche) que l'information majoritairement présente dans le signal est comprise entre les échelles d'ondelettes 6 et 10. Eqalement, on observe deux cônes pour les échelles d'ondelettes de 1 à 6 correspondant au pic de turbidité présent à 100h et 320h pour l'événement 11 (Figure 57). Concernant le résidu de prévision (graphique de droite) on observe que l'information précédemment identifiée pour les échelles d'ondelettes de 6 à 10, ne se retrouve pas dans le résidu de modélisation. On en conclut que cette information a bien été intégrée par le modèle. En revanche, le signal majoritairement présent dans le résidu se retrouve au niveau des deux pics de turbidité situés entre les échelles 1 à 6. Cela signifie que les erreurs maximales se situent au niveau des 2 pics de turbidité. Il est à noter que les couleurs sur les 2 graphiques sont déterminées automatiquement en fonction des plages de valeurs du signal. Or le résidu est moins intense que le signal, ainsi les erreurs sur les pics apparaissant dans le résidu sont en réalité d'une puissance inférieure à celle du signal original. Contrairement à ce que pourrait suggérer le jaune clair présent dans le résidu au niveau des pics, jaune plus clair que celui du signal original.



Figure 56 : Résultat de la décomposition en ondelettes continue de la turbidité observée pour l'événement 11 (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Les échelles de couleurs ne correspondent pas à des valeurs absolues mais à des valeurs relatives, on ne peut donc pas comparer les puissances respectives entre les deux graphiques, mais seulement les puissances au sein d'un même graphique. Au travers de cette analyse, il apparaît donc que le résidu contient de l'information qui peut être associée à une échelle temporelle et donc à un phénomène ou à un processus. Cette analyse a été effectuée sur les autres évènements de turbidité et a abouti à une conclusion similaire pour chacun d'entre eux. Les graphiques relatifs à l'interprétation des différents évènements sont présents en Annexe 6. Afin de pouvoir continuer plus précisément, nous allons tenter, par l'analyse en ondelettes discrète, d'identifier les composantes du signal d'origine (turbidité observée) permettant de mieux décrire le résidu de prédiction.

6.2.2 ANALYSE MULTIRESOLUTION

Considérons par exemple l'évènement 11, sa durée est de 600h, la décomposition dyadique conduit à 9 niveaux de décomposition (de D1 à D9) et une approximation S9. Sur la Figure 57, nous présentons en (a) les turbidités observées et prédites à 12h par le modèle à deux branches dirigé. On peut noter que le modèle ne relève pas certaines variations de la turbidité et *a contrario* représente à certains instants des variations absentes du signal de la turbidité observée. Si l'on effectue la différence entre la turbidité observée et la turbidité prévue (b), on peut noter des variations importantes de la turbidité au moment des pics de turbidité.

Après avoir comparé plusieurs combinaisons de niveau de décomposition avec le résidu de prédiction, il ressort une forte concordance de ce dernier avec l'addition des niveaux de décomposition D1, D2 et D3 (Figure 57(c)).

Cette constatation renforce l'analyse visuelle réalisée à partir de l'ondelette continue. On pouvait y noter une information persistante dans le résidu pour les échelles allant de 2 à 6. En revanche, l'utilisation de l'ondelette discrète a mis en avant que les niveaux les plus régulièrement utilisés pour la représentation du résidu de prédiction pour les différents évènements de turbidité sont les niveaux D1, D2, et D3. Ainsi, les niveaux de décomposition D4, D5 et D6 sont moins prégnants dans le résidu par cette interprétation. Tout comme pour l'analyse en ondelettes continue, l'analyse précédente a été effectuée sur les autres évènements de turbidité et ont apporté une conclusion semblable pour chaque évènement. Les résultats de ces analyses sont présentés en annexe 9.



Figure 57 : Analyse du résidu du modèle dirigé à deux branches pour la prédiction à 12h de l'événement 11. (a) Turbidités prédite et observée. (b) Résidu de prévision. (c) Addition des niveaux de décomposition D1, D2 et D3.

Ainsi, une des voies d'amélioration du modèle de prévision, pourrait consister à mieux prendre en compte certaines fréquences. Grâce à ces observations, nous nous orienterons donc vers la création de deux modèles différents. Le premier modèle sera un modèle général composé d'un sous-modèle pour chaque niveau de décomposition. Le deuxième modèle sera composé de trois sous-modèles s'attachant respectivement à la prévision des niveaux de décomposition D1, D2 et D3 et d'un autre sous-modèle visant à prédire le signal composé de l'addition des niveaux de décomposition de D4 à D10 et du résidu (S). L'architecture de ces modèles sera décrite plus en détail dans le paragraphe 6.4.

Ainsi, contrairement aux travaux précédemment cités qui utilisent toutes les échelles d'ondelettes sans distinction, l'analyse du résidu nous a permis de cibler plus précisément comment utiliser chacun des niveaux de décomposition.

6.3 ENCADREMENT DE LA MISE EN PLACE DU MODELE

Dans cette partie, nous nous attachons à définir les limites d'utilisation de l'analyse multirésolution dans le cadre spécifique de la prédiction. Nous verrons dans un premier temps que la décomposition en ondelettes entraîne un effet de bord qui ne peut pas être négligé dès lors que l'on effectue de la prévision. L'impact de cet effet de bord sur les modèles dirigés et non-dirigés, et les moyens de pallier ce problème seront ensuite présentés.

6.3.1 LA DECOMPOSITION DU SIGNAL

Le fait de disposer de signaux qui ne sont pas infinis crée des perturbations dans le calcul des coefficients d'ondelettes lorsque l'on s'approche du début ou de la fin du signal. Ces perturbations sont appelées "effets de bord". Ces derniers ont un impact sur une durée plus ou moins longue en fonction de l'échelle considérée. De manière logique, la taille de l'effet de bord sera plus faible pour les premiers niveaux de décomposition (hautes fréquences), que pour les derniers niveaux de décomposition (basses fréquences).

Par exemple, sur la Figure 58 nous observons le premier niveau de décomposition (D1) de deux signaux correspondant à l'événement 11, pris sur deux durées différentes. Le premier signal correspond à une chronique de 1024h comprenant l'intégralité de l'événement 11 (Ligne pointillée noire sur la Figure 58). Le deuxième signal correspond au même évènement (d'une durée de 1024h), mais tronqué, d'une durée de 512h. Appliquons l'analyse multirésolution à ces deux signaux. Pour le premier signal (1024h = 2^{10} h), on obtient ainsi 10 niveaux de décomposition et une approximation ; alors que pour le deuxième signal ($512h = 2^9$ h), on obtient 9 niveaux de décomposition et une approximation. On peut ainsi comparer les 9 premiers niveaux de décomposition des deux signaux (1024h et 512h). Dans notre cas, la comparaison du premier niveau de décomposition sera suffisante pour illustrer notre propos.

Ainsi, sur la Figure 58, la ligne en tirets bleus correspond à l'échelle calculée sur le signal d'une durée de 1024h, la ligne en trait plein rouge correspond à l'échelle calculée pour le signal d'une durée de 512h. Comme il est visible sur la Figure 58, les lignes rouges et tirets bleus sont superposés entre o et 100h, c'est pour cela que l'on ne voit que la ligne rouge tracée par-dessus la ligne en tirets bleus. Au-delà de 100h, on observe une évolution différente de ces deux signaux. Pour une meilleure visibilité de cette divergence, la Figure 59 présente un zoom sur la zone encadrée. Il apparaît sur la Figure 59 que les 2 échelles sont significativement différentes sur une durée de 8h, à la fin de la ligne rouge, correspondant à la durée de l'effet de bord.



Figure 58 : Illustration de l'effet de bord issu de la décomposition en ondelettes (1^{er} niveau de décomposition soit D1) de deux signaux issus de l'évènement 11. Signal original d'une durée de 1024h (ligne pointillée noire) ; niveau de décomposition D1 du signal d'une durée de 1024h (ligne tiretée bleue); niveau de décomposition D1 du signal d'une durée de 512h (ligne pleine rouge). Le graphique présente un zoom sur les 300 premières heures afin de permettre une meilleure lisibilité.



Figure 59 : Agrandissement de la zone verte signalée sur la Figure 58. Niveau de décomposition D1 du signal d'une durée de 1024h (ligne en tirets bleus); niveau de décomposition D1 du signal d'une durée de 512h (trait plein rouge).

Ainsi, dans le cas d'un modèle dirigé, si nous cherchons à prévoir 1h à l'avance la turbidité et que nous bouclons le modèle avec 4h d'historique de turbidité, un problème peut être identifié. En effet, lors de son apprentissage, si nous nous plaçons dans la configuration illustrée par la figure 59 à l'instant k = 115h où le modèle tente de prédire la turbidité à l'heure 116, en connaissant parfaitement le futur, c'est à dire en utilisant la courbe bleue, le bouclage du modèle incorpore des données ne présentant pas d'effet de bord. En revanche, pour la même configuration, si le modèle est en situation réelle, c'est à dire s'il ne connaît pas l'avenir, il est dans la configuration de la courbe rouge ; ainsi, la rétroaction prend en compte le signal affecté par un effet de bord. On comprend alors que le modèle n'ayant pas été calibré en utilisant des données représentant cet effet de bord, l'utilisation d'un modèle dirigé induit une prise en compte de données différentes par rapport à celles utilisées lors de son apprentissage.

6.3.2 DEFAUTS DU MODELE DIRIGE ET AVANTAGE DU NON DIRIGE

Sur la base des observations précédentes, la suite des travaux s'attachera à créer et dimensionner un réseau de neurones pour chaque niveau de décomposition. A ce niveau, nous aurons la possibilité de créer deux architectures différentes, des modèles de types perceptron multicouche (cf. Chapitre 3: matériel et méthode) ou bien des modèles deux branches (cf. Chapitre 5: modèles prédictif de la turbidité à 12h et 24h). Compte tenu des résultats du chapitre 5, nous avons choisi d'améliorer le modèle à deux branches qui semblait le mieux prendre en compte les composantes de l'évapotranspiration (modèles prédictifs de la turbidité à 12h et 24h). Le dimensionnement de modèles basés sur l'architecture du perceptron multicouche n'a pas été effectué. Ainsi, le modèle à deux branches pourra être décliné en deux utilisations ; le modèle dirigé et le modèle non-dirigé. Comme indiqué dans le chapitre 4 (matériel et méthode), dans le cas d'un modèle dirigé, la sortie mesurée aux pas de temps précédents est appliquée en entrée du modèle pour les pas de temps précédant l'horizon de prévision.

Ainsi à cause de l'effet de bord intrinsèque à l'analyse multirésolution, il apparaît qu'il est risqué d'utiliser la combinaison d'un modèle dirigé et de l'analyse multirésolution pour effectuer la prévision ou la simulation d'un signal en temps réel. Dans le cas d'une telle application, il serait alors nécessaire de veiller à ce que le retour d'information effectué concerne des informations mesurées suffisamment lointaines de l'instant de calcul pour que les effets de bord puissent être négligés.

Cependant, il est possible d'envisager une alternative pour les modèles de prévision dirigés en utilisant la turbidité observée et non le niveau de décomposition que l'on cherche à prédire comme entrée du modèle afin de donner une information sur l'état du système en s'affranchissant de l'effet de bord en entrée du modèle.

Néanmoins, dans le cas d'un modèle non-dirigé, on utilise les données prédites par le modèle comme entrée du système aux pas de temps précédent l'instant de calcul. Ainsi, la combinaison d'un modèle non-dirigé et de l'analyse multirésolution est possible car c'est alors la composante estimée qui est renvoyée en entrée du modèle. Dans ce cas, lors de la conception du modèle, grâce à l'utilisation de données historiques et non mesurées en temps réel, la totalité des événements étant connus, les effets de bord provoqués par la décomposition en échelles ne sont présents qu'au début et à la fin de chaque événement. Il faut donc faire attention à ce que les parties intéressantes des signaux se trouvent au milieu de ce dernier, afin de pouvoir négliger les effets de bord.

6.3.3 MISE EN PLACE DES MODELES

ANALYSE MULTIRESOLUTION UTILISEE EN CONJONCTION AVEC DES RESEAUX DE NEURONES

Comme vu dans le paragraphe précédent, il est important de ne pas appliquer en temps réel au modèle des données pouvant présenter un effet de bord. Ainsi, les sous-modèles inspirés du fonctionnement "dirigés" ne sont pas réellement dirigés par les valeurs de leur variable de sortie aux pas de temps précédents, du fait des effets de bord mis en évidence précédemment, mais sont alimentés par la valeur de la turbidité observée aux pas de temps précédents. Cette information est évidemment bien moins adaptée que celle de la variable prédite. Il est donc supposé que les résultats seront moins bons dans cette configuration. Pour ôter toute ambiguïté dans les noms, nous appellerons ce mode : "T-dirigé" pour rappeler qu'il est dirigé par les mesures de turbidité et pas de la variable prédite.

Dans notre cas d'étude, nous utiliserons les précipitations brutes : non décomposées en entrée du modèle. Nous avons identifié dans le chapitre 4 (analyses des données), différents événements possédant des tailles différentes. Pour la suite des travaux, nous cherchons à calibrer des modèles pour chaque niveau de décomposition. Il est donc important que tous nos événements soient de la même durée afin d'obtenir le même nombre de niveaux de décomposition. Nous avons donc re-découpé ces derniers afin d'obtenir des événements d'une durée égale de 1024h. Ces événements re-découpés ont ensuite été décomposés en utilisant une ondelette de Daubechies de coefficient 20 (Daubechies, l. 1992). La taille de ces événements nous a ainsi permis d'obtenir 10 niveaux de décomposition ainsi qu'un résidu de décomposition. Nous calibrons un sous-modèle pour chacun de ces niveaux de décomposition. Deux modèles sont conçus : un T-dirigé et un non-dirigé.

De plus, comme nous travaillons avec des évènements de plus grande durée par rapport à ceux utilisés au chapitre 5 (modèles prédictifs de la turbidité à 12h et 24h), nous devons redimensionner des modèles du chapitre 5 sélectionnés pour cette étude (modèles deux branches dirigé et non-dirigé réalisant la prédiction à 12h). En effet, la durée des évènements peut jouer un rôle sur la qualité finale de la modélisation et ainsi modifier la qualité des modèles sans que cette dernière soit associée à l'intégration de la multirésolution. Les architectures des modèles définies et calibrées dans le chapitre 5 seront reprises et utilisées avec les données redimensionnées. Nous ne réaliserons pas cependant de redimensionnement des modèles en rapport avec la nouvelle taille des évènements.

ARCHITECTURE ET DONNEES CHOISIES

Deux types de modèle ont été conçus : le premier est le modèle appelé "total", il est composé de 11 sousmodèles calculant chacun les composantes D1, D2, D3, D4, D5, D6, D7, D8, D9, D10 et S. Ce modèle est réalisé en deux variantes : T-dirigé et non dirigé. La sortie est obtenue par la somme des sorties des sous-modèles.

Un second type de modèle a été conçu en prenant en compte la connaissance acquise en analysant les résidus de prédiction obtenus à partir des modèles réalisés au chapitre 5. Ce type de modèle est appelé "adapté", ce qui signifie adapté en fonction de l'analyse réalisée sur les résidus. Ce second type de modèle comporte 4 sous modèles : les trois premiers correspondent aux composantes D1, D2, D3, obtenues chacune par un sous-modèle, et d'un autre sous-modèle identifiant la somme de composantes D(4-10+S).

On pourra noter qu'*in fine* il n'est nécessaire que de concevoir 12 sous-modèles, en dirigé et non dirigé, comme indiqué dans le Tableau 7, les deux modèles "total" et "adapté" se composant alors de ces 24 sous-modèles de manière différente.

6.4 RESULTATS

6.4.1 MODELES NON-DIRIGES

Le Tableau 8 présente les architectures définies pour chacun des sous-modèles non-dirigés après leur sélection. On observe que tous les sous-modèles possèdent une largeur de fenêtre temporelle d'entrée pour l'évapotranspiration de trois pas de temps, d'un neurone pour la couche cachée relative à l'évapotranspiration et d'un ordre de 1. Les sous-modèles propres à chaque niveau de décomposition présentent des historiques de pluies allant de 25 à 50 pas de temps. Pour la couche cachée relative aux pluies, le nombre de neurones varie entre 5 et 10. Le nombre de neurones pour la couche de sortie, varie entre 5 et 15, indiquant une certaine complexité. Les seuls hyper-paramètres que nous pouvons donc mettre en relation avec l'échelle d'ondelettes sont les nombres de neurones cachés. Les sous-modèles D1 à D4 (relatifs aux hautes fréquences) semblent nécessiter un plus grand nombre de neurones cachés que pour les autres sous-modèles : entre 8 et 9 pour la couche cachée des pluies et entre 10 et 15 pour la couche de sortie. Pour le sous-modèle D(4-10+S), il semble que ce dernier utilise des hyper-paramètres égaux au maximum utilisé par les autres sous-modèles concernant la taille de la fenêtre des données d'entrée (soit une fenêtre de 50 pour les pluie et 3 pour l'ETP) et, à l'inverse, des paramètres égaux au minimum utilisé par les autres sous-modèles concernant le nombre de neurones (soit 5 pour la couche cachée relative aux pluies et 6 pour la couche de sortie). Ainsi, ce modèle, semble se caler sur les modèles présentant la plus faible complexité. En effet l'augmentation de la complexité du modèle dépend principalement de l'augmentation du nombre de neurones.

		Taille de l	a fenêtre	Nom	bre de neur	ones		
		Entrée pluie	Entrée ETP	Couche cachée Pluie	Couche cachée ETP	Couche de sortie	Ordre	N° évènement d'arrêt
	Dı	25	3	9	1	10	1	7
	D2	35	3	8	1	15	1	17
	D3	50	3	8	1	5	1	17
	D4	25	3	9	1	13	1	7
~	D5	25	3	9	1	5	1	7
Sous-	D6	25	3	5	1	5	1	7
Non-dirigés	D7	45	3	5	1	5	1	6
Non-unges	D8	40	3	5	1	6	1	6
	Dg	40	3	6	1	6	1	6
	D10	25	3	6	1	5	1	18
	S	25 3		10	1	11	1	6
	D(4-10+S)	50	3	5	1	6	1	3

Tableau 8 : Tableau récapitulatif des architectures des modèles non-dirigés.

Grâce à ces sous-modèles, assemblés pour former les modèles "total" et "adapté", nous avons effectué la prédiction à l'horizon 12h. Les résultats sont illustrés sur l'événement 11, Figure 6o. Sur cette figure, il apparaît que le modèle total et le modèle adapté n'effectuent pas de meilleures prévisions que celles obtenues précédemment au chapitre 5. Il est toutefois visible que le modèle adapté semble plus performant que le modèle total.



Figure 60 : Résultats des modèles non-dirigés : (a) deux branches, (b) deux branches total, (c) deux branches adapté.

Le Tableau 9 présente les critères de qualité calculés pour l'événement 11. On peut voir que pour la restitution du pic, c'est le modèle "deux branches 1024" qui est le plus performant mais que c'est en revanche le modèle total qui est le plus performant pour la synchronisation du pic.

		Modèles avec multirésolution						
	Deux branches 1024	Modèle total (D1 à D10 + S)	Modèle adapté (D1 à D3 + [D4-D10+S])					
Nash	0,46	0,24	0,39					
Pd	9	1	8					
Pp	26	18	22					

Tableau 9 : Résultats des critères de Nash, de pourcentage de restitution (*P_p*) et de délais du pic (*P_d*) pour les différents modèles à deux branches non-dirigé avec un horizon de prévision de 12h.

6.4.2 MODELES T-DIRIGES

En ce qui concerne les modèles T-dirigés, le Tableau 10 présente les sous-modèles obtenus à l'issue du processus de sélection. Les mêmes commentaires que pour les modèles non dirigés peuvent être effectués. Pour les modèles T-dirigés, il apparaît plus clairement qu'il y a une différence dans le nombre de neurones cachés pour les modèles prédisant les hautes fréquences et les modèles prédisant les basses fréquences. La prévision des hautes fréquences est plus difficile à faire que celle des basses fréquences.

		Taille de l	a fenêtre	No	mbre de	neurones		
		Entrée pluie	Entrée ETP	Couche cachée Pluie	Couche cachée ETP	Couche de sortie	Ordre	N° évènement d'arrêt
	Dı	25	3	13	1	13	5	7
	D2	35	3	13	1	15	5	17
	D3	50	3	13	1	5	1	17
	D4	25	3	10	1	13	5	7
	D5	25	3	9	1	5	3	7
Sous-	D6	25	3	5	1	5	5	7
modèles	D7	45	3	5	1	5	5	6
T-dirigés	D8	40	3	5	1	6	5	6
	Dg	40	3	6	1	6	5	6
	D10	25	3	6	1	5	5	18
	S	25	3	13	1	11	5	6
	D(4- 10+S)	50	3	5	1	6	5	3

Tableau 10 : Tableau récapitulatif des paramètres des modèles dirigés.

Les résultats obtenus pour la prédiction de l'événement 11, présentés en Figure 61, sont plus diversifiés que précédemment. En effet, il apparaît rapidement que la prédiction du modèle total (ayant un critère de Nash de 0,75) est de moins bonne qualité que la prédiction du modèle adapté (ayant un critère de Nash de 0,86). Cependant le critère de Nash n'étant pas le seul critère à prendre en compte, nous présentons dans le Tableau 11, les résultats de prévision évalués sur l'ensemble des critères déjà utilisés.

		Modèles avec multirésolution						
	Deux branches 1024	Modèle total (D1 à D10 + S)	Modèle adapté (D1 à D3 + [D4-D10+S])					
Nash	0,91	0,85	0,86					
P _d	0	0	13					
P _p	72	33	65					

Tableau 11 : Résultats des critères de Nash, de pourcentage de restitution (*P_p*) et de retard du pic (*P_d*) pour les différents modèles à deux branches dirigés avec un horizon de prévision de 12h.

Le modèle total présente une relativement bonne synchronisation du pic majeur de turbidité mais l'amplitude de ce dernier n'est pas bien restituée. Pour le modèle adapté, la synchronisation du pic est moins bonne mais l'amplitude est mieux respectée. En comparaison avec les résultats du modèle présenté au chapitre 5 (modèles prédictifs de la turbidité à 12h), il apparaît que pour le modèle nouvellement créé, le pic majeur de turbidité présente une amplitude plus faible que pour le modèle deux branches dirigé 1024 avec des variations au sein du pic plus importantes.





Figure 61: Résultats des modèles T-dirigés : (a) deux branches, (b) deux branches total, (c) deux branches adapté.

6.4.3 DISCUSSION

On ne peut définitivement conclure sur l'examen d'un seul évènement, le Tableau 12 présente donc les critères de Nash des différents modèles utilisant la multirésolution, calculés sur les évènements de turbidité de notre base de données. Les évènements 3, 6, 7, 17 et 18 ne sont pas testés car ils sont utilisés au moins une fois en tant qu'ensemble d'arrêt par les modèles utilisant l'analyse multirésolution et il n'est alors pas possible d'obtenir des résultats de prédictions pour ces derniers car il manque la prédiction d'au moins un niveau de décomposition pour l'assemblage du signal final.

Si nous nous intéressons au critère de Nash des modèles T-dirigés utilisant la multirésolution, il est visible que ces derniers n'apportent pas d'amélioration par rapport aux modèles dirigés établis dans le chapitre <u>5</u> (modèles prédictifs de la turbidité à <u>12</u>h). A L'exception de l'évènement <u>2</u> ou la différence de critère de Nash est fortement négligeable, le modèle deux branches dirigé présente toujours les meilleurs résultats de critère de Nash. Pour les modèles T-dirigés, nous voyons que le modèle total ne permet pas une amélioration du critère de Nash. Le modèle adapté présente un critère de Nash légèrement supérieur pour l'événement <u>11</u> mais un critère de Nash inférieur au modèle à deux branches dirigé <u>1024</u> pour les autres événements.

Les modèles non dirigés présentent des résultats plus mitigés. En effet, le modèle présentant le plus fréquemment les meilleurs scores de Nash est le modèle couplé avec des scores de Nash proches de 0,7. On notera deux exceptions; une pour l'évènement 11 ou le modèle à deux branches 1024 est le meilleur et une pour l'évènement 16 ou le meilleur modèle se trouve être le modèle total. Le modèle couplé non-dirigé se trouve bien meilleur que le modèle à deux branches non-dirigé 1024. Ceci peut s'expliquer par une meilleure prise en compte des variations basses fréquences qui permettent d'obtenir une prédiction avec une plus forte amplitude que pour le modèle à deux branches non-dirigé 1024.

	Deux bra	nches 1024		Modèles avec multirésolution						
	Dirigé	Non-dirigé	Modèle total	(D1 à D10 + S)	Modèle adapté (D1 à D3 + [D4- D10+S])					
	,	,	T-dirigé	Non-dirigé	T-dirigé	Non-dirigé				
Nash T2	0,93	0,63	0,84	0,52	0,94	0,75				
Nash T10	0,56	0,65	0,50	0,59	0,19	0,69				
Nash T11	0,91	0,46	0,85	0,24	0,86	0,39				
Nash T16	0,75	Х	0,47	0,42	0,45	0,34				
Nash T22	0,89	0,65	0,76	0,44	0,87	0,71				

Tableau 12 : Résultats des critères de Nash pour les différents modèles à deux branches avec un horizon de prévision de 12h.

On peut également vérifier ces interprétations en étudiant les résultats des différents modèles à l'aide des critères d'amplitude du pic (Pp) et du retard du pic (Pd) comme présenté dans les Tableau 13 et Tableau 14.

	Deux bran	ches 1024	Modèles avec multirésolution						
	Dirigé	Non-dirigé	Modèle total	(D1 à D10 + S)	Modèle adapté (D1 à D3 + [D4-D10+S])				
	5	5	T-dirigé	Non-dirigé	T-dirigé	Non-dirigé			
P_{p} T2	90	37	53	30	81	33			
<i>P</i> _p T10	80	38	36	14	39	22			
<i>P</i> _p T11	78	26	36	18	66	22			
<i>P</i> _p T16	88	28	35	16	45	13			
<i>P</i> _p T22	155 131		78	103	186	148			

Tableau 13: Résultats des critères de pourcentage de pic (*P_P*) pour les différents modèles à deux branches avec un horizon de prévision de 12h.

	Deux bran	iches 1024	Modèles avec multirésolution						
	Dirigé	Non-dirigé	Modèle total	(D1 à D10 + S)	Modèle adap [D4-D	uté (D1 à D3 + 10+S])			
	,	,	T-dirigé	Non-dirigé	T-dirigé	Non-dirigé			
$P_d T_2$	0	-17	6	-23	0	-23			
<i>Pd</i> T10	4	12	18	-2	4	1			
Pd T11	-3	9	0	1	13	3			
<i>P</i> _d T16	1	-37	16	-11	7	-10			
<i>P</i> _d T22	-1	14	12	-6	-2	5			

Tableau 14: Résultats des critères de pourcentage de pic (*P*_d) pour les différents modèles à deux branches avec un horizon de prévision de 12h.

Il apparaît donc au travers de ces résultats que le modèle dirigé n'utilisant pas la multirésolution est plus performant. En revanche, concernant les modèles non-dirigés, le prétraitement utilisant l'analyse multi-résolution semble apporter quelques améliorations.

Pour conclure définitivement sur l'intérêt d'ajouter un prétraitement réalisé par l'analyse multirésolution, certaines expériences que nous n'avons pas pu réaliser dans le temps imparti à cette thèse manquent encore :

- Mesurer la perte/gain de performance des modèles « total » et « adapté » par rapport à un modèle conçu avec des évènements de 1024 valeurs.
- Effectuer le même travail avec le perceptron multicouche
- Effectuer un travail équivalent, y compris avec le modèle T-dirigé, mais en utilisant un forçage calculé par la "taverlette", issue d'un simple filtrage, sans effet de bord et causal, qui permet de travailler en prévision. Cette méthode a été proposée par Taver dans sa thèse (Taver, V., 2014).

6.5 CONCLUSION DU CHAPITRE 6

Ce chapitre s'est attaché à présenter comment l'analyse multirésolution pouvait aider le modèle neuronal à mieux représenter les différentes échelles temporelles au sein du signal. Plusieurs architectures ont été conçues puis testées : le modèle à deux branches, sa version décomposée selon toutes les composantes dyadiques, et sa version comportant des regroupements de composantes afin de mieux s'adapter à l'erreur de prévision préalablement analysée.

Du fait de l'importance majeure des effets de bord sur l'utilisation des modèles de type dirigé en prévision, nous avons dû proposer une nouvelle façon d'implémenter ce principe, appelée T-dirigé, pour le rendre utilisable en prévision. De ce fait, et en dépit de la proposition faite, les résultats obtenus pour le modèle T-dirigé n'ont pas été à la hauteur de nos attentes.

Un premier travail envisagé serait d'effectuer les mêmes travaux en utilisant le perceptron multicouche afin d'étudier le comportement de cette architecture.

Pour les modèles présentement étudiés, il est apparu que la simplification de la modélisation que nous cherchions à réaliser a en réalité été une complexification. La fonction de transfert du système et la fonction de transfert de la méthode de décomposition du signal ont été combinées et ont résulté pour le modèle en une fonction de transfert plus complexe que la fonction de transfert du système (Annexe 13). Si nous voulons simplifier la fonction de transfert, il nous faut décomposer les données d'entrées de la même manière que les données de sorties. Ainsi, il est nécessaire comme nous l'avons présenté de trouver des méthodes de décomposition du signal n'entrainant pas d'effets de bord.

Plusieurs pistes de recherche ont donc été ouvertes mais des travaux complémentaires doivent encore être menés avant de pouvoir conclure définitivement, par exemple refaire la même étude en utilisant la méthode dite "taverlette". Dans le même objectif, il serait intéressant d'explorer les différentes méthodes de

décomposition du signal n'entrainant pas d'effet de bord afin de les intégrer aux modèles et de tester les performances des modèles ainsi obtenus.

Le principal apport de ce travail, à notre sens, est d'avoir montré que les effets de bord empêchent d'appliquer simplement l'analyse multirésolution au modèle dirigé, qui est celui qui est plébiscité dans les études de prévision utilisant des pré- ou post-traitements avec les ondelettes discrètes. Il apparaît ainsi que la majeure partie des études publiées depuis une dizaine d'années et montrant un grand bénéfice de l'analyse multirésolution ne peuvent être utilisées en temps réel. Ce qui réduit de fait leur application opérationnelle.

Chapitre 7 : Vers un indicateur de la présence de produits phytosanitaires

CHAPITRE 7 : VERS UN INDICATEUR DE LA PRESENCE DE PRODUITS PHYTOSANITAIRES

7.1 INTRODUCTION

La turbidité en elle-même, tels qu'elle est mesurée, et que nous avons tenté de prédire au moyen de différentes approches, n'est pas spécifiquement par elle-même le facteur de danger sanitaire. De par les processus impliqués dans les variations de turbidité, elle constitue un indicateur de contamination potentiel. Parmi ces contaminations, les produits phytosanitaires constituent une préoccupation majeure. La question de l'association entre turbidité et produit phytosanitaire qui permettrait d'utiliser la turbidité comme un indicateur de leur occurrence est posée ici. De nombreuses études sont effectuées afin de contrôler les concentrations en produits phytosanitaires dans l'eau souterraine et de surface afin de satisfaire les normes règlementaires. Comme nous l'avons vu dans le chapitre 1, un suivi des concentrations souterraines est également effectué par les Agences de l'eau dans le cadre de la DCE visant à contrôler et atteindre un bon état des masses d'eau souterraines. Dans notre cas d'étude, la connaissance des concentrations en produits phytosanitaires est un enjeu crucial pour la CODAH afin d'éviter une possible contamination de l'eau destinée à l'alimentation en eau potable. Comme il a été montré dans le chapitre 4, les données actuellement archivées présentent des échantillonnages avec un pas de temps ne permettant pas d'identifier de manière systématique de possibles variations rapides (quelques jours).

Nous allons donc, dans un premier temps, comparer deux méthodes d'analyses des produits phytosanitaires qui sont : les échantillons d'eau et les échantillonneurs passifs de type POCIS.

Par la suite nous tenterons d'identifier un proxy des produits phytosanitaires afin de permettre une meilleure gestion du pompage d'Yport plus en accord les variations en produits phytosanitaires.

7.2 ECHANTILLONNAGE

7.2.1 CHOIX DE L'EMPLACEMENT DES CAPTEURS

La mise en place d'échantillonneurs passifs nécessite d'avoir accès au conduit karstique. Malheureusement, cette opération est difficilement envisageable dans notre cas d'étude pour les raisons suivantes :

- le conduit karstique est noyé,
- l'accès au conduit est très difficile et nécessite une logistique importante (descente verticale dans le conduit et nécessité l'intervention de plongeurs).
- les vitesses d'écoulements dans le conduit karstique peuvent être importantes ; elles ont été mesurées entre 1,8 et 2,4 m/s lors de la campagne de plongée de 2012 au droit du forage d'Yport (Chedeville, S., 2014). Outre le danger pour les plongeurs, de telles vitesses entraînent un risque d'arrachement du matériel de mesure.

Cependant, comme évoqué dans le chapitre 2, la mesure et l'archivage de la turbidité et de la conductivité de l'eau dans le conduit karstique sont réalisées depuis 2013. Les mesures sont effectuées en continu grâce à la mise en place d'une pompe indépendante des pompes alimentant la station d'Yport (rappelons que ces dernières sont arrêtées lors des épisodes turbides ou de pollutions). Cette pompe supplémentaire alimente en continu un réservoir, situé à l'entrée de la station, dans lequel les mesures de turbidité et de conductivité sont effectuées. Comme indiqué sur le schéma de la Figure 62, c'est dans ce réservoir qu'ont été placés les POCIS. Il convient de préciser que le cheminement de l'eau permet une mesure en milieu peu turbulent mais avec un débit suffisant pour éviter l'encrassement des POCIS ou bien la formation d'un biofilm. En sus, l'absence de lumière joue également un rôle important pour éviter le développement d'un biofilm.

Chapitre 7 : Vers un indicateur de la présence de produits phytosanitaires



Figure 62 : Dispositif de mesure de la turbidité et de la conductivité mis en place en 2013.

Afin de mieux appréhender le fonctionnement du système de collecte et de mesure, voici un descriptif du cheminement de l'eau dans le système présenté en Figure 62 :

- L'eau arrive dans un premier temps dans le système par le tube d'alimentation situé en haut à gauche du schéma (1). Le débit d'alimentation peut être réglé arbitrairement avec la vanne rouge en amont de l'entrée d'eau dans le réservoir.
- L'eau se heurte à la paroi face au point d'entrée de l'eau dans le système. L'eau perd alors de sa vélocité et n'a comme passage possible que l'ouverture située dans la partie inférieure du réservoir.
- A ce stade, suivant la position de la vanne de purge, l'eau peut emprunter deux cheminements ; (i) si la vanne de purge (3) est en position ouverte, l'eau passe par l'ouverture au fond du réservoir et est extraite du système, (ii) si la vanne de purge est en position fermée, l'eau remonte en direction du trop-plein et est ensuite extraite du système.

Il est à noter que le débit n'est pas obligatoirement constant et peut varier suivant l'ouverture de la vanne d'alimentation.

PROTOCOLE DE PRELEVEMENT

Comme indiqué sur la Figure 62, les POCIS sont placés au fond du réservoir, dans le courant de l'eau dont la vitesse n'est jamais nulle. Les POCIS sont ainsi placés dans la zone optimale du dispositif.

Lors du renouvellement des POCIS et le prélèvement des échantillons d'eau, le protocole suivant a été respecté :

- Prélèvement des POCIS et rinçage à l'eau distillée si une campagne précède notre intervention
- Ouverture de la vanne de purge pour vider et nettoyer le réservoir
- Rinçage à deux reprises de la bouteille de prélèvement avec l'eau brute
- Prélèvement de l'échantillon d'eau au niveau de la vidange du système (1L soit deux fois 500mL pour avoir un réplica).
- Fermeture de la vanne de vidange pour permettre un remplissage du réservoir
- Déploiement des trois POCIS installés sur leur support de maintien (Figure 63).

Les différents déploiements effectués durant ces travaux de thèse ont été réalisés selon ce protocole afin de permettre une reproductibilité des mesures et une diminution de l'impact des conditions d'échantillonnage sur les résultats.
Suite à ces déploiements, plusieurs observations ont été faites ;

- Le positionnement des POCIS a induit une différence de concentration lors des analyses. Le POCIS situé proche de l'arrivée d'eau comportait généralement une concentration plus élevée que les deux autres. Cependant, il est à noter que le positionnement des POCIS ne pouvait pas être fait différemment dans le réservoir à cause de la présence du tube de vidange (Figure 62 et Figure 63).
- L'encrassement des POCIS, bien que souvent visible n'a pas été suffisant pour interférer avec l'accumulation des produits phytosanitaires. Les POCIS ayant été les plus encrassés sont visibles sur la Figure 63.
- Il n'y a pas eu de formation de biofilm.



Figure 63 : Positionnement et visualisation de l'encrassement des POCIS.

Ainsi, l'environnement choisi, à l'exception de la plus forte accumulation réalisée par le POCIS proche de l'arrivée d'eau, n'a pas d'impact majeur sur les résultats des analyses et nous avons pu obtenir des résultats pour chaque campagne de mise en place des POCIS.

7.2.2 CAMPAGNES D'ECHANTILLONNAGE

Les premières campagnes de prélèvements avaient pour objectif de mesurer les variations des concentrations pour les produits phytosanitaires durant les périodes susceptibles de voir se produire des pics de turbidité. Malheureusement, aucun pic de turbidité supérieur à 30 NTU n'est survenu durant cette première campagne (déploiement 1 et 2 -> "mai" et "juillet"). Pour la deuxième campagne (déploiement 3 à 7 -> "hiver") et la troisième campagne (déploiement 8 à 16 -> "printemps"/"été"), l'accent a été mis sur l'observation des variations saisonnières des concentrations en produits phytosanitaires. Ainsi 16 déploiements ont été réalisés durant ces travaux de thèse selon le calendrier suivant (chaque puce représente une campagne d'échantillonnage continue):

- Du 04/05/2016 au 20/05/2016 soit 16 jours pour le déploiement libellé "mai" (1^{er} déploiement). Comprenant 3 POCIS et deux échantillons d'eau. ["mai (1)" et "mai (2)"].
- Du o6/07/2016 au 28/07/2016 soit 22 jours pour le déploiement libellé "juillet "(2^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau. ["juillet (1)" et "juillet (2)"].
- Du 13/10/2016 au 28/10/2016 soit 15 jours pour le déploiement libellé "hiver (1)" (3^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau. ["hiver (1)" et "hiver (1-2)"].
- Du 28/10/2016 au 18/11/2016 soit 21 jours pour le déploiement libellé "hiver (2)" (4^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["hiver (1-2)" et "hiver (2-3)"].
- Du 18/11/2016 au 02/12/2016 soit 14 jours pour le déploiement libellé Hiver (3) (5^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["hiver (2-3)" et "hiver (3-4)"].
- Du 02/12/2016 au 20/12/2016 soit 18 jours pour le déploiement libellé "hiver (4)" (6^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["hiver (3-4)" et "hiver (4-5)"].

- Du 20/12/2016 au 10/01/2017 soit 21 jours pour le déploiement libellé "hiver (5)" (7^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["hiver (4-5)" et Hiver "(5)"].
- Du 28/03/2017 au 11/04/2017 soit 14 jours pour le déploiement libellé "printemps (1)" (8^{éme} déploiement). Comprenant 3 POCIS et deux échantillons d'eau ["printemps (1)" et "printemps (1-2)"].
- Du 11/04/2017 au 02/05/2017 soit 21 jours pour le déploiement libellé "printemps (2)" (9^{éme} déploiement). Comprenant 3 POCIS et deux échantillons d'eau ["printemps (1-2)" et "printemps (2-3)"].
- Du 02/05/2017 au 22/05/2017 soit 20 jours pour le déploiement libellé "printemps (3)" (10^{éme} déploiement). Comprenant 3 POCIS et deux échantillons d'eau ["printemps (2-3)" et "printemps (3-4)"].
- Du 22/05/2017 au 08/06/2017 soit 17 jours pour le déploiement libellé "printemps (4)" (11^{éme} déploiement). Comprenant 3 POCIS et deux échantillons d'eau ["printemps (3-4)" et "printemps (4-5)"].
- Du 08/06/2017 au 06/07/2017 soit 28 jours pour le déploiement libellé "printemps (5)" (12^{éme} déploiement). Comprenant 3 POCIS et deux échantillons d'eau ["printemps (4-5)" et "printemps (5)/été (1)"].
- Du o6/07/2017 au 24/07/2017 soit 18 jours pour le déploiement libellé "été (1)" (13^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["printemps (5)/été (1)" et "été (1-2)"].
- Du 24/07/2017 au 14/08/2017 soit 21 jours pour le déploiement libellé "été (2)" (14^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["été (1-2)" et "été (2-3)"].
- Du 14/08/2017 au 04/09/2017 soit 21 jours pour le déploiement libellé "été (3)" (14^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["été (2-3)" et "été (3-4)"].
- Du 04/09/2017 au 25/09/2017 soit 21 jours pour le déploiement libellé "été (4)" (15^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["été (3-4)" et "été (4-5)"].
- Du 25/09/2017 au 09/10/2017 soit 14 jours pour le déploiement libellé "été (5)" (16^{éme} déploiement).
 Comprenant 3 POCIS et deux échantillons d'eau ["été (4-5)" et "été (5)"].

Ainsi 42 échantillons d'eau (21 échantillons avec un réplica) et 51 POCIS (17 triplicas de POCIS mis en places) ont été analysés. On notera que pour les déploiements 13 et 14, l'échantillon d'eau commun entre ces deux campagnes (24/07/2017 soit "été (1-2)") n'a pu être analysé car le flacon s'est cassé durant le transport. On notera également qu'en temps normal, comme précisé dans le chapitre 3, l'utilisation des taux d'accumulation des POCIS (*Rs*) permettent de calculer une accumulation par unité de temps.

Cependant, en raison de l'envoi de POCIS avec la mauvaise référence par notre fournisseur, la conversion de la quantité de produit par gramme de phase en mg/L n'a pas pu être effectuée pour tous les échantillons. En effet, la phase utilisée dans les POCIS des déploiements 3, 4 et 5 (soit "hiver (1)", "hiver (2)" et "hiver (3)") est une phase de type PESTICIDE® et non PHARMA®. Une fois l'erreur identifiée et remontée à notre fournisseur, il a été jugé préjudiciable pour notre étude de transposer les taux d'accumulation obtenus pour les POCIS « PHARMA® » aux POCIS "PESTICIDE®" même si ces derniers possèdent des capacités de rétention intéressantes vis-à-vis des produits phytosanitaires cibles.

Pour plus de fiabilité, il aurait été nécessaire de déterminer les taux d'accumulation pour les POCIS PESTICIDES® mais cette procédure n'était pas envisageable avec les ressources et le temps imparti à nos travaux. Ainsi, seule la quantité de produits phytosanitaires par masse de phase par jour sera utilisée dans l'interprétation des résultats.

Cette partie présente les résultats obtenus pour l'analyse des échantillons d'eau et l'analyse de la phase des POCIS. Dans un premier temps nous comparerons les résultats obtenus avec les échantillons d'eau et les POCIS. Nous verrons au travers de cette comparaison que les POCIS semblent apporter une information un peu différente sur la qualité de l'eau que les échantillons d'eau ponctuels. A la suite de cette partie nous effectuerons pour chaque méthode d'échantillonnage la comparaison des résultats d'analyse avec l'occurrence des précipitations, la conductivité et la turbidité. Cette partie nous permettra d'apporter de nouveaux éléments d'appuyant l'intérêt des POCIS. Nous verrons également que la turbidité n'est pas un bon indicateur de la présence de produits phytosanitaires.

7.3.1 COMPARAISON DE LA METHODE D'ECHANTILLONNAGE (ECHANTILLONS D'EAU VS. POCIS)

RESULTATS DES ECHANTILLONS D'EAU

Dans cette partie, nous présentons les résultats d'analyse des 42 échantillons d'eau correspondant aux 21 échantillonnages effectués en doublon. Pour l'obtention des résultats d'analyses, des triplicats des analyses sont effectués et la moyenne des résultats pour un échantillon est utilisée pour obtenir la concentration finale en produits phytosanitaires. On effectue ainsi 126 analyses pour les échantillons d'eau brute puis effectuons une moyenne des triplicats et obtenons 21 résultats pour les 32 produits analysés. Ces résultats sont représentés en Figure 64.

Sur cette figure les concentrations inférieures au seuil de quantification sont ramenées à la valeur de ce seuil. De même pour le seuil de détection. Pour les molécules non détectées, leur concentration est fixée à o ng/L. Après avoir analysé la Figure 64, nous effectuerons une analyse par molécule de la fréquence d'occurrence en se basant sur les seuils de détection puis les seuils de quantification (Figure 65 et Figure 66).

CONCENTRATIONS EN PRODUITS PHYTOSANITAIRES

Concernant l'analyse des échantillons d'eau, il est visible dans la Figure 64 que toutes les molécules ne sont pas nécessairement retrouvées lors de chacun des déploiements. On notera la présence des écarts type calculés via les triplicats d'analyses. Ainsi, on observe un très fort écart type pour l'échantillon "hiver (5) ". Ce fort écart type est le résultat d'une mauvaise homogénéisation du flacon avant son analyse. Ainsi, en divisant le prélèvement en deux flacons distincts, les duplicats ne contenaient pas une concentration identique. Le premier duplicat contenait une concentration en produits bien inférieure à celle de l'autre flacon. Comme dit précédemment, nous n'avons pas de résultats pour le déploiement "été (1-2)" car le flacon s'est brisé durant le transport. Nous représentons tout de même cet échantillon pour garder une continuité dans le graphique.

En analysant le nombre de produits identifiés durant chaque déploiement, II apparait une forte disparité. En effet, on observe un faible nombre de produits phytosanitaires pour les déploiements suivants : "mai (1)", "mai (2)", "juillet (1)", "juillet (2)", "hiver (1-2)", "hiver (2-3)", "hiver (3-4)", "hiver (4-5)", "hiver (5)", "printemps (5)/été (1)", "été (2-3)", "été (3-4)", "été (4-5)" et "été (5)". A contrario, un nombre plus important de produits phytosanitaires est observé pour les déploiements ; "hiver (1)", "printemps (1)", "printemps (1-2)", "printemps (2-3)", "printemps (3-4)" et "printemps (4-5)". Cette disparité peut être expliquée par l'utilisation de nombreux produits phytosanitaires en vue de la préparation du sol à la fin de l'hiver et le traitement des plantations durant le printemps. Cette hypothèse se confirme par la présence en grande partie d'herbicide (couleurs froides vertes et bleu sur la Figure 64) durant les déploiements de : "mai (1)", "mai (2)", "juillet (1)", "juillet (2)", "hiver (1-2)", "hiver (2-3)", "hiver (3-4)", "hiver (4-5)" et "hiver (5)" puis de l'utilisation plus importante de fongicide (couleurs chaudes rouges et jaunes sur la Figure 64) une fois que les plantes ont débuté leur développement pour les déploiements : "printemps (1)", "printemps (1-2)", "printemps (2-3)", "été (3-4)", "été (4-5)" et "été (5)". Par la suite, durant les déploiements "printemps (5)/été (1)", "été (2-3)", "été (3-4)", "été (4-5)" et "été (5)", on retrouve une utilisation plus faible des fongicides. On notera également la présence d'insecticides (en noir) dans les résultats d'analyses pour les déploiements "printemps (1)", "printemps (1-2)", "printemps (2-3)", "printemps (2-3)", "printemps (2-3)", "printemps (2-3)", "printemps (2-3)", "printemps

(3-4)". Cela confirme également la présence de plantes en cours de développement que les agriculteurs cherchent à protéger.

Pour la majorité des produits, les concentrations observées sont inférieures à 30 ng/L. Seules quelques exceptions présentent des valeurs nettement plus élevées telles que le DIA (déisopropylatrazine) pour le déploiement "juillet (1)", avec une concentration de 588 ng/L. Le tébuconazole quant à lui atteint une concentration exceptionnelle de 1885 ng/L pour la campagne "printemps (1-2)" et une concentration de 210 ng/L pour la campagne "printemps (3-4)". Le tébuconazole présente également une concentration de 62 ng/L pour la campagne de "printemps (4)".

Ces fortes concentrations observées pour le DEA (déséthylatrazine) et le tébuconazole sont certes très importantes mais restent cependant inférieures aux 2 µg/L préconisés pour l'alimentation en eau potable. Ainsi, ces deux produits (DEA et tébuconazole), semblent être dus à une pollution ponctuelle. Cette hypothèse est confirmée par les résultats des POCIS que nous verrons par la suite.



Figure 64 : Résultats des analyses des échantillons d'eau. Les produit : de couleur bleue correspondent aux herbicides interdits d'utilisation en France, de couleur verte les herbicides autorisés à l'utilisation en France, de couleur rouge les fongicides interdits d'utilisation en France, de couleur jaune les fongicides autorisés à l'utilisation en France et en noir les insecticides autorisés à l'utilisation en France. Les noms soulignés correspondent aux produits de dégradation du produit les précédents dans la liste du graphique (lecture de gauche à droite et de haut en bas). La concentration en tebuconazole n'est pas affichée à sa vraie valeur afin de ne pas "écraser" toutes les autres valeurs.

Par ailleurs, la concentration en DEA dépasse régulièrement les 20 ng/L qui reste dans le même ordre de grandeur que la concentration en atrazine (>10 ng/L). Le DEA étant un produit de décomposition de l'atrazine, il est normal d'observer une variation similaire de leurs concentrations. Cependant, cette variation ne présente pas de similarité pour les campagnes de "mai (1)", "mai (2)", "juillet (1)", "juillet (2)" et "hiver (1)". Cette disparité dans leur variation de concentrations peut être expliquée par des conditions environnementales (température,

humidité du sol, pH, cycle de vie des micro-organismes du sol,...), plus ou moins favorables à la décomposition de l'atrazine⁴.

FREQUENCES D'OCCURRENCE DES PRODUITS PHYTOSANITAIRES

Dans ce paragraphe, nous cherchons à identifier les molécules le plus fréquemment retrouvées dans les analyses et ainsi tenter de caractériser la signature de notre eau. Ainsi, le taux de détection (Figure 65) et de quantification (Figure 66) ont été calculés pour chacune des molécules en prenant tous les échantillons d'eau analysés. L'échantillon d'eau du 24/07/2017 qui a été brisé lors du transport n'est pas pris en compte. Cette partie nous permettra également par la suite de comparer les résultats obtenus avec les échantillons d'eau et les POCIS. Nous verrons ainsi quelle méthode nous permet de détecter ou de quantifier au mieux les produits présents dans l'eau souterraine de notre zone d'étude.

La Figure 65 représente le taux de dépassement du seuil de détection de chacune des molécules. Trois grandes classes sont définies en fonction de leur taux de dépassement du seuil de détection:

- Classe 1. Les molécules dont le taux de détection dépasse 80% (atrazine, DEA, simazine et DIA)
- Classe 2. Les molécules dont le taux de détection est compris entre 80% et 60% (oxadixyl, DET et flazasulfuron)



Classe 3. Les molécules dont le taux de détection est inférieur à 60%

Figure 65 : Classement des molécules suivant leur taux de dépassement du seuil de détection (analyse des échantillons d'eau).

Ainsi, les molécules identifiées les plus fréquemment (taux de détection > 80%) sont : le DEA, l'atrazine, la simazine et le DIA. Ces quatre produits sont également majoritairement retrouvés lors de l'analyse des POCIS. Le DIA et le DEA étant des produits de décomposition de l'atrazine, il est normal de les trouver régulièrement au vu du taux de détection de 100% de l'atrazine.

Lors de l'analyse des échantillons d'eau, 23 produits au total ont dépassé les seuils de détection au moins une fois.

⁴ Le temps de demi-vie de l'atrazine varie entre 1 et 60 jours suivants les conditions environnementales et biologiques responsables de sa dégradation (Tomlin, C., 1994).





Figure 66 : Classement des produits suivant leur fréquence de dépassement du seuil de quantification (analyse des échantillons d'eau).

En ce qui concerne la fréquence de quantification (Figure 66), en considérant les mêmes classes que celles fixées précédemment, 3 produits ont un taux de détection supérieur à 80% (atrazine, DEA et simazine), 2 produits ont un taux de détection compris entre 80% et 60% (DIA, et flazasulfuron) et enfin, 7 produits avec un taux de détection inférieure à 60%. Ainsi, 12 produits ont dépassé au moins une fois le seuil de quantification.

Ainsi les produits pour lesquels la concentration réelle a été la plus fréquemment identifiée sont les produits de la classe 1 (>80%) soit l'atrazine et le DEA.

Ainsi, nous avons vu dans cette partie que l'analyse des échantillons d'eau nous a permis de détecter 23 produits dans le milieu naturel ; 12 produits sur les 23 produits détectés ont pu être quantifiés. Nous avons également vu que l'échantillonnage à l'aide de prélèvements d'eau ponctuels peu fréquents (moins d'une fois par jour) présente une forte sensibilité aux variations éphémères de concentrations. De plus, cette méthode ne s'avère pas être la plus robuste pour étudier la qualité moyenne d'une nappe d'eau dans le temps. Cette méthode présente un fort risque de sous échantillonnage étant donné les périodes de non analyses séparant deux prélèvements (de l'ordre de 15 jours).

RESULTATS DE L'UTILISATION DES POCIS

Contrairement aux échantillons d'eau, prélevés ponctuellement dans le temps, les POCIS permettent d'avoir une accumulation des substances au cours du temps de leur exposition. Ainsi ils enregistrent une valeur moyenne. Une pollution ponctuelle est donc gardée en mémoire, même si son intensité est diminuée par la moyenne, mais l'information est toujours présente. Par ailleurs, un produit en faible concentration, inférieure au seuil de détection, mais présent en continu au cours du temps est accumulé au sein du POCIS, sa concentration, du fait de l'accumulation, peut alors dépasser le seuil de détection à l'issue de l'analyse du POCIS. Les POCIS, par l'accumulation qu'ils réalisent quand ils sont dans le milieu, peuvent donc permettre de détecter des produits dont la concentration. Les résultats obtenus par l'analyse des POCIS sont représentés en Figure 67. Comme précédemment, sur cette figure, les concentrations inférieures au seuil de quantification sont ramenées à la valeur de ce seuil. Il en va de même pour le seuil de détection. Nous analyserons également la fréquence de dépassement du seuil de détection et du seuil de quantification.

CONCENTRATIONS EN PRODUITS PHYTOSANITAIRES

La Figure 67 illustre les résultats obtenus pour l'analyse de la phase des POCIS. Contrairement aux résultats obtenus pour l'analyse des échantillons d'eau, seul un produit (le tébuconazole) présente une concentration

élevée (45 ng/g de phase/jour) par rapport aux autres produits pour la campagne de printemps (1). Cette forte concentration fait écho aux résultats obtenus pour le prélèvement d'eau printemps (1-2) qui présentait, pour rappel, une concentration de 1 885 ng/L de tébuconazole. Ainsi, cette concentration est bien le témoin d'un pic de concentration en tébuconazole ayant eu lieu lors de la prise d'échantillon. En effet, cette concentration, moyennée sur la durée de la mise en place du POCIS (21 jours), conduit à une valeur de 45 ng/g de phase/jour, ce qui signifie que cette valeur très élevée a aussi été très ponctuelle (1 à 2 jour).

En ce qui concerne les concentrations de 210 ng/L et de 62 ng/L relevées via les échantillons d'eau pour les campagnes printemps (3-4) et printemps (4-5), on peut observer, au travers de l'analyse des POCIS, une augmentation de la concentration en tébuconazole entre la campagne printemps (3) (0,5 ng/g de phase/jour) et la campagne printemps (4-5) (2,3 ng/g de phase/jour). Ces concentrations n'étant pas fortement élevées pour l'analyse des POCIS des déploiements printemps (3) et printemps (4), ceci confirme que ces deux concentrations en tébuconazole (210 ng/L et 62 ng/L) correspondent également à un pic de contamination et non à une contamination importante qui perdure dans le temps.

Sur la Figure 67, il ressort que deux substances sont majoritairement présentes : le DEA et l'atrazine. On observe également une variation de la concentration en produits phytosanitaires au cours du temps que l'on peut rattacher à une saisonnalité. On observe donc une diminution de la concentration en DEA et en atrazine au cours de la période hivernale puis une augmentation forte au début de la période printanière suivi durant la période printanière et estivale de diminutions et augmentations successives. Les autres molécules présentes dans les analyses affichent une tendance similaire mais avec des concentrations moins importantes.

L'utilisation des taux de rétention (*Rs*) aurait pu nous permettre d'équilibrer les concentrations cependant, l'utilisation de POCIS étant une méthode semi-quantitative, il n'apparait pas intéressant d'effectuer une telle conversion. Ainsi, seules les variations seront interprétées et la comparaison systématique des concentrations entres POCIS et échantillons d'eau ne sera pas effectuée.





Figure 67 : Résultats des analyses des POCIS. Les produits : de couleur bleue correspondent aux herbicides interdits d'utilisation en France, de couleur verte les herbicides autorisés à l'utilisation en France, de couleur rouge les fongicides interdits d'utilisation en France, de couleur jaune les fongicides autorisés à l'utilisation en France et en noir les insecticides autorisés à l'utilisation en France. Les noms soulignés correspondent aux produits de dégradation du produit les précédant dans la liste du graphique (lecture de gauche à droite et de haut en bas). La concentration en tébuconazole n'est pas affichée à sa vraie valeur afin de ne pas "écraser" toutes les autres valeurs.

FREQUENCES D'OCCURENCE DES PRODUITS PHYTOSANITAIRES

En suivant le même procédé que pour les échantillons d'eau et en utilisant les mêmes valeurs pour définir les classes de taux de détection, l'interprétation de la Figure 68 conduit aux résultats suivants :

- Classe 1. 6 substances avec un taux de détection supérieur à 80% (atrazine, DEA, DIA, simazine, diuron et DET)
- Classe 2. 4 substances avec un taux de détection compris entre 80% et 60% (acetochlore, tébuconazole, isoproturon et oxadixyl)
- Classe 3. 15 substances avec un taux de détection inférieur à 60%

Ainsi, deux produits sont identifiés en plus par rapport aux échantillons d'eau (soit 25 produits) ce qui illustre l'avantage de l'utilisation des POCIS par rapport aux seuls échantillons d'eau prélevés à des instants si peu rapprochés dans le temps (14 jours au minimum).





Figure 68 : Classement des molécules suivant leur fréquence de dépassement du seuil de détection (analyse des POCIS).

En ce qui concerne les taux de quantification via l'analyse des POCIS présentée en Figure 69, 20 produits ont dépassé le seuil de quantification. En comparaison avec la Figure 66 concernant les échantillons d'eau, 8 produits de plus sont quantifiés. En regardant la répartition des classes, certaines observations peuvent être soulignées :

- 1. 4 substances avec un taux de quantification supérieur à 80% (atrazine, DEA, DIA et simazine)
- 2. aucune substance avec un taux de quantification compris entre 80% et 60%
- 3. 16 substances avec un taux de quantification inférieur à 60%

Ainsi, l'atrazine, le DEA, le DIA et la simazine sont, là encore, majoritaires (proche de 100% de quantification). En se basant sur ces observations, nous avons décidé pour la suite de nos travaux (Paragraphes 7.3.2 et 7.3.3) de sélectionner ces 4 produits auxquels nous ajoutons le flazasulfuron qui se trouve constamment compris dans la seconde classe (taux de détection/quantification compris entre 80% et 40%).



Figure 69 : Classement des produits suivant leur fréquence de dépassement du seuil de quantification (analyse des POCIS).

COMPARAISON ENTRE ECHANTILLONNAGES PONCTUELS ET MESURES INTEGRATIVES (POCIS)

Nous avons donc vu que, grâce aux POCIS, il a été possible de quantifier un plus grand nombre de produits phytosanitaires. Grâce à l'accumulation de substances spécifiques, les POCIS permettent de concentrer les polluants et ainsi de dépasser les seuils analytiques. Ainsi, la Figure 70 présente le nombre d'analyses, tous produits confondus, ayant dépassé le seuil de détection ou le seuil de quantification.



Figure 70 : Nombre d'analyses supérieures aux seuils de détection et de quantification sur la totalité des déploiements.

Ainsi, pour les POCIS 457 analyses sont supérieures au seuil de détection contre 339 pour les échantillons d'eau. Quant aux seuils de quantification, on dénombre 366 analyses dépassant les seuils pour les POCIS et seulement 203 pour les échantillons d'eau. Ainsi, les POCIS présentent toujours un nombre d'analyses quantifiées supérieur au nombre des analyses quantifiées provenant des échantillons d'eau.

CONCLUSION SUR LES DEUX METHODES D'ANALYSES

L'interprétation des résultats obtenus par les deux méthodes met en avant l'intérêt des POCIS pour l'étude des variations en continu de la qualité de l'eau naturelle. En effet les échantillonnages ponctuels d'eau avec une période de mesure trop grande sont moins capables de représenter les pics de concentrations d'une pollution accidentelle qui a fort peu de chances d'être échantillonnée. De plus, comme leurs seuils de détection/quantification sont plus élevés que ceux des POCIS, ils sont moins à même de décrire un état de contamination faible de l'eau, notamment une pollution chronique. A contrario, les POCIS, placés dans l'eau durant une durée suffisamment longue (ici 14 jours au minimum), de par leur processus d'accumulation, sont plus sensibles aux très faibles concentrations.

Les parties qui vont suivre (Partie 7.3.2 et 7.3.3) ont pour but d'identifier la relation entre les produits phytosanitaires et les différentes variables environnementales telles que les précipitations, la conductivité et la turbidité. Nous verrons que l'utilisation des POCIS apporte un avantage certain dans cette mise en relation.

7.3.2 ANALYSES PONCTUELLES DES ECHANTILLONS D'EAU

Pour commencer, cette partie regroupe les comparaisons entre les concentrations des produits phytosanitaires obtenues par les méthodes d'échantillonnages recommandées pour l'eau brute (prise d'échantillons d'eau), et les variables environnementales telles que les précipitations, la turbidité et la conductivité électrique.

Par souci de lisibilité, les graphiques de cette partie et dans la partie 7.3.3, ne présentent les concentrations que pour les 5 produits phytosanitaires suivants : atrazine, DEA, DIA, simazine et flazasulfuron. Comme indiqué à la fin du de la partie 7.3.1 ; ces produits ont été sélectionnés car ils présentent de très forts taux de détection/quantification.

Pour la comparaison des concentrations en produits phytosanitaires avec les précipitations, la valeur prise en compte pour chaque variable correspond aux précipitations cumulées durant le déploiement précédent. Dans le cas particulier des déploiements de "mai (1)", "juillet (1)", "hiver (1)" et "printemps (1)", ces derniers n'ont pas de déploiement les précédant directement dans le temps. Pour ces déploiements, le cumul des précipitations mesurées durant les 20 jours précédents a été utilisé.

Pour la comparaison des concentrations en produits phytosanitaires avec la turbidité et la conductivité, nous avons utilisé les valeurs moyennes calculées sur les mêmes périodes que pour les précipitations.

Ainsi, cette partie aura pour objectif de tenter d'évaluer si le lessivage des sols dû aux précipitations entraine à la fois des produits phytosanitaires, et des particules contribuant à la turbidité. Si cette hypothèse est vraie, la turbidité pourrait alors être utilisée comme étant un "proxy" de la pollution de l'eau par les produits phytosanitaires. En sus de cette analyse, nous avons ajouté l'interprétation de la relation entre la conductivité et la concentration en produits phytosanitaires, postulant qu'il pouvait également exister un lien entre ces deux variables.

RECHERCHE D'UNE RELATION ENTRE PRECIPITATIONS ET CONCENTRATION EN PRODUITS PHYTOSANITAIRES DANS LES ECHANTILLONS PONCTUELS D'EAU

Sur la Figure 71, nous avons présenté la pluviométrie et les concentrations des 5 produits phytosanitaires sélectionnés. On note que les pics élevés de précipitations observés durant le déploiement de "hiver (2-3)" et "été (4-5)" (gros triangles jaunes) ne peuvent pas être mis simplement en relation avec une valeur importante des concentrations en produits phytosanitaires. De plus, le pic de DEA (598 ng/L) du déploiement "juillet (1)" ne peut être associé à un pic de pluviométrie observé durant les jours précédents.



Figure 71 : Visualisation des précipitations mesurées au moment de l'échantillonnage et la concentration trouvée via l'analyse des échantillons d'eau.

Cependant, une analyse unique de ces 5 produits ne peut nous permettre de conclure sur le possible lien entre précipitation et concentration en produits phytosanitaires. Pour combler ce manque, la Figure 72 présente les coefficients de corrélation avec les précipitations pour chacun des produits phytosanitaires. Les coefficients de corrélations sont calculés sur les campagnes d'hiver et de printemps/été car pour ces campagnes, il existe plus de 2 mesures et permettent ainsi de calculer un meilleur coefficient de corrélation que pour les campagnes d'hiver et de printemps/été car ces deux campagnes d'hiver et de printemps/été separément pour les campagnes d'hiver et de printemps/été car ces deux campagnes sont séparées par un période durant laquelle nous n'avons pas analysé les concentrations en produits phytosanitaires.



Figure 72: Coefficient de corrélation entre les précipitations et les concentrations en produits phytosanitaires trouvée via l'analyse des échantillons d'eau.

Ainsi, au regard de la Figure 72, il apparait que le DIA présente un coefficient de corrélation de 0,7 pour la campagne d'hiver puis un coefficient de -0,2 pour la campagne de printemps/été. Ainsi, ce produit de décomposition semble être plus en relation avec les précipitations durant l'hiver par rapport au printemps et à l'été. Au regard de la Figure 71, cette forte relation est causée par la forte pluviométrie du déploiement "hiver (2-3)" et faible pluviométrie du déploiement "hiver (5)". Cette forte corrélation ne se retrouvera pas par la suite, ce qui nous conforte dans la présence d'un artefact de corrélation. Cette interprétation peut être appliquée également pour l'acétochlore. Pour les autres produits phytosanitaires, il apparait qu'ils ne possèdent pas un fort coefficient de corrélation avec les précipitations (coefficient de corrélation inférieur à 0,5).

Ainsi, il a été montré qu'en utilisant la méthode d'échantillonnage réglementaire avec une fréquence d'échantillonnage d'environ 2 échantillons par mois, il n'est pas possible d'identifier une relation entre les concentrations en produits phytosanitaires et le cumul établi sur 21 jours des précipitations précédant ce prélèvement.

RECHERCHE D'UNE RELATION ENTRE TURBIDITE ET CONCENTRATION EN PRODUITS PHYTOSANITAIRES DANS LES ECHANTILLONS PONCTUELS D'EAU

Sur la Figure 73, les concentrations en produits phytosanitaires sont mises en relation avec la turbidité de l'échantillon d'eau. Sur la Figure 73, on observe une possible relation inverse entre le DIA et la simazine avec la turbidité durant la campagne d'hiver. Pour la campagne de printemps/été, cette relation semble toujours présente mais plus faible. Ainsi, cette relation induirait l'hypothèse que plus l'eau est turbide, plus la concentration en produits phytosanitaires serait faible.



Figure 73 : Visualisation de la turbidité mesurée au moment de l'échantillonnage et la concentration trouvée via l'analyse des échantillons d'eau.

Pour confirmer cette relation, la Figure 74 présente les coefficients de corrélation pour les campagnes d'hiver et de printemps/été entre les concentrations en produits phytosanitaires et la turbidité moyenne durant les périodes de déploiement.



Figure 74: Coefficient de corrélation entre la turbidité et les concentrations en produits phytosanitaires trouvée via l'analyse des échantillons d'eau.

Au regard de la Figure 74, l'hypothèse émise précédemment d'une relation inverse entre turbidité et concentration en produits phytosanitaires est vérifiée pour la majeur partie des produits phytosanitaires. On peut également noter une corrélation plus importante de la turbidité avec les produits phytosanitaires durant la campagne d'hiver. Cette relation inverse peut s'expliquer par une faible utilisation des produits phytosanitaires en hiver associée à un développement végétal plus faible et ainsi une turbidité de l'eau plus importante due au ruissèlement. Comme dit précédemment, l'utilisation de produits phytosanitaires augmente durant le printemps et l'été alors que les plantes sont développées limitant ainsi le ruissèlement et donc la turbidité. Malheureusement, le coefficient de corrélation ne dépasse que très rarement o,5 et nous ne pouvons pas conclure sur un lien net entre turbidité et concentration en produits phytosanitaires. Par contre la relation inverse tend à apparaître : une corrélation inverse entre précipitations efficaces (celles qui s'infiltrent) et les produits phytosanitaires durant l'hiver.

Il est à noter que la turbidité présente une très forte corrélation durant la campagne d'hiver avec le flazasulfuron et le penconazole. Cette corrélation importante n'est pas à prendre en compte car les variations de concentration de ces deux produits phytosanitaires sont très faibles (±0,2ng/L) et peuvent provenir d'une plus ou moins bonne analyse des échantillons résultant de la turbidité.

Ainsi, au regard de la Figure 73 et de la Figure 74, une relation existe bien entre turbidité et produits phytosanitaires mais cette dernière n'étant que faible, ou même contraire, l'utilisation de la turbidité comme proxy de la présence de produits phytosanitaires, ne peut être utilisée. Cette analyse pourrait être limitée à la région d'étude pour laquelle il pleut régulièrement en hiver quand peu de produits phytosanitaires sont épandus.

RECHERCHE D'UNE RELATION ENTRE CONDUCTIVITE ET CONCENTRATION EN PRODUITS PHYTOSANITAIRES DANS LES ECHANTILLONS PONCTUELS D'EAU

Comme indiqué en début de partie, nous avons jugé intéressant d'analyser la relation pouvant lier la conductivité et les produits phytosanitaires. En effet, la conductivité de l'eau souterraine et celle de l'eau de surface sont très différentes. L'eau de surface possède une conductivité proche 500 μ S/cm alors que la conductivité de l'eau souterraine dans la nappe de la craie est généralement de l'ordre de 600 μ S/cm. Ainsi, lors d'une arrivée d'eau de surface dans l'eau souterraine, la conductivité électrique va diminuer. Grâce à ce phénomène, lors de la seule mesure de la conductivité électrique de l'eau souterraine, il nous est donc possible d'identifier, mais pas de quantifier, les arrivées d'eau de surface. La Figure 75 représente les concentrations des 5 produits phytosanitaires sélectionnés au cours des différents déploiements, ainsi que la conductivité de



l'échantillon d'eau. La mesure de la conductivité moyenne est effectuée sur les mêmes périodes que pour la turbidité et les précipitations.

Figure 75 : Conductivité mesurée au moment de l'échantillonnage et concentration trouvée via l'analyse des échantillons d'eau.

On peut ainsi voir que durant la campagne d'hiver, la conductivité est proche de 600 μ S/cm, ce qui selon notre hypothèse représente le témoin d'un faible apport d'eau de surface. Durant la campagne de printemps/été, la conductivité chute pour atteindre des valeurs proches de 560 μ S/cm, ce qui indique que l'eau mesurée dans le conduit karstique correspond à un mélange l'eau souterraine (fortement conductrice) et d'eau de surface (peu conductrice). Ceci est en accord avec les résultats de la Figure 71 pour laquelle nous avions observé de plus fortes précipitations pour la campagne de printemps/été et de plus faibles précipitations pour la campagne d'hiver. On observe ainsi au niveau des produits phytosanitaires une plus forte concentration durant les périodes de faibles conductivité et a contrario, une plus faible concentration durant les périodes de forte conductivité. Au regard de la Figure 75 il semblerait exister une relation dans les variations de concentration en produits phytosanitaires et les variations de conductivité pour la campagne de printemps/été. Pour confirmer cette hypothèse la Figure 76 présente les coefficients de corrélation pour les concentrations de chaque produit phytosanitaire avec la conductivité pour les campagnes d'hiver et de printemps/été.

La première chose visible sur la Figure 76 est la réelle existence d'une corrélation entre conductivité et produits phytosanitaires pour la campagne de printemps/été. Contrairement à la Figure 74 et la Figure 72, les coefficients de corrélation dépassent régulièrement 0,5 pour la campagne de printemps/été. Cette corrélation s'explique encore une fois par l'apport d'eau important durant cette période. En effet, même si la quantité de produits phytosanitaires utilisée est plus importante par rapport à l'hiver, l'apport en eau des pluviométries est beaucoup plus important. Ainsi, bien que la concentration moyenne en produits phytosanitaires soit plus élevée dans l'eau souterraine, l'apport d'eau de surface va entrainer une dilution ponctuelle des produits phytosanitaires lors des pics d'apport d'eau.

Le fait que la conductivité soit plus en relation avec les produits phytosanitaires par rapport aux précipitations que la turbidité provient du fait qu'elle apporte une information plus robuste sur la quantité d'eau en provenance de la surface. En effet, les précipitations vont être affectées par l'évapotranspiration, le ruissellement et l'infiltration et ainsi subir de nombreuses modifications avant d'affecter l'eau souterraine. Pour la turbidité, cette dernière va se trouver affectée par le couvert végétal et la disponibilité à l'arrachement des particules du sol, et ne sera pas un indicateur aussi robuste que la conductivité pour l'identification d'arrivée d'eau de surface.





Figure 76 : Coefficient de corrélation entre la conductivité et les concentrations en produits phytosanitaires trouvée via l'analyse des échantillons d'eau.

Pour la campagne d'hiver, il est visible sur la Figure 76 de forts coefficients de corrélation négatifs pour 6 produits phytosanitaires (atrazine, DEA, diuron, isoproturon, simazine et terbuthylazine). Ces fortes corrélations sont causées par le pic de concentration de ces produits phytosanitaires mesurés lors de l'échantillonnage ponctuel, associés à une chute de la conductivité durant le déploiement "hiver (5)". Dû au caractère trop peu rapproché des prélèvements ponctuels, il est possible que ces corrélations ne représentent pas bien la réalité pour la campagne d'hiver. En effet, et si le point "hiver (5)" est supprimé pour le calcul des coefficients de corrélation, on obtient une corrélation de l'ordre de +0,2.

Ainsi, on observe bien un lien existant entre la concentration en produits phytosanitaires et la conductivité électrique. Cette relation étant la plus forte des trois étudiées (précipitation, turbidité et conductivité), la conductivité apparait comme le meilleur proxy pour les produits phytosanitaires. Il faut toute fois noter que ce proxy n'est efficient que durant le printemps et l'été, qu'il ne fonctionne pas durant l'hiver, et que son efficacité reste à prouver pour l'automne. Cependant, ce proxy reste fort intéressant car il est efficient durant la période d'utilisation des produits phytosanitaires.

CONCLUSION SUR L'ANALYSE DES ECHANTILLONS PONCTUELS D'EAU

L'analyse d'échantillons d'eau prélevés régulièrement à des intervalles rapprochés (par exemple toutes les 12h) serait l'idéal pour décrire la qualité de l'eau. Malheureusement les moyens techniques actuels ne permettent pas de réaliser ces analyses à un prix accessible pour tous les sites de prélèvement en eau potable de France. Dans notre étude, nous avons poussé l'échantillonnage à une périodicité proche de deux échantillons d'eau par mois. Les résultats d'une telle fréquence d'analyse ont montré que les précipitations, la turbidité et la conductivité possédaient chacun un lien plus ou moins important avec la concentration en produits phytosanitaires. La question initialement posée était la suivante; est-il possible d'utiliser la turbidité comme proxy des produits phytosanitaires ?

Les résultats de cette partie ont clairement montré un lien inverse existant entre la concentration en produits phytosanitaires et la turbidité. Ce lien étant plus fort durant l'hiver en comparaison avec le printemps et l'été, ce qui s'explique par l'importance du couvert végétal influençant la turbidité durant le printemps et l'été. Cependant, cette relation n'a pas été jugée suffisamment robuste pour une utilisation de la turbidité en tant que proxy des produits phytosanitaires.

Une analyse de la relation entre la concentration en produits phytosanitaires et la conductivité a permis l'obtention d'un "semi proxy" des produits phytosanitaires. En effet, la conductivité électrique s'est révélée fortement corrélée avec la concentration en produits phytosanitaires durant le printemps et l'été correspondant à la période d'utilisation la plus importante et la plus diversifiée des produits phytosanitaires mais également avec les précipitations les plus importantes. Ainsi, la conductivité joue le rôle d'un indicateur de facteur de dilution des produits phytosanitaires dans l'eau souterraine.

En synthèse, il apparait que la turbidité ne peut être utilisée comme proxy de la concentration en produits phytosanitaires mais qu'à défaut, il est possible au printemps et en été d'utiliser la conductivité électrique comme "semi proxy" des produits phytosanitaires.

La question restant en suspens est la suivante : est-il possible que les relations précédemment observées soient faussées par les risques de non mesure des variations de concentrations en produits phytosanitaires entre les prélèvements d'échantillons ?

Pour répondre à cette dernière question, nous avons effectué les même analyses que dans la partie précédente mais en utilisant les résultats d'analyses obtenus avec les POCIS.

7.3.3 ANALYSES DES POCIS

Comme précédemment, les figures de cette partie, présentent les moyennes et les écarts types calculés sur les résultats d'analyses obtenus via les triplicas des POCIS. De plus, les conductivités et turbidités indiquées correspondent à la moyenne calculée sur la période de déploiements des POCIS. Les précipitations quant à elles correspondent au cumul sur la période de déploiements du POCIS.

RECHERCHE D'UNE RELATION ENTRE PRECIPITATIONS ET CONCENTRATION EN PRODUITS PHYTOSANITAIRES EN UTILISANT LES POCIS

Comme dans la partie précédente, la Figure 77 présente les concentrations des 5 produits phytosanitaires sélectionnés, relevées par les POCIS. Ces concentrations sont mises en relation avec les précipitations cumulées sur la période du déploiement.

Sur la Figure 77, on peut noter une tendance générale à avoir de fortes concentrations en début de campagne d'échantillonnage puis une diminution progressive au cours des campagnes. Les précipitations ne présentent pas de variations similaires. Cependant, en approfondissant l'observation, on peut voir deux molécules se démarquer dans leur corrélation avec les précipitations : l'atrazine et le DEA. En effet, durant la campagne de printemps/été on note, en passant successivement d'une mesure à une autre, une relation inverse de la concentration en atrazine et DEA avec les précipitations. Lorsque la pluviométrie diminue, la concentration en produits phytosanitaires augmente. Cette relation fait écho à l'observation faite dans la partie précédente. Il semble possible que la concentration moyenne en produits phytosanitaires diminue lorsque les précipitations sont plus importantes à cause du phénomène de dilution. De nouveau, pour permettre une analyse plus détaillée de la relation entre la concentration en produits phytosanitaires et les précipitations, la Figure 78 présente les coefficients de corrélation pour chaque produit phytosanitaire.





Figure 77 : Visualisation des précipitations cumulées mesurées durant les déploiements et les concentrations trouvées via l'analyse des POCIS.

Ainsi, grâce à la Figure 78 notre hypothèse précédemment émise concernant l'anticorrélation de la concentration de l'atrazine et du DEA avec les précipitations est vérifiée. Cette même relation peut également s'appliquer à : l'atrazine, le diuron, l'isoproturon et l'azoxystrobine qui possèdent comme le DEA et l'atrazine un coefficient de corrélation supérieur à -0,5 pour la campagne de printemps/été.



Figure 78 : Coefficient de corrélation entre les précipitations et les concentrations en produits phytosanitaires via l'analyse des POCIS.

Pour l'isoproturon, le terbuthylazine, le tébuconazole et l'imidaclopride, on observe une corrélation supérieure à 0,5 pour la campagne d'hiver. Cette forte corrélation est due à un pic de concentration de ces produits phytosanitaires pour de déploiement "hiver (2)" qui se trouve synchronisé avec le pic de précipitation durant ce même déploiement. Cependant, ce pic de pluviométrie ne peut être mis en relation avec une diminution de conductivité importante, ainsi, ces précipitations ne semblent pas avoir affecté l'eau souterraine

au niveau du conduit karstique de la station d'Yport. Ainsi, il est difficile de pencher sur une réelle relation entre ce pic de concentration en produits phytosanitaires et une augmentation des précipitations efficaces.

Nous avons donc pu observer une corrélation plus importante entre produits phytosanitaires et précipitations lorsque l'on utilise des POCIS par rapport à l'utilisation d'échantillons d'eau. Cette plus forte corrélation est à mettre en relation avec l'aspect moyennant des POCIS qui lisse les variations ponctuelles et ainsi permet une corrélation plus fiable. Cependant, l'utilisation des POCIS a permis de bien mettre en évidence le phénomène de dilution induit par l'apport d'eau en surface. Ainsi même si l'eau de surface va se charger en produits phytosanitaires lors du ruissèlement, il est possible que cette dernière contribue avant tout à diluer la concentration moyenne présente dans l'eau souterraine. Ceci peut nous amener à nous poser une question : l'apport d'eau de surface va-t-il induire une augmentation des concentrations en produits phytosanitaires? Nous allons continuer nos analyses avec la turbidité et la conductivité et nous reviendrons à cette question une fois ces analyses effectuées.

RECHERCHE D'UNE RELATION ENTRE CONDUCTIVITE ET TURBIDITE EN PRODUITS PHYTOSANITAIRES EN UTILISANT LES POCIS

La Figure 79 met en relation les concentrations des 5 produits phytosanitaires sélectionnés avec la turbidité moyenne relevée durant les déploiements. Pour la turbidité, il est plus difficile de trouver une relation avec la concentration en produits phytosanitaires. Le produit semblant le plus corrélé avec la turbidité est l'atrazine. En effet, durant la campagne de printemps, il semblerait que l'atrazine présente une corrélation inverse avec la turbidité mais cette corrélation semble plus discutable durant l'été du fait de la faible variation de la turbidité durant les déploiements "été (1), "été (2)" et "été (3)".





Pour confirmer cette possible corrélation, la Figure 80 présente les coefficients de corrélation des différents produits phytosanitaires avec la turbidité. Il apparait donc bien une anti-corrélation relativement importante (-0.57) entre la concentration en atrazine et la turbidité. On ne retrouve ensuite cette corrélation que pour l'acétochlore. On peut ensuite noter une corrélation plus importante (supérieure à 0,5) pour le DIA, le linuron et le penconazole durant la campagne d'hiver. Cependant, il apparait étrange une telle corrélation de ces trois produits phytosanitaires sans une corrélation des autres produits phytosanitaires. Cette corrélation étrange apparait également pour le chlortoluron et l'imidaclopride avec un coefficient de corrélation supérieur à -0,5 durant la campagne d'hiver sans une telle corrélation de la part des autres produits avec la turbidité.

Ce phénomène de corrélation pourrait être mis en relation avec la forte non linéarité de la turbidité. En effet, la turbidité n'est pas influencée par le seul facteur de l'apport d'eau par les précipitations. La turbidité aura alors une corrélation avec les concentrations en produits phytosanitaires beaucoup plus variable. De plus, on peut ajouter à cette variation, une utilisation différée entre les différents produits phytosanitaires et avec des doses variables d'un produit.



Figure 80 : Coefficient de corrélation entre la turbidité et les concentrations en produits phytosanitaires trouvée via l'analyse des POCIS.

Ainsi, sur ces analyses, il est visible que la relation entre la concentration en produits phytosanitaires et la turbidité sur une période est plus difficile à interpréter que lors de mesures ponctuelles. La turbidité apparait donc comme inutilisable en tant que proxy des contaminations en produits phytosanitaires dans l'eau souterraine en utilisant les POCIS comme méthode d'échantillonnage. De notre interprétation, comme la méthode de mesure (POCIS) est plus fiable que les échantillons d'eaux ponctuelles, nous nous heurtons à une compréhension plus fine du comportement des produits phytosanitaires, compréhension qui n'a pu être menée à son terme dans le temps imparti à ce travail de thèse.

RECHERCHE D'UNE RELATION ENTRE CONDUCTIVITE ET CONCENTRATION EN PRODUITS PHYTOSANITAIRES EN UTILISANT LES POCIS

Pour finir, la Figure 81 met en relation les concentrations des 5 produits phytosanitaires sélectionnés avec la conductivité moyenne relevée durant les déploiements. Comme pour les échantillons d'eau, on observe que la conductivité moyenne de l'eau souterraine est plus forte durant l'hiver et plus faible durant le printemps et l'été. Sur cette figure, il est facile d'observer une corrélation inverse entre la concentration en produits phytosanitaires et la conductivité. Cette relation pourrait être directement liée au processus de dilution des concentrations en produits phytosanitaires de l'eau souterraine par les arrivées d'eau de surface. On peut ainsi appuyer l'hypothèse selon laquelle, l'eau de surface va principalement jouer un rôle de dilution de l'eau souterraine même si cette dernière se charge en produits phytosanitaires lors du ruissèlement. Pour confirmer la relation observée et permettre de l'étendre aux autres produits phytosanitaires, la Figure 82 présente les coefficients de corrélation entre les concentrations en produits phytosanitaires en produits phytosanitaires de l'eau souters produits phytosanitaires, la Figure 82 présente les coefficients de corrélation entre les concentrations en produits phytosanitaires et la conductivité électrique.





Figure 81 : Visualisation de la conductivité moyenne mesurée durant les déploiements et les concentrations trouvées via l'analyse des POCIS.

On peut voir sur la Figure 82 que la relation entre la concentration en atrazine et la conductivité est bel est bien présente et peut s'appliquer à de nombreux produits phytosanitaires; l'atrazine, le DEA, le DIA, le diuron, le DCPU, le DCPMU, l'isoproturon, le chlortoluron, le flazasulfuron, le linuron, le propyzamide, le dimethomorphe, l'epoxiconazole, le metalaxyl et le tébuconazole. On notera ensuite une forte corrélation inverse (supérieure à -0,5) de l'acétochlore, de l'alachlor et du chlortoluron pour la campagne d'hiver. Cette corrélation inverse est à mettre en relation avec une utilisation plus précoce de ces produits vers la fin de l'hiver en vue d'une préparation des terres agricoles par rapport aux autres produits tels que l'atrazine qui présente un pic d'utilisation au début du printemps. Attention, il est important de rappeler qu'une période de deux mois sans mesure sépare les campagnes d'hiver et de printemps/été.



Figure 82 : Coefficient de corrélation entre la conductivité et les concentrations en produits phytosanitaires trouvées via l'analyse des POCIS.

Ainsi, l'utilisation de la conductivité comme semi proxy de la concentration en produits phytosanitaires apparait de nouveau comme possible. Cette dernière nous informe alors plus sur la dilution des concentrations en produits phytosanitaires de l'eau souterraine que sur les apports quantitatifs.

CONCLUSION SUR L'ANALYSE DES POCIS

L'analyse des POCIS avait pour objectif de centrer l'analyse sur les relations des concentrations moyennes de l'eau souterraine avec les précipitations, la turbidité et la conductivité. Un questionnement avait été soulevé lors de l'analyse des échantillons d'eau et de leur mise en relation avec les précipitations, la turbidité et la conductivité : est-il possible que les relations précédemment observées (partie 7.3.2) soient faussées par les risques de non mesures des variations de concentrations en produits phytosanitaires entres les prélèvements d'échantillons ponctuels ?

Les résultats de cette partie ont clairement montré un lien inverse existant entre la concentration des produits phytosanitaires et les précipitations contrairement à la partie 7.3.2 qui n'avait pas permis de mettre en relation concentration en produits phytosanitaires et précipitations. Le lien nouvellement identifié semble plus fort et provenir de l'effet de régularisation des POCIS. Ainsi, il a été émis l'hypothèse d'une dilution des concentrations en produits phytosanitaires par les arrivées d'eau de surface.

Par la suite, contrairement à la conclusion de la partie 7.3.2 , il n'a pas été possible de mettre en avant une relation claire entre turbidité moyenne et concentration en produits phytosanitaires moyenne.

Pour finir, la dernière analyse a mis en avant, comme précédemment dans la partie 7.3.2, que la concentration en produits phytosanitaires présentait une relation avec la conductivité électrique. Ainsi, lors d'apports d'eau de surface, la conductivité électrique diminue et la concentration en produits phytosanitaires, sous l'effet d'une dilution, diminue également.

7.4 CONCLUSION DU CHAPITRE 7

Ce chapitre avait pour objectif d'identifier un proxy des concentrations en produits phytosanitaires au moyen de deux méthodes d'acquisition de données en concentration en produits phytosanitaires : les échantillonnages ponctuels d'eau et les échantillonneurs passifs que sont les POCIS. Ce proxy avait pour but de permettre un suivi des variations de concentrations en produits phytosanitaires sans nécessiter l'augmentation de la fréquence des analyses ponctuelles des produits phytosanitaires dans l'eau souterraine.

Pour répondre à ces deux objectifs, plusieurs campagnes d'analyses mêlant échantillonnages d'eau et déploiements de POCIS ont été effectuées. C'est en tout 42 échantillons d'eau (21 échantillons avec un réplica) et 51 POCIS (17 triplicas de POCIS mis en places) qui ont été analysés. Il n'a pas été possible de réaliser un échantillonnage durant l'automne dans le temps imparti. Nous considérons que cette campagne devrait toutefois être réalisée afin de compléter notre étude.

L'interprétation des résultats obtenus par l'analyses des échantillons d'eau et de POCIS nous a permis de montrer que les deux méthodes d'échantillonnage (échantillons d'eau et POCIS) présentent chacune leurs avantages et leurs inconvénients. Les échantillons d'eau, permettent une analyse des concentrations ponctuelles, analyse ne pouvant pas être effectuée avec les POCIS du fait de leur déploiement durant une période de 15 à 20 jours. Ainsi, les POCIS permettent d'observer une concentration cumulée mais ne sont pas adaptés pour détecter des pics de concentrations.

Il a également été mis en évidence grâce au POCIS, la possibilité de détecter et de quantifier des concentrations très faibles en produits phytosanitaires non détectables ou quantifiables avec des échantillons d'eau ponctuels. Ainsi, les échantillons d'eau ont relevé la présence de 23 produits dans l'eau souterraine au cours des différentes campagnes alors que 25 produits ont pu être détectés grâces au POCIS. De plus, seulement 12 produits ont pu être quantifiés avec les échantillons d'eau lorsque 20 produits ont pu être quantifiés grâces aux POCIS. Ainsi, les POCIS ont montré leur large avantage pour la caractérisation de la qualité moyenne de

l'eau souterraine face aux échantillons d'eau. En revanche, les échantillons d'eau permettent d'identifier les variations rapides de concentration en produits phytosanitaires.

La conclusion de cette partie est donc que l'association des deux pratiques d'échantillonnage représente la meilleure alternative et que les deux méthodes sont complémentaires.

Par la suite, la mise en relation des concentrations en produits phytosanitaires issues de l'analyse des échantillons d'eau et des POCIS avec les précipitations, la turbidité et la conductivité a permis de pointer plusieurs résultats que nous pouvons qualifier d'intermédiaires : la conductivité semble être un meilleur indicateur de concentration en polluants que la turbidité, mais il faudrait encore croiser cette information avec celle des précipitations et effectuer ce croisement sur plus d'années pour conclure définitivement. En effet, toutes ces grandeurs sont fortement variables en fonction des produits, des saisons et des pratiques agricoles qui peuvent varier au cours des années, et auxquelles nous n'avons pu avoir accès.

Nous pensons donc dans un premier temps à la réalisation d'analyses des précédents proxys pour les différentes molécules identifiées. A la suite de ces travaux, nous pensons que l'utilisation d'un multi-proxy serait plus judicieuse étant donnée la variabilité des résultats obtenus. Bien entendu, il serait fort intéressant de mieux connaitre les pratiques agricoles opérant sur notre terrain d'étude et à cette fin, nous envisageons la réalisation d'une enquête auprès des agriculteurs afin de connaitre plus en détail les variations des pratiques agricoles sur notre terrain d'étude.

CONCLUSION

La turbidité en Normandie est un problème qui concerne plus de 50% des installations chargées de la production d'eau potable. La turbidité est communément admise comme possédant un fort lien avec la concentration en produits phytosanitaires. Effectuer la prévision de la turbidité présente donc un double enjeu; (i) diminuer les coûts de traitement de l'eau engendrés par les matières en suspension, (ii) diminuer les risques de contamination de l'eau potable par des produits phytosanitaires. Nos travaux se sont concentrés sur le cas particulier du captage d'alimentation en eau potable d'Yport. Cependant, il est possible d'extrapoler la majeure partie de nos résultats aux autres installations de productions d'eau potable en Normandie. En effet, l'homogénéité de l'occupation des sols ainsi que des formations géologiques superficielles et souterraines n'engendrent pas une forte variabilité des fonctionnements hydrogéologiques. Ainsi ces travaux de thèse avaient pour objectif d'étudier la turbidité afin de permettre : (i) la réalisation de modèles prédictifs de la turbidité dans l'eau souterraine (ii) l'identification d'un indicateur des contaminations en produits phytosanitaires.

Tout d'abord nous avons montré dans le chapitre 1, au travers d'un état de l'art, que la turbidité pouvait apporter de précieuses informations sur le fonctionnement d'un hydrosystème, mais que la mesure de cette variable était sujette à de nombreuses difficultés. En définissant la turbidité comme l'indicateur de la présence de particules dans l'eau, il est apparu que les différentes propriétés de ces matières en suspension (forme, taille, couleur, composition, ...) entraînaient des turbidités variables pour une même concentration en matière en suspension. Une fois ce problème exposé, nous avons montré que les matières en suspensions pouvaient être d'origine souterraine ou bien superficielle. Ce sont alors ces matières d'origine superficielle qui peuvent être associées à des arrivées d'eau chargées en produits phytosanitaires. Ces interprétations nous ont amené à penser que le modèle prévisionnel de la turbidité pouvait ainsi présenter une double utilité. La dernière étape du chapitre 1 a permis de faire un recensement des modèles existants pour la prévision de la turbidité. Nos travaux de prévision ont alors montré leur aspect novateur au sein de la communauté scientifique. En effet, très peu de modèles se sont intéressés à la prédiction de la turbidité et, à notre connaissance, aucun modèle n'a été trouvé pour la prévision infra-journalière de la turbidité de l'eau souterraine en utilisant les données de précipitation.

Suite à cet état de l'art, nous avons effectué dans le chapitre 2 un descriptif; climatique, hydrologique, hydrogéologique, et d'occupation des sols du bassin versant d'Yport afin d'apporter les connaissances suffisantes pour la bonne compréhension et la bonne interprétation de nos futurs résultats. Nous avons ainsi pu voir que notre site d'étude présentait une variabilité des précipitations influencée par un climat océanique au nord-ouest et continental au sud-est. L'occupation du sol (75% d'origine agricole) et ses propriétés (croûte de battance) associée aux formations géologiques (bétoires et karsts) entraînent une forte vulnérabilité de la ressource en eau souterraine. Par la suite, nous avons effectué le recensement, la description et la critique des données à notre disposition; pluviométrie, conductivité, turbidité, niveau piézométrique et concentration en produits phytosanitaires. Nous avons émis l'hypothèse que la périodicité faible des analyses des produits phytosanitaires entraînait une forte sensibilité à des variations ponctuelles non représentatives de la qualité de l'eau souterraine. Pour pouvoir aller plus en avant dans nos travaux, nous avons présenté dans le chapitre 3, les outils et méthodes que nous avons utilisés par la suite : corrélation, décomposition en ondelettes, analyse multirésolution, réseaux de neurones, POCIS, analyse des produits phytosanitaires.

Les premiers travaux utilisant les données à notre disposition ont ensuite été présentés dans le chapitre 4. Cette partie de nos travaux nous a permis en grande partie d'effectuer une sélection des données pouvant être utilisées pour la modélisation. Avec l'étude des précipitations et leur mise en relation avec les variations de turbidité, nous avons mis en avant que seules 6 des 11 stations pluviométriques pouvaient être utilisées dans nos modèles. Ce sont ainsi les stations de; Froberville, Annouville, Goderville, Anglesqueville, Mannevilette et Ethainus qui ont été sélectionnées. L'utilisation d'analyses corrélatoires nous a révélé un temps de réponse du bassin d'Yport de l'ordre de 36 heures. En revanche, un ré-échantillonnage des données nous a permis d'identifier l'importance de l'humidité du bassin sur l'occurrence des pics de turbidité. L'étude de la fréquence d'échantillonnage des produits phytosanitaires a conforté notre hypothèse selon laquelle la fréquence et le mode d'échantillonnage entrainaient une forte sensibilité des observations par rapport aux variations ponctuelles des concentrations en produits phytosanitaires.

C'est avec toutes ces informations que nous avons pu effectuer la conception des modèles neuronaux présentés dans le chapitre 5. C'est ainsi 32 modèles qui ont été dimensionnés et testés pour la prévision de la turbidité. Au travers de ces 32 modèles, l'observation majeure était la capacité des réseaux de neurones à prédire la turbidité à 12h et 24h. Nous avons également pu observer que nos modèles présentaient de bien meilleurs résultats pour la prédiction d'un dépassement de seuil plutôt que la prédiction de l'occurrence et du maximum d'un événement de turbidité. Parmi ces résultats, nous avons pu observer que l'utilisation de pluies spatialisées augmentait la complexité des modèles sans toutefois augmenter leur qualité, voire même dans certains cas, la diminuer. Il est également apparu que les modèles dirigés et les modèles non dirigés pouvaient être associés pour obtenir de meilleurs résultats de prédiction. En ce qui concerne les données, l'utilisation de données horaires a montré un plus fort intérêt par rapport à l'utilisation de données à 6 heures. L'utilisation de ces différents pas de temps nous a permis, dans le cas des données à 6 heures, de montrer une incertitude de prédiction beaucoup moins importante lors de l'utilisation dans le cas des meilleurs modèles semblait encore importante.

C'est ainsi que dans le chapitre 6, nous nous sommes proposé d'améliorer les modélisations en associant les réseaux de neurones avec la décomposition en ondelettes continue. En travaillant, dans un premier temps, sur le résidu de modélisation, il a été mis en avant que ce dernier comportait une part d'information non prise en compte par le modèle. Par la suite nous avons, avant de nous lancer dans la modélisation, mis en avant l'impact de l'effet de bord engendré par la décomposition en ondelettes continue, dans le cas de l'utilisation des signaux décomposés en entrée des modèles. Il est ressorti de ces travaux qu'une décomposition trop importante de la turbidité ne permettait pas une amélioration des résultats de prévision. En revanche, la réalisation de modèles en lien avec un assemblage de différents niveaux de décomposition représentait une voie d'amélioration pour les modèles prévisionnels de la turbidité. Ainsi, ces travaux préliminaires sur le couplage des réseaux de neurones et de la décomposition en ondelettes ont prouvé l'intérêt de poursuivre les recherches.

Le chapitre final de ce manuscrit (chapitre 7) nous a permis de présenter les résultats obtenus au travers des campagnes d'échantillonnages de produits phytosanitaires ayant eu lieu pendant cette thèse. Au travers de ce chapitre, nous avons pu montrer l'utilité de réaliser des échantillonnages à l'aide de POCIS combinés avec les méthodes d'échantillonnages ponctuels. Les premières observations ont montré que les échantillonneurs passifs (POCIS) permettaient d'identifier un plus grand nombre de produits phytosanitaires dans l'eau souterraine en comparaison avec l'échantillonnage ponctuel : pratique actuellement utilisée pour l'analyse de la qualité de l'eau souterraine. Par la suite, les campagnes effectuées nous ont permis d'observer une saisonnalité des concentrations des produits phytosanitaires en relation avec les pratiques agricoles. La comparaison des évolutions des concentrations en produits phytosanitaires avec les évolutions de la turbidité nous a amené à conclure que la turbidité ne pouvait pas être utilisée pour prédire la présence de produits phytosanitaires. En revanche, il est apparu que la conductivité se présentait comme un bon candidat pour être utilisée comme indicateur des pollutions de surface, couplée avec la prise en compte des pratiques agricoles.

Pour finir, ce travail a ouvert plusieurs chantiers novateurs : prévision de la turbidité par réseaux de neurones, étude des liens entre les produits phytosanitaires et les variables environnementales comme la pluie, la conductivité électrique, la turbidité. Nous avons proposé plusieurs pistes intéressantes qui restent à approfondir tant pour les travaux de modélisation associant les réseaux de neurones et l'analyse multirésolution que pour la définition de la conductivité électrique comme élément indicateur de la contamination par des produits phytosanitaires.

Enfin, nos travaux ont suggéré un certain nombre de pistes de recherche que nous considérons particulièrement intéressantes à développer. Parmi celles-ci: (i) explorer de façon plus détaillé le lien entre

contenu informatique des données et modélisation par apprentissage statistique (diversifier les méthodes de décomposition ainsi que de modélisation), (ii) développer une approche multi-proxy pour permettre l'identification des contaminations par les produits phytosanitaires et enfin (iii) développer une approche de modélisation prédictive de l'occurrence de ce type de contaminations.

BIBLIOGRAPHIE

Abbott, D. C., Harrison, R. B., Tatton, J. G., & Thomson, J. (1965). Organochlorine pesticides in the atmospheric environment. Nature, 208(5017), 1317-1318. DOI: 10.1111/j.1749-6632.1969.tb15835.x

Adamowski, J., & Sun, K. (2010). Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. Journal of Hydrology, 390(1), 85-91. DOI: 10.1016/j.jhydrol.2010.06.033

Aderhold, D., & Nordmeyer, H. (1995). Leaching of herbicides in soil macropores as a possible reason for groundwater contamination.

Alvarez, D. A., Petty, J. D., Huckins, J. N., Jones-Lepp, T. L., Getting, D. T., Goddard, J. P., & Manahan, S. E. (2004). Development of a passive, in situ, integrative sampler for hydrophilic organic contaminants in aquatic environments. Environmental Toxicology and Chemistry: An International Journal, 23(7), 1640-1648. DOI: 10.1897/03-603

Amraoui, F., Razack, M., & Bouchaou, L. (2003). Turbidity dynamics in karstic systems. Example of Ribaa and Bittit springs in the Middle Atlas (Morocco). Hydrological sciences journal, 48(6), 971-984. DOI: 10.1623/hysj.48.6.971.51418

Anctil, F., & Coulibaly, P. (2004). Wavelet analysis of the interannual variability in southern Québec streamflow. Journal of climate, 17(1), 163-173. DOI: 10.1175/1520-0442(2004)017<0163:WAOTIV>2.0.CO;2

Artigue, G. (2012). Prévision des crues éclair par réseaux de neurones: généralisation aux bassins non jaugés (Doctoral dissertation, Montpellier 2).

Artigue, G., Johannet, A., Borrell, V., & Pistre, S. (2012). Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin (southern France). Natural Hazards and Earth System Sciences, 12(11), 3307-3324. DOI: 10.5194/nhess-12-3307-2012

Atteia, O., & Kozel, R. (1997). Particle size distributions in waters from a karstic aquifer: from particles to colloids. Journal of Hydrology, 201(1-4), 102-119. DOI: 10.1016/S0022-1694(97)00033-4

Auckenthaler, A., Raso, G., & Huggenberger, P. (2002). Particle transport in a karst aquifer: natural and artificial tracer experiments with bacteria, bacteriophages and microspheres. Water Science and Technology, 46(3), 131-138. DOI: 10.2166/wst.2002.0072

Audra, P., Mocochain, L., Camus, H., Gilli, É., Clauzon, G., & Bigot, J. Y. (2004). The effect of the Messinian Deep Stage on karst development around the Mediterranean Sea. Examples from Southern France. Geodinamica acta, 17(6), 389-400. DOI: 10.3166/ga.17.389-400

Audus, L. J. (1949). The biological detoxication of 2: 4-dichlorophenoxyacetic acid in soil. Plant and Soil, 2(1), 31-36. DOI: 10.1007/BF01344145

Bakalowicz, M. (2008, May). Le milieu karstique: études et perspectives, identification et caractérisation de la ressource. In Colloque Hydrogéologie et Karst au travers des travaux de Michel Lepillers (pp. 11-27).

Barriuso, E., Calvet, R., Schiavon, M., & Soulas, G. (1996). Les pesticides et les polluants organiques des sols. Etude et gestion des sols, 3(4), 279-296.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. IEEE Transactions on Information theory, 39(3), 930-945. DOI: 10.1109/18.256500

Beaudeau, P., Leboulanger, T., Lacroix, M., Hanneton, S., & Wang, H. Q. (2001). Forecasting of turbid floods in a coastal, chalk karstic drain using an artificial neural network. Ground Water, 39(1), 109-118. DOI: 10.1111/j.1745-6584.2001.tb00356.x

Beestman, G. B., & Deming, J. M. (1974). Dissipation of acetanilide herbicides from soils. Agronomy Journal, 66(2), 308-311. DOI: 10.2134/agronj1974.00021962006600020035x

Berthet, L., Perrin, C., Andréassian, V., Tangara, M., Loumagne, C., Ramos, M. H., Le Moine, N., Lerat, J. & Javelle, P. (2008, November). Comment passer d'un modèle hydrologique à un système de prévision des crues? Ecueils liés à la structure des modèles et aux échelles d'espace et de temps. In Colloque SHF-191e CST Prévisions hydrométéorologiques (pp. 10-p).

Bin Omar, A. F., & Bin MatJafri, M. Z. (2009). Turbidimeter design and analysis: a review on optical fiber sensors for the measurement of water turbidity. Sensors, 9(10), 8311-8335. DOI: 10.3390/s91008311

Bishop, J. K. (1986). The correction and suspended particulate matter calibration of Sea Tech transmissometer data. Deep Sea Research Part A. Oceanographic Research Papers, 33(1), 121-134. DOI: 10.1016/0198-0149(86)90111-1

Boiffin, J., Pary, F., & Peyré, Y. (1986). Systèmes de production, systèmes de culture et risques d'érosion dans le Pays de Caux.

Bopp, S. (2004). Development of a passive sampling device for combined chemical and toxicological long-term monitoring of groundwater (Doctoral dissertation, UFZ-Umweltforschungszentrum Leipzig-Halle GmbH).

Bowers, D. G., & Binding, C. E. (2006). The optical properties of mineral suspended particles: A review and synthesis. Estuarine, Coastal and Shelf Science, 67(1), 219-230. DOI: 10.1016/j.ecss.2005.11.010

Bracq, P., Hanich, L., Delay, F., & Crampon, N. (1992). Mise en évidence par traçage d'une relation rapide, entre la surface et les eaux souterraines, liée à des phénomènes de dissolution dans la craie du Boulonnais (nord de la France). Bulletin de la Société géologique de France, 163(2), 195-203.

Bunt, J. A., Larcombe, P., & Jago, C. F. (1999). Quantifying the response of optical backscatter devices and transmissometers to variations in suspended particulate matter. Continental shelf research, 19(9), 1199-1220. DOI: 10.1016/S0278-4343(99)00018-7

Calba, F. (1980). Hydrogéologie du karst crayeux du pays de Caux (France): étude de deux bassins (Doctoral dissertation, Paris).

Capriel, P., Haisch, A., & Khan, S. U. (1985). Distribution and nature of bound (nonextractable) residues of atrazine in a mineral soil nine years after the herbicide application. Journal of agricultural and food chemistry, 33(4), 567-569. DOI: 10.1021/jf00064a004

Chedeville, S. (2014). Étude sédimentaire d'une carotte faite au captage d'Yport. (Seine Maritime, France). (Rapport d'étude).

Chédeville, S., Laignel, B., Rodet, J., Todisco, D., Fournier, M., Dupuis, E., ... & Hanin, G. (2015). The sedimentary filling in the chalk karst of the Northwestern Paris Basin (Normandy, France): Characterization, origin and hydro-sedimentary behaviour. Zeitschrift für Geomorphologie, 59(1), 79-101. DOI: 10.1127/0372-8854/2014/0139

Chesters, G., Simsiman, G. V., Levy, J., Alhajjar, B. J., Fathulla, R. N., & Harkin, J. M. (1989). Environmental fate of alachlor and metolachlor. In Reviews of environmental contamination and toxicology (pp. 1-74). Springer New York. DOI: 10.1007/978-1-4684-7092-5_1

Cooper, R. J., Jenkins, J. J., & Curtis, A. S. (1990). Pendimethalin volatility following application to turfgrass. Journal of Environmental Quality, 19(3), 508-513. DOI: 10.2134/jeq1990.00472425001900030026x

Coulibaly, P., & Burn, D. H. (2004). Wavelet analysis of variability in annual Canadian streamflows. Water Resources Research, 40(3). DOI: 10.1029/2003WR002667

Coulibaly, P., & Burn, D. H. (2005). Spatial and temporal variability of Canadian seasonal streamflows. Journal of Climate, 18(1), 191-210. DOI: 10.1175/JCLI-3258.1

Creed, E. L., Pence, A. M., & Rankin, K. L. (2001). Inter-comparison of turbidity and sediment concentration measurements from an ADP, an OBS-3, and a LISST. In OCEANS, 2001. MTS/IEEE Conference and Exhibition (Vol. 3, pp. 1750-1754). IEEE. DOI: 10.1109/OCEANS.2001.968098

Darras, T. (2016) Prévision des crues rapides par apprentissage statistique. (Doctoral dissertation, Alés)

Darras, T., Borrell Estupina, V., Vayssade, B., Johannet, A., & Pistre, S. (2015). Identification of spatial and temporal contributions of rainfalls to flash floods using neural network modelling: case study on the Lez basin (southern France). Hydrology and Earth System Sciences, 19(10), 4397-4410. DOI: 10.5194/hess-19-4397-2015

Darras, T., Johannet, A., Vayssade, B., Kong-A-Siou, L. & Pistre S. (2014). Influence of the Initialization of Multilayer Perceptron for Flash Floods Forecasting: How Designing a Robust Model, conference paper ITISE 2014 (Grenade). pp. 687-698.

Daubechies, I. (1992). Ten lectures on wavelets. Society for industrial and applied mathematics.

Delay, F., Porel, G., & de Marsily, G. (1997). Predicting solute transport in heterogeneous media from results obtained in homogeneous ones: an experimental approach. Journal of contaminant hydrology, 25(1-2), 63-84. DOI: 10.1016/S0169-7722(96)00020-4

Downing, J. P., Sternberg, R. W., & Lister, C. R. B. (1981). New instrumentation for the investigation of sediment suspension processes in the shallow marine environment. Marine Geology, 42(1), 19-34. DOI: 10.1016/0025-3227(81)90156-0

Dreyfus, G. (2005). Neural networks: methodology and applications. Springer Science & Business Media, 497.

Dreyfus, G., Samuelides, M., Martinez, J.-M., Gordon, M.-B., Badran, F., Thiria, S., Hérault, L. (2008). Apprentissage statistique. Eyrolles. 450

Durand, F. & Petit, V. (1997). Guide pour l'aménagement des captages destinés à l'alimentation en eau potable et leur périmètre de protection immédiate. Rapport BRGM/RR-39473-FR. 38 p.

Dussart-Baptista, L., Massei, N., Dupont, J. P., & Jouenne, T. (2003). Transfer of bacteria-contaminated particles in a karst aquifer: evolution of contaminated materials from a sinkhole to a spring. Journal of Hydrology, 284(1-4), 285-295. DOI: 10.1061/40698(2003)17

El Azzi, D. (2012). Transfert de polluants organiques et inorganiques dans les hydrosystèmes en période de crue: interactions avec les matières en suspension et la matière organique (Doctoral dissertation, Université de Toulouse, Université Toulouse III-Paul Sabatier).

El Boukili, A. (2016). Contribution à l'étude de la photolyse et l'adsorption de l'herbicide mefenpyrdiethyl dans différents sols céréaliers marocains. (Thèse de doctorat, Rabat)

El Janyani, S., Dupont, J. P., Massei, N., Slimani, S., & Dörfliger, N. (2014). Hydrological role of karst in the Chalk aquifer of Upper Normandy, France. Hydrogeology journal, 22(3), 663-677. DOI: 10.1007/s10040-013-1083-z

Foufoula-Georgiou, E., Kumar, P., Mukerji, T., & Mavko, G. (1995). Wavelets in geophysics. Pure and Applied Geophysics, 145(2), 374-375.

Fournier, M., Massei, N., Bakalowicz, M., Dussart-Baptista, L., Rodet, J., & Dupont, J. P. (2007). Using turbidity dynamics and geochemical variability as a tool for understanding the behavior and vulnerability of a karst aquifer. Hydrogeology Journal, 15(4), 689-704. DOI: 10.1007/s10040-006-0116-2

Foy, C. L. (1964). Volatility and tracer studies with alkylamino-s-triazines. Weeds, 103-108. DOI: 10.2307/4040606

Fusi, P. (1991). Adsorption et décomposition catalytique des pesticides sur les minéraux argileux-XXIème Congrés du Groupe français des Pesticides. Nancy Brabois, 45-72.

Gaillard, T., Hauchard, E., & Roux, J. C. (2012) Les Fontaines d'Yport (Seine-Maritime), émergences majeures du littoral normand crayeux: Exploitation et vulnérabilité de la ressource en eau. Ressources et gestion des aquifères littoraux. Cassis 2012. 235-244.

Gardner, W. D., Biscaye, P. E., Zaneveld, J. R. V., & Richardson, M. J. (1985). Calibration and comparison of the LDGO nephelometer and the OSU transmissometer on the Nova Scotian Rise. Marine Geology, 66(1), 323-344. DOI: 10.1016/0025-3227(85)90037-4

Gartner, J. W. (2004). Estimating suspended solids concentrations from backscatter intensity measured by acoustic Doppler current profiler in San Francisco Bay, California. Marine Geology, 211(3), 169-187. DOI: 10.1016/j.margeo.2004.07.001

Geman, S., Bienenstock, E., & Doursat, R. (1992). Neural networks and the bias/variance dilemma. Neural computation, 4(1), 1-58. DOI: 10.1162/neco.1992.4.1.1

Glotfelty, D. E., Taylor, A. W., Turner, B. C., & Zoller, W. H. (1984). Volatilization of surface-applied pesticides from fallow soil. Journal of Agricultural and Food Chemistry, 32(3), 638-643. DOI: 10.1021/jf00123a053

Glysson, G. D., & Gray, J. R. (2002). Turbidity and Other Sediment Surrogates Workshop (Proceedings. Reno, Nevada, USA, April/May 2002). US Geological Survey Circular, 1250.

Gray, J. R., & Gartner, J. W. (2009). Technological advances in suspended-sediment surrogate monitoring. Water resources research, 45(4). DOI: 10.1029/2008WR007063

Grébil, G., Novak, S., Perrin-Ganier, C., & Schiavon, M. (2001). La dissipation des produits phytosanitaires appliqués au sol. Ingénieries Numéro Spécial «Phytosanitaires: transfert, diagnostic et solutions correctives, 31-44.

Gril, J. J., Gouy, V., & Carluer, N. (1999). Processus de transfert superficiel des produits phytosanitaires, de la parcelle au bassin versant. La Houille Blanche, (5), 76-80.

Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. IEEE transactions on Neural Networks, 5(6), 989-993. DOI: 10.1109/72.329697

Haghe, J. P. (1998). Les eaux courantes et l'état en France (1789-1919). Du contrôle institutionnel a la fétichisation marchande (Thèse de doctorat, Paris).

Hanin, G. (2010). Contrôles structural et hydrogéologique sur la dynamique d'un champ captant en contexte crayeux karstique et sa sensibilité aux variations du signal climatique: Implications en matière de vulnérabilité de la ressource (Doctoral dissertation, Rouen).

Hardenbergh, W. A. (1938). Water Supply and Purification, Internat. Textbook Co., Scranton, Pa, 280.

Hatcher, A., Hill, P., Grant, J., & Macpherson, P. (2000). Spectral optical backscatter of sand in suspension: effects of particle size, composition and colour. Marine Geology, 168(1), 115-128. DOI: 10.1016/S0025-3227(00)00042-6

Hauchard, E. (2001). De la dynamique non lineaire a la dynamique du relief en geomorphologie. Application aux bassins-versants de la marge nord occidentale du bassin de paris (Doctoral dissertation, Rouen).

Hauchard, E. (2002). Plan de prévention des risques d'inondation par débordement et ruissellement. Méthodologie pour la définition de l'aléa ruissellement: Approche hydrogéomorphologique, DDE, 76.

Hauchard, E., & Laignel, B. (2008). Morphotectonic evolution of the north-western margin of the Paris Basin. Zeitschrift für Geomorphologie, 52(4), 463-488. DOI: 10.1127/0372-8854/2008/0052-0463

Hauchard, E., Laignel, B., & Delahaye, D. (2002). Proposition d'un nouveau schéma structural du Nord-Ouest du bassin de Paris reposant sur l'analyse fractale des réseaux de thalwegs et les données récentes de la géologie régionale. Comptes Rendus Geoscience, 334(4), 295-302. DOI: 10.1016/S1631-0713(02)01733-9

Haar, A. (1910). Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, 69(3), 331-371. DOI: 10.1007/BF01456326

Herman, E. K., Tancredi, J. H., Toran, L., & White, W. B. (2007). Mineralogy of suspended sediment in three karst springs. Hydrogeology Journal, 15(2), 255-266. DOI: 10.1007/s10040-006-0108-2

Holdaway, G. P., Thorne, P. D., Flatt, D., Jones, S. E., & Prandle, D. (1999). Comparison between ADCP and transmissometer measurements of suspended sediment concentration. Continental shelf research, 19(3), 421-441.

Hole, J.P., & Roux, J.C. (1978). Vulnérabilité aux pollutions du bassin hydrogéologique des sources d'Yport. Rapport BRGM. 81p.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366. DOI: 10.1016/0893-6080(89)90020-8

Houria, B., Mahdi, K., & Zohra, T. F. (2014). PSO-ANNs based suspended sediment concentration in Ksob basin, Algeria. Journal of Engineering and Technology Research, 6(8), 129-136. DOI: 10.5897/JETR2014.0524

Hubbard, B. B. (1995). Ondes et ondelettes: la saga d'un outil mathématique. Pour la science.

Huckins, J. N., Manuweera, G. K., Petty, J. D., Mackay, D., & Lebo, J. A. (1993). Lipid-containing semipermeable membrane devices for monitoring organic contaminants in water. Environmental science & technology, 27(12), 2489-2496. DOI: 10.1021/es00048a028

Husson, G. P. (1996). Historique de l'alimentation en eau potable de la ville de Paris. Journal européen d'hydrologie, 27(2), 97-108.

Ibrahim, I. (2013). Etude de l'applicabilité des échantillonneurs passifs POCIS et Chemcatcher pour le suivi des pesticides en milieux aquatiques (Doctoral dissertation, Ecole Nationale Supérieure des Mines de Saint-Etienne).

Iglesias, C., Torres, J. M., Nieto, P. G., Fernández, J. A., Muñiz, C. D., Piñeiro, J. I., & Taboada, J. (2014). Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain. Water resources management, 28(2), 319-331. DOI: 10.1007/s11269-013-0487-9

Jacquet, J. (1960). Applications de la méthode de l'hydrogramme unitaire à quelques cours d'eau français. La Houille Blanche, 857-871.

Jago, C. F., & Bull, C. F. J. (2000). Quantification of errors in transmissometer-derived concentration of suspended particulate matter in the coastal zone: implications for flux determinations. Marine Geology, 169(3), 273-286. DOI: 10.1016/S0025-3227(00)00079-7

Jaunky, A. (2000). Etude de la volatilisation des pesticides incorporés dans un sol agricole: études expérimentales et modélisation (Doctoral dissertation).

Johannet, A. (2011). Modélisation par apprentissage statistique des systemes naturels, ou en interaction avec un environnement naturel. Applications aux karsts, crues éclair et en robotique (*Doctoral dissertation*, *Université Pierre et Marie Curie-Paris VI*), 83 p.

Johannet, A., Siou, L. K. A., Mangin, A., Estupina, V. B., Pistre, S., & Bertin, D. (2012a). Prediction of spring discharge by neural networks using orthogonal wavelet decomposition. In Neural Networks (IJCNN), The 2012 International Joint Conference on (pp. 1-8). IEEE. DOI: 10.1109/IJCNN.2012.6252620

Johannet, A., Vayssade, B., & Bertin, D. (2008). Neural networks: from black box towards transparent box. Application to evapotranspiration modeling. International Journal of Computational Intelligence, 4(3), 163-170. DOI: 10.1999/1307-6892/15004

Joy, R. J. (2004). Cholera, Chloroform, and the Science of Medicine: A Life of John Snow. Bulletin of the History of Medicine, 78(4), 896-897. DOI: 10.1353/bhm.2004.0174

Juignet, P. (1980). Transgressions-régressions, variations eustatiques et influences tectoniques de l'Aptien au Maastrichtien dans le Bassin de Paris occidental et sur la Bordure du Massif Armoricain. Cretaceous Research, 1(4), 341-357. DOI: 10.1016/0195-6671(80)90043-9

Kallache, M., Rust, H. W., & Kropp, J. (2005). Trend assessment: applications for hydrology and climate research. Nonlinear Processes in Geophysics, 12(2), 201-210.

Kang, S., & Lin, H. (2007). Wavelet analysis of hydrological and water quality signals in an agricultural watershed. Journal of Hydrology, 338(1-2), 1-14. DOI: 10.1016/j.jhydrol.2007.01.047

Kearney, P. C., Sheets, T. J., & Smith, J. W. (1964). Volatility of seven s-triazines. Weeds, 83-87. DOI: 10.2307/4040600

Khammari, B. A. (2013). Etude BAC du captage d'Yport en vue des prescriptions du « Grenelle ». Avis de l'hydrogéologue agrée (Août 2010-Mars 2013). 22.

Khan, S. U., & Hamilton, H. A. (1980). Extractable and bound (nonextractable) residues of prometryn and its metabolites in an organic soil. Journal of Agricultural and Food Chemistry, 28(1), 126-132. DOI: 10.1021/jf60227a013

Kisi, O., & Cimen, M. (2011). A wavelet-support vector machine conjunction model for monthly streamflow forecasting. Journal of Hydrology, 399(1-2), 132-140. DOI: 10.1016/j.jhydrol.2010.12.041

Kisi, O., Dailr, A. H., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. Journal of Hydrology, 450, 48-58. DOI: 10.1016/j.jhydrol.2012.05.031

Kitanidis, P. K., & Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 2. Applications and results. Water Resources Research, 16(6), 1034-1044. DOI: 10.1029/WR016i006p01034

Klöppel, H., Haider, J., & Kördel, W. (1994). Herbicides in surface runoff: a rainfall simulation study on small plots in the field. Chemosphere, 28(4), 649-662. DOI: 10.1016/0045-6535(94)90216-X

Kneller, B., & Buckee, C. (2000). The structure and fluid mechanics of turbidity currents: a review of some recent studies and their geological implications. Sedimentology, 47(s1), 62-94. DOI: 10.1046/j.1365-3091.2000.047s1062.x

Kong-A-Siou, L., Fleury, P., Johannet, A., Estupina, V. B., Pistre, S., & Dörfliger, N. (2014). Performance and complementarity of two systemic models (reservoir and neural networks) used to simulate spring discharge and piezometry for a karst aquifer. Journal of Hydrology, 519, 3178-3192. DOI: 10.1016/j.jhydrol.2014.10.041

Kong-A-Siou, L., Johannet, A., Borrell V., Pistre, S., (2012). Optimization of the generalization capability for rainfall–runoff modeling by neural networks: the case of the Lez aquifer (southern France). Environ Earth Sci 65, 2365–2375. DOI: 10.1007/s12665-011-1450-9

Kong-A-Siou, L., Johannet, A., Borrell, V., Pistre, S., (2011). Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). Journal of Hydrology 403, 367–380.DOI: 10.1016/j.jhydrol.2011.04.015

Koriko, M., Tchegueni, S., Koledzi, K. E., Tchangbedji, G., Hafidi, M., & Sarakha, M. (2014). Cinétique de la photodégradation induite du profenofos (PFF) en solution aqueuse par excitation des substances humiques naturelles des sols/Kinetics of the Photodegradation of profenofos (PFF) induced by humic substances in aqueous solution. Journal de la Société Ouest-Africaine de Chimie, 38, 73.

Kothyari, U. C., & Jain, S. K. (1997). Sediment yield estimation using GIS. Hydrological sciences journal, 42(6), 833-843. DOI: 10.1080/02626669709492082

Kothyari, U. C., Tiwari, A. K., & Singh, R. (1997). Estimation of temporal variation of sediment yield from small catchments through the kinematic method. Journal of Hydrology, 203(1-4), 39-57. DOI: 10.1016/S0022-1694(97)00084-X

Kot-Wasik, A., Zabiegała, B., Urbanowicz, M., Dominiak, E., Wasik, A., & Namieśnik, J. (2007). Advances in passive sampling in environmental studies. Analytica chimica acta, 602(2), 141-163. DOI: 10.1016/j.aca.2007.09.013

Labat, D. (2005). Recent advances in wavelet analyses: Part 1. A review of concepts. Journal of Hydrology, 314(1-4), 275-288. DOI: 10.1016/j.jhydrol.2005.04.003

Labat, D. (2006). Oscillations in land surface hydrological cycle. Earth and Planetary Science Letters, 242(1-2), 143-154. DOI: 10.1016/j.epsl.2005.11.057

Labat, D. (2008). Wavelet analysis of the annual discharge records of the world's largest rivers. Advances in Water Resources, 31(1), 109-117. DOI: 10.1016/j.advwatres.2007.07.004

Labat, D. (2010). Cross wavelet analyses of annual continental freshwater discharge and selected climate indices. Journal of Hydrology, 385(1-4), 269-278. DOI: 10.1016/j.jhydrol.2010.02.029

Labat, D., Ababou, R., & Mangin, A. (1999a). Analyse en ondelettes en hydrologie karstique. 1re partie: analyse univariée de pluies et débits de sources karstiques. Comptes Rendus de l'Academie des Sciences Series IIA Earth and Planetary Science, 12(329), 873-879.

Labat, D., Ababou, R., & Mangin, A. (1999b). Analyse en ondelettes en hydrologie karstique. 2e partie: analyse en ondelettes croisée pluie-débit. Comptes Rendus de l'Académie des Sciences-Series IIA-Earth and Planetary Science, 329(12), 881-887. DOI: 10.1016/S1251-8050(00)88501-8

Labat, D., Ababou, R., & Mangin, A. (2000). Rainfall–runoff relations for karstic springs. Part I: convolution and spectral analyses. Journal of Hydrology, 238(3-4), 123-148. DOI: 10.1016/S0022-1694(00)00321-8

Labat, D., Ababou, R., & Mangin, A. (2000). Rainfall–runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. Journal of hydrology, 238(3-4), 149-178. DOI: 10.1016/S0022-1694(00)00322-X

Labat, D., Ababou, R., & Mangin, A. (2000a). Rainfall—runoff relations for karstic springs. Part I: convolution and spectral analyses. Journal of Hydrology, 238(3-4), 123-148. DOI: 10.1016/S0022-1694(00)00321-8

Labat, D., Ababou, R., & Mangin, A. (2001). Introduction of Wavelet Analyses to Rainfall/Runoffs Relationship for a Karstic Basin: The Case of Licq-Atherey Karstic System (France). Groundwater, 39(4), 605-615. DOI : 10.1111/j.1745-6584.2001.tb02348.x Labat, D., Ababou, R., & Mangin, A. (2002). Analyse multirésolution croisée de pluies et débits de sources karstiques. Comptes Rendus Géoscience, 334(8), 551-556. DOI: 10.1016/S1631-0713(02)01795-9

Labat, D., Ronchail, J., Callede, J., Guyot, J. L., De Oliveira, E., & Guimaraes, W. (2004). Wavelet analysis of Amazon hydrological regime variability. Geophysical Research Letters, 31(2). DOI: 10.1029/2003GL018741

Lacroix, M., Rodet, J., Wang, H. Q., Masséi, N., & Dupont, J. P. (2000). Origine des matières en suspension dans un système aquifère karstique: apports de la microgranulométrie. Comptes Rendus de l'Academie des Sciences-Series IIA-Earth and Planetary Science, 330(5), 347-354. DOI: 10.1016/S1251-8050(00)00161-0

Laignel B., (1997). Les altérites à silex de l'Ouest du Bassin de Paris : caractérisation lithologique, genèse et utilisation potentielle comme granulats. Thèse de Doctorat. Université de Rouen. 219p.

Laignel B., (2003). Caractérisation et dynamique érosive de systèmes géomorphologiques continentaux et crayeux. Exemple de l'Ouest du Bassin de Paris dans le contexte nord-ouesteuropéen. Mémoire de HDR. Université de Rouen. 143p.

Lawler, D. M. (2005). Spectrophotometry: turbidimetry and nephelometry. In Encyclopedia of Analytical Science, Elsevier, 2nd ed, 10 volumes, 343-351.(pp. 343-351). DOI: 10.1016/B0-12-369397-7/00718-4

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. Quarterly of applied mathematics, 2(2), 164-168.

Lode, O., Eklo, O. M., Holen, B., & Svensen, A. (1995). Pesticides in precipitation in Norway. Science of the total environment, 160, 421-431. DOI: 10.1016/0048-9697(95)04375-B

Mahler, B. J., Personné, J. C., Lods, G. F., & Drogue, C. (2000). Transport of free and particulate-associated bacteria in karst. Journal of Hydrology, 238(3-4), 179-193. DOI: 10.1016/S0022-1694(00)00324-3

Makwana, J. J., & Tiwari, M. K. (2014). Intermittent streamflow forecasting and extreme event modelling using wavelet based artificial neural networks. Water resources management, 28(13), 4857-4873. DOI: 10.1007/s11269-014-0781-1

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence, 11(7), 674-693. DOI: 10.1109/34.192463

Mangin, A. (1975). Contribution à l'étude hydrodynamique des aquifères karstiques (Doctoral dissertation, Université de Dijon).

Mangin, A. (1981). Apport des analyses correlátoires et spectrales croisées dans la connaissance des systemes hydrologiques. CR Acad. Sci. Paris, 293(2), 1011-1014.

Mangin, A. (1998). L'approche hydrogéologique des karsts. Speleochronos, (9), 3-26.

Maqsoud, A. (1996). Approche hydrologique et hydrochimique du caractère karstique éventuel d'hydrosystèmes souterrains de la craie du bassin de Paris (Doctoral dissertation, Lille).

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. Journal of the society for Industrial and Applied Mathematics, 11(2), 431-441. DOI: 10.1137/0111030

Massei, N. (2001). Transport de particules en suspension dans l'aquifère crayeux karstique et à l'interface craie/alluvions (Doctoral dissertation, Rouen).

Massei, N., Dieppois, B., Hannah, D. M., Lavers, D. A., Fossa, M., Laignel, B., & Debret, M. (2017). Multitime-scale hydroclimate dynamics of a regional watershed and links to large-scale atmospheric circulation: Application to the Seine river catchment, France. Journal of Hydrology, 546, 262-275. DOI: 10.1016/j.jhydrol.2017.01.008

Massei, N., Dupont, J. P., Mahler, B. J., Laignel, B., Fournier, M., Valdes, D., & Ogier, S. (2006). Investigating transport properties and turbidity dynamics of a karst aquifer using correlation, spectral, and wavelet analyses. Journal of hydrology, 329(1-2), 244-257. DOI: 10.1016/j.jhydrol.2006.02.021

Massei, N., Wang, H. Q., Dupont, J. P., Rodet, J., & Laignel, B. (2003). Assessment of direct transfer and resuspension of particles during turbid floods at a karstic spring. Journal of hydrology, 275(1-2), 109-121. DOI: 10.1016/S0022-1694(03)00020-9

Mégnien, C. (1979). Hydrogéologie du centre du bassin de Paris: contribution à létude de quelques aquifères principaux (Doctoral dissertation, Paris).

Miettinen, I., Martikainen, P., Vartiainen, T., & Lötjönen, S. (1993). Biochemical and chemical degradation of 3-chloro-4-(dichloromethyl)-5-hydroxy-2 (5H)-furanone (MX) in surface and drinking water. Chemosphere, 27(9), 1707-1718. DOI: 10.1016/0045-6535(93)90151-T

Miller, J. L., Wollum, A. G., & Weber, J. B. (1997). Degradation of carbon-14-atrazine and carbon-14metolachlor in soil from four depths. Journal of environmental quality, 26(3), 633-638. DOI: 10.2134/jeq1997.00472425002600030007x

Millet, M., & Bedos, C. (2017). La contamination de l'atmosphère par les produits phytosanitaires: protéger les végétaux des attaques de pesticides nuisibles. 2268-3798.

Miquel, M. G. (2003). Office parlementaire d'évaluation des choix scientifiques et technologiques. La qualité de l'eau et de l'assainissement en France. 2, 23-24.

Muet, P. & Vier, E. (2006). Bilan et analyse de la mise en œuvre des périmètres de protection des captages AEP en milieu karstique. Synthèse des préconisations en faveur de l'amélioration des démarches de protection. Agence de l'Eau RMC. 28p.

Mulia, I. E., Tay, H., Roopsekhar, K., & Tkalich, P. (2013). Hybrid ANN–GA model for predicting turbidity and chlorophyll-a concentrations. Journal of Hydro-environment Research, 7(4), 279-299. DOI: 10.1016/j.jher.2013.04.003

Nalley, D., Adamowski, J., & Khalil, B. (2012). Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). Journal of Hydrology, 475, 204-228. DOI: 10.1016/j.jhydrol.2012.09.049

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10(3), 282-290. DOI: 10.1016/0022-1694(70)90255-6

Nebbache, S. L. M. A., Loquet, M., Vinceslas-Akpa, M., & Feeny, V. (1997). Turbidité et micro-organismes dans une eau de source en milieu karstique. European journal of soil biology, 33(2), 89-103.

Nerrand, O., Roussel-Ragot, P., Personnaz, L., Dreyfus, G., & Marcos, S. (1993). Neural networks and nonlinear adaptive filtering: unifying concepts and new algorithms. Neural computation, 5(2), 165-199. DOI: 10.1162/neco.1993.5.2.165

Nieto, P. G., García-Gonzalo, E., Fernández, J. A., & Muñiz, C. D. (2014). Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: A case study in Northern Spain. Ecological Engineering, 73, 192-200. DOI: 10.1016/j.ecoleng.2014.09.042

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005b). Which potential evapotranspiration input for a lumped rainfall–runoff model? : Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. Journal of hydrology, 303(1-4), 290-306. DOI: 10.1016/j.jhydrol.2004.08.026

Oudin, L., Michel, C., & Anctil, F. (2005a). Which potential evapotranspiration input for a lumped rainfall-runoff model? : Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? Journal of Hydrology, 303(1-4), 275-289. DOI: 10.1016/j.jhydrol.2004.08.025

Paschke, A., Brümmer, J., & Schüürmann, G. (2007). Silicone rod extraction of pharmaceuticals from water. Analytical and bioanalytical chemistry, 387(4), 1417-1421. DOI: 10.1007/s00216-006-0986-0

Pereira-Machado, V. 2017. Analyses des produits phytosanitaires de l'eau brute du captage d'alimentation en eau potable d'Yport. (Rapport de stage, Alès).

Petit, V., Berard, P., Lachassagne, P., Lemordant, Y., Talbo, H. (1998). Les périmètres de protection en milieu fissuré : exemples et principes d'application. Rapport BRGM/RR-40289-FR. 55 p.

Pronk, M., Goldscheider, N., & Zopfi, J. (2007). Particle-size distribution as indicator for fecal bacteria contamination of drinking water from karst springs. Environmental Science & Technology, 41(24), 8400-8405. DOI: 10.1021/es071976f

Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M., & Nourani, V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Science of the total environment, 407(17), 4916-4927. DOI: 10.1016/j.scitotenv.2009.05.016

Rao, Y. S., Krisha, B., & Venkatesh, B. (2014). Wavelet based neural networks for daily stream flow forecasting. Int. J. Emerg. Technol. Adv. Eng, 4(1), 307-317.

Rathinasamy, M., & Khosa, R. (2012). Multiscale nonlinear model for monthly streamflow forecasting: a wavelet-based approach. Journal of Hydroinformatics, 14(2), 424-442. DOI: 10.2166/hydro.2011.130

Raveton, M., Ravanel, P., Serre, A. M., Nurit, F., & Tissut, M. (1997). Kinetics of uptake and metabolism of atrazine in model plant systems. Pest Management Science, 49(2), 157-163. DOI: 10.1002/(SICI)1096-9063(199702)49:2<157::AID-PS517>3.0.CO;2-M

Ristori, G.G., & Fusi, P. (1995). Adsorption mechanisms and abiotic catalytic transformations of some agrochemicals by clay minerals. Environmental Impacts of Soil Component Interactions: Land Quality, Natural and Anthropogenic Organics, 1, 337.

Roche, P. A. (2002). L'eau. Revue d'économie financière, 21-37. DOI: 10.3406/ecofi.2002.3740

Rodet J. (1992). La craie et ses karsts. *Centre normand d'étude du karst et des cavités du sous-sol et Groupe Seine*, CNRS (Caen), 560 p.

Rodet, J. (1993). Le rôle des formations quaternaires dans le drainage karstique: l'exemple des Craies du bassin de Paris. Quaternaire, 4(2), 97-102. DOI: 10.3406/quate.1993.1998

Rousseaux, S. (2001). Diversité taxonomique, génotypique et physiologique des souches bactériennes dégradant l'atrazine dans les sols: impacts d'apports de boues de station d'épuration et essai de bioremédiation (Thèse de doctorat, Dijon).

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533-538. DOI: 10.1038/323533a0

Ryan, M., & Meiman, J. (1996). An Examination of Short-term Variations in Water Quality at a Karst Spring in Kentucky. Groundwater, 34(1), 23-30. DOI: 10.1111/j.1745-6584.1996.tbo1861.x

SAFEGE (2012) Étude du Bassin d'Alimentation du forage d'Yport "Phase 1 :Délimitation du bassin d'alimentation et analyse de la vulnérabilité intrinsèque du basin". 112.

Sang, Y. F. (2013). A review on the applications of wavelet transform in hydrology time series analysis. Atmospheric research, 122, 8-15. DOI: 10.1016/j.atmosres.2012.11.003

Savary, M., Johannet, A., Massei, N., Dupont, J. P., & Hauchard, E. (2016, June). Operational Turbidity Forecast Using Both Recurrent and Feed-Forward Based Multilayer Perceptrons. In International Work-Conference on Time Series Analysis (pp. 243-256). Springer, Cham. DOI: 10.1007/978-3-319-55789-2_17

Schiavon, M., Barriuso, E., Portal, J. M., Andreux, F., Bastide, J., & Coste, C. (1990). Etude du devenir de deux substances organiques utilisées dans les sols, l'une massivement (l'atrazine) et l'autre à l'état trace (le metsulfuron-méthyl), à l'aide de molécules marquées au 14C. Rapport SRETIE/MERE 7219, 75 p.

Schroll, R., Langenbach, T., Cao, G., Dörfler, U., Schneider, P., & Scheunert, I. (1992). Fate of [14C] terbutylazine in soil-plant systems. Science of the total environment, 123, 377-389. DOI: 10.1016/0048-9697(92)90161-K

Shevenell, L., & McCarthy, J. F. (2002). Effects of precipitation events on colloids in a karst aquifer. Journal of Hydrology, 255(1-4), 50-68. DOI: 10.1016/S0022-1694(01)00510-8

Shiri, J., & Kisi, O. (2010). Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. Journal of Hydrology, 394(3-4), 486-493. DOI: 10.1016/j.jhydrol.2010.10.008

Slimani, S. (2009). Les Fonctionnements hydrologiques de l'aquifère karstifié de la Craie de Haute-Normandie: conséquences des contextes morphologiques et structuraux (Doctoral dissertation, Rouen).

Smith, A. H., Bates, M. N., Steinmaus, C. M., & von Ehrenstein, O. (2003). Arsenic epidemiology and drinking water standards. In Abstracts of the 5th Congress of Toxicology in Developing Countries.
Soedergren, A. (1987). Solvent-filled dialysis membranes simulate uptake of pollutants by aquatic organisms. Environmental science & technology, 21(9), 855-859. DOI: 10.1021/es00163a001

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society. Series B (Methodological), 111-147.

Tasli, S. (1995). Devenir de l'atrazine en culture de maïs: études en plein champ- site de La Côte Saint André, Isère, France- et au laboratoire. Recherche de solutions alternatives de traitement (Thèse de doctorat).

Taver, V. (2014). Caractérisation et modélisation hydrodynamique des karsts par réseaux de neurones. Application à l'hydrosystème du Lez (Doctoral dissertation, Université Montpellier 2).

Tomlin, C. (1994). The pesticide manual. Incorporating the agrichemicals handbook: A world compendium. 10th Edition. British Crop Protection Council publication, Farnham, U.K. 1341 p.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. Bulletin of the American Meteorological society, 79(1), 61-78. DOI: 10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Toukourou, M., Johannet, A., Dreyfus, G., & Ayral, P. A. (2011). Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of "Cévenol flash floods". Applied Intelligence, 35(2), 178-189. DOI: 10.1007/s10489-010-0210-y

Towler, E., Rajagopalan, B., Summers, R. S., & Yates, D. (2010). An approach for probabilistic forecasting of seasonal turbidity threshold exceedance. Water Resources Research, 46(6). DOI: 10.1029/2009WR007834

Tramblay, Y., Saint-Hilaire, A., & Ouarda, T.B. (2007). Modelling extreme suspended sediment concentrations in North America: frequency analysis and correlations with watershed characteristics. Water Quality and Sediment Behaviour of the Future: Predictions for the 21st Century, (314), 20.

Valdès, D. (2005). Variabilite spatiale du fonctionnement hydrologique de l'aquifère karstique de la craie (Département de l'Eure, France) (Doctoral dissertation, Rouen).

Voltz, M., & Louchart, X. (2001). Les facteurs-clés de transfert des produits phytosanitaires vers les eaux de surface. Ingénieries Numéro Spécial «Phytosanitaires: transfert, diagnostic et solutions correctives, 45-54.

Vrana, B., Mills, G. A., Dominiak, E., & Greenwood, R. (2006). Calibration of the Chemcatcher passive sampler for the monitoring of priority organic pollutants in water. Environmental Pollution, 142(2), 333-343. DOI: 10.1016/j.envpol.2005.10.033

Wang, W., & Ding, J. (2003). Wavelet network model and its application to the prediction of hydrology. Nature and Science, 1(1), 67-71.

Wicks, J. M., & Bathurst, J. C. (1996). SHESED: a physically based, distributed erosion and sediment yield component for the SHE hydrological modelling system. Journal of Hydrology, 175(1-4), 213-238. DOI: 10.1016/S0022-1694(96)80012-6

Winkelmann, D. A., & Klaine, S. J. (1991). Degradation and bound residue formation of four atrazine metabolites, deethylatrazine, deisopropylatrazine, dealkylatrazine and hydroxyatrazine, in a western Tennessee soil. Environmental toxicology and chemistry, 10(3), 347-354. DOI: 10.1002/etc.5620100307

Zhang, H., Davison, W., & Grime, G. W. (1995). New in-situ procedures for measuring trace metals in pore waters. In Dredging, Remediation, and Containment of Contaminated Sediments. ASTM International. DOI: 10.1520/STP15999S

Limites de qualité des eaux destinées à la consommation humaine, à l'exclusion des eaux conditionnées (extrait de l'arrêté du 11 janvier 2007 relatif aux limites et références de qualité des eaux brutes et des eaux destinées à la consommation humaine mentionnées aux articles R. 1321-2, R. 1321-3, R. 1321-7 et R. 1321-38 du code de la santé publique).

Limites de qualité (LQ) : Paramètres microbiologiques								
Paramètres	Valeurs	Unités	Commentaires					
Escherichia coli (E.Coli)	0	/100 ml						
Entérocoques	0	/100 ml						

Li	mites de qu	alité (LQ) :	Paramètres chimiques
Paramètres	Valeurs	Unités	Commentaires
Acrylamide	0,1	μg/l	
Antimoine	5	μg/l	
Arsenic	10	μg/l	
Baryum	0,7	mg/l	
Benzène	1	μg/l	
Benzo(a)pyrène	0,01	μg/l	
Bore	1	mg/l	
Bromates	10	μg/l	La valeur la plus faible possible inférieure à cette limite doit être visée sans pour autant compromettre la désinfection. La limite de qualité est fixée à 25 µg/L jusqu'au 25 décembre 2008. Toutes les mesures appropriées doivent être prises pour réduire le plus possible la concentration de bromates dans les eaux destinées à la consommation humaine, au cours de la période nécessaire pour se conformer à la limite de qualité de 10 µg/L.
Cadmium	5	μg	
Chrome	50	μg/l	
Chlorure de vinyle	0,5	μg/l	
Cuivre	2	mg/l	
Cyanures totaux	50	μg/l	
1,2- dichloroéthane	3	μg/l	
Epichlorhydrine	0,1	μg/l	
Fluorures	1,5	mg/l	
Hydrocarbures aromatiques polycycliques (HAP)	0,1	μg/l	Pour la somme des composés suivants : fluoranthène, Aromatiques Benzo[b]fluoranthène, Benzo[k]fluoranthène, Polycycliques Benzo[ghi]pérylène, Indéno[1,2,3-cd]pyrène.
Mercure total	1	μg/l	
Microcystine-LR	1	μg/l	A rechercher en cas de prolifération algale dans les eaux brutes.
Nickel	20	μg/l	

Nitrates (NO ₃ -)	50 mg/l		La somme de la concentration en nitrates divisée
			par 50 et de celle en nitrités divisée par 3 doit rester
	0.5		En cortio dos installations do traitoment. la
	0,5	iiig/i	Ell sol de des installations de traitement, la
Destisides (as a substance			egale a 0,10 mg/L.
Pesticides (par substance	0,1	µg/i	Par « pesticides », on entend :
individuelle).			– les insecticides organiques ;
			– les herbicides organiques ;
			– les fongicides organiques ;
			– les nematocides organiques ;
			– les acaricides organiques ;
			– les algicides organiques ;
Aldrine, dieldrine, heptachlore,	0,03	μg/l	– les rodenticides organiques ;
heptachlorépoxyde (par			 – les produits antimoisissures organiques ;
substance individuelle).			 les produits apparentés (notamment les
			régulateurs de croissance)
			et leurs métabolites, produits de dégradation et de
			réaction pertinents.
Total pesticides	0,5	μg/l	Par "total pesticides", on entend la somme de tous
	_		les pesticides individualisés et quantifiés.
Blamb			
FIOLID	10	μ	décembre ages l'es mesures appropriées pour
			réduire prograssivement la concentration en plemb
			dans las aquiv destinées à la concentration en pionib
			bumaine au cours de la période péressaire pour co
			conformer a la limite de qualité de 10 μ g/L sont
			precisees aux articles R. 1321-55 et R. 1321-49
			(arrete d'application). Lors de la mise en œuvre des
			mesures destinees a atteindre cette valeur, la
			priorité est donnée aux cas où les concentrations en
			plomb dans les eaux destinées à la consommation
			humaine sont les plus élevées.
Selenium	10	μg/I	
Tétrachloroéthylène et	10	μg/l	Somme des concentrations des paramètres
trichloro- éthylène.			spécifiés.
Total trihalométhanes (THM)	100	μg/l	La valeur la plus faible possible inférieure à cette
			valeur doit être visée sans pour autant
			compromettre la désinfection. Par « total
			trihalométhanes », on entend la somme de :
			chloroforme, bromoforme, dibromochlorométhane
			et bromodichlorométhane. La limite de qualité est
			fixée à 150 g/L jusqu'au 25 décembre 2008. Toutes
			les mesures appropriées doivent être prises pour
			réduire le plus possible la concentration de THM
			dans les eaux destinées à la consommation
			humaine, au cours de la période nécessaire pour se
			conformer à la limite de qualité

Turbidité	1	NFU	La limite de qualité est applicable au point de mise
			en distribution, pour les eaux visées à l'article R.
			1321-37 et pour les eaux d'origine souterraine
			provenant de milieux fissurés présentant une
			turbidité périodique importante et supérieure à 2,0
			NFU. En cas de mise en œuvre d'un traitement de
			neutralisation ou de re-minéralisation, la limite de
			qualité s'applique hors augmentation éventuelle de
			turbidité due au traitement. Pour les installations
			qui sont d'un débit inférieur à 1 000 m3/j ou qui
			desservent des unités de distribution de moins de 5
			ooo habitants, la limite de qualité est fixée à 2,0
			NFU jusqu'au 25 décembre 2008. Toutes les
			mesures appropriées doivent être prises pour
			réduire le plus possible la turbidité, au cours de la
			période nécessaire pour se conformer à la limite de
			qualité de 1,0 NFU.

Référence de qualité des eaux destinées à la consommation humaine, à l'exclusion des eaux conditionnées (extrait de l'arrête du 11 janvier 2007 relatif aux limites et références de qualité des eaux brutes et des eaux destinées a la consommation humaine mentionnées aux articles r. 1321-2, r. 1321-3, r. 1321-7 et r. 1321-38 du code de la sante publique).

Référence de qualité (RQ) : paramètres microbiologiques									
Paramètres	Valeurs	Unités	Commentaires						
Bactéries coliformes	0	/100mL							
Bactéries sulfitoréductrices y compris les spores	0	/100mL	Ce paramètre doit être mesuré lorsque l'eau est d'origine superficielle ou influencée par une eau d'origine superficielle. En cas de non- respect de cette valeur, une enquête doit être menée sur la distribution d'eau pour s'assurer qu'il n'y a aucun danger potentiel pour la santé humaine résultant de la présence de micro-organismes pathogènes, par exemple Cryptosporidium.						

Référence de qualité (RQ) : paramètres chimiques et organoleptiques									
Paramètres	Valeurs	Unités	Commentaires						
Aluminium total	200	μg/L							
Ammonium	0,1	mg/L	S'il est démontré que l'ammonium à une origine naturelle, la valeur est de 0,5 mg/l pour les eaux souterraines						
Chlore libre et total	Absence d'odeur et de saveur désagréable et pas de changement anormal								
Cuivre	1	mg/L							
Chlorites	0,2	mg/L	Sans compromettre la désinfection, la valeur la plus faible possible doit être visée.						
Chlorures	250	mg/L	Les eaux ne doivent pas être agressives.						
Couleur	15	mg/L Pt- Co	Acceptable pour les consommateurs et aucun changement anormal.						
Conductivité	180 à 1000	μS/cm	à 20°C Les eaux ne doivent pas être corrosives.						
Concentration en ions hydrogène	6,5 à 9	Unités pH	Les eaux ne doivent pas être agressives.						
Carbone Organique Total (COT)	2	mg/L	Oxydabilité au permanganate de potassium mesurée après 10 minutes en milieu acide.						
Fer total	200	μg/L							
Manganèse	50	μg/L							

Numération de germes aérobies revivifiables à 22°C et 37°C	Variation dans un rapport de 10 par		
	rapport à la		
	valeur		
	nadituelle		

Référence de qualité (RQ) : paramètres indicateurs de radioactivité									
Paramètres	Valeurs	Unités	Commentaires						
Activité alpha globale			En cas de valeur supérieure à 0,10 Bq/L, il est procédé à l'analyse des radionucléides spécifiques définis dans l'arrêté mentionné à l'article R. 1321-20.						
Activité bêta globale résiduelle			En cas de valeur supérieure à 1,0 Bq/L, il est procédé à l'analyse des radionucléides spécifiques définis dans l'arrêté mentionné à l'article R. 1321-20.						
Dose totale indicative (DTI)	0,1	mSv/an	Le calcul de la DTI est effectué selon les modalités définies à l'article R. 1321-20.						
Tritium	100	Bq/L	La présence de concentrations élevées de tritium dans l'eau peut être le témoin de la présence d'autres radionucléides artificiels. En cas de dépassement de la référence de qualité, il est procédé à l'analyse des radionucléides spécifiques définis dans l'arrêté mentionné à l'article R. 1321-20.						

Limites de qualité des eaux brutes de toute origine utilisées pour la production d'eau destinée à la consommation humaine (extrait de l'arrête du 11 janvier 2007 relatif aux limites et références de qualité des eaux brutes et des eaux destinées à la consommation humaine mentionnées aux articles r. 1321-2, r. 1321-3, r. 1321-7 et r. 1321-38 du code de la sante publique)

Groupes de paramètres	Paramètres	Limites de qualité	Unités				
Paramètres organoleptiques	Couleur (Pt) (1)	200	mg/L				
Paramètres physico-	Chlorures (Cl ⁻) ⁽¹⁾	200	mg/L				
chimiques liés à la structure	Sodium (Na+) (1)	200	mg/L				
naturelle des eaux	Sulfates (SO42 ⁻) ⁽¹⁾	250	mg/L				
	Taux de saturation en oxygène dissous pour les eaux superficielles (O2) ⁽¹⁾	< 30	%				
	Température ⁽¹⁾ (2)	25	°C				
Paramètres concernant les substances indésirables	Agents de surface réagissant au bleu de méthylène (lauryl-sulfate de sodium)	0,5	mg/L				
	Ammonium (NH ₄ +)	4	mg/L				
	Baryum (Ba) pour les eaux superficielles	1	mg/L				
	Carbone organique total (COT) ⁽¹⁾ (3)	10	mg/L				
	Hydrocarbures dissous ou émulsionnés	1	mg/L				
	Nitrates pour les eaux superficielles (NO3 ⁻)	50	mg/L				
	Nitrates pour les autres eaux (NO3 ⁻)	100	mg/L				
	Phénols (indice phénol) (C6H5OH)	0,1	mg/L				
	Zinc (Zn)						
Paramètres concernant les	Arsenic (As)	100	μg/L				
substances toxiques	Cadmium (Cd)	5	μg/L				
	Chrome total (Cr)	50	μg/L				
	Cyanures (CN⁻)	50	μg/L				
	Hydrocarbures aromatiques polycycliques (HAP) : Somme des composés suivants : fluoranthène, benzo[b]fluoranthène, benzo[k]fluoranthène, benzo[a]pyrène, benzo[g,h,i]pérylène et indéno[1,2,3-cd]pyrène	1	μg/L				
	Mercure (Hg)	1	μg/L				
	Plomb (Pb)	50	μg/L				
	Sélénium (Se)	10	μg/L				
Pesticides	Par substances individuelles, y compris les métabolites	2	μg/L				
	Total	5	μg/L				
Paramètres microbiologiques	Entérocoques	10 000	/100mL				
	Escherichia coli.	20 000	/100mL				
(1) L'avis de l'Agence français requis pour les paramètres no	e de sécurité sanitaire des aliments mentionné à l'article tés (1). Toutefois. l'avis de l'Agence francaise de sécurité	R. 1321-7 (II) n' sanitaire des a	est pas liments				

s pour les paramètres notés (1). Toutefois, l'avis de l'Agence française de sécurité sanitaire des alimen est sollicité lorsque la ressource en eau utilisée est de l'eau de mer.

(2) La limite de qualité pour le paramètre température ne s'applique pas dans les départements d'outre-mer.
 (3) Le plan de gestion des ressources en eau prévu à l'article R. 1321-42 n'est pas requis pour les paramètres notés (3).

Limites de qualité des eaux douces superficielles utilisées pour la production d'eau destinée à la consommation humaine (extrait de l'arrête du 11 janvier 2007 relatif aux limites et références de qualité des eaux brutes et des eaux destinées à la consommation humaine mentionnées aux articles r. 1321-2, r. 1321-3, r. 1321-7 et r. 1321-38 du code de la sante publique).

Groupes de	Paramètres Groupe						Unités	
paramètres		A	1	A	2	A3		
		G	Ι	G	I	G	Ι	
Paramètres	Couleur (Pt)	10	20	50	100	50	200	mg/L
organoleptiques	Odeur (facteur de dilution à 25 °C)	3		10		20		mg/L
Paramètres physicochimiques liés à la structure	Chlorures (Cl⁻)	200		200		200		mg/L
naturelle des eaux	Conductivité	1 000		1 000		1000		μS/cm à 20 °C
		1 100		1 100		1 1 0 0		S/cmà
	Demande biochimique en oxygène (DBO5) à 20 °C sans nitrification (O2)	< 3		< 5		< 7		mg/L
	Demande chimique en oxygène (DCO) (O ₂)					30		mg/L
	Matières en suspension	25						mg/L
	рН	6,5-8,5		5,5-9		5,5-9		unités pH
	Sulfates (SO ₄ ^{2–})	150	250	150	250	150	250	mg/L
	Taux de saturation en oxygène dissous (O ₂)	> 70		> 50		> 30		%
	Température	22	25	22	25	22	25	°C
Paramètres concernant les substances indésirables	Agents de surface réagissant au bleu de méthylène (lauryl- sulfate de sodium)	0,2		0,2		0,5		mg/L
	Ammonium (NH ₄ +)	0,05		1	1,5	2	4	mg/L
	Azote Kjeldhal (N)	1		2		3		mg/L
	Baryum (Ba)		0,1		1		1	mg/L
	Bore (B)	1		1		1		mg/L
	Cuivre (Cu)	0,02	0,05	0,05		1		mg/L
	Fer dissous sur échantillon filtré à 0,45 µm	0,1	0,3	1	2	1		mg/L
	Fluorures (F [−])	0,7/1	1,5	0,7/1,7		0,7/1,7		mg/L
	Hydrocarbures dissous ou émulsionnés		0,05		0,2	0,5	1	mg/L
	Manganèse (Mn)	0,05		0,1		1		mg/L
	Nitrates (NO ₃ ⁻)	25	50		50		50	mg/L

	Phénols (indice phénol) (C₀H₅OH)		0,001	0,001	0,005	0,01	0,1	mg/L
	Phosphore total (P ₂ O ₅)	0,4		0,7		0,7		mg/L
	Substances extractibles au chloroforme	0,1		0,2		0,5		mg/L
	Zinc (Zn)	0,5	3	1	5	1	5	mg/L
Paramètres	Arsenic (As)		10		50	50	100	μg/L
concernant les	Cadmium (Cd)	1	5	1	5	1	5	μg/L
toxiques	Chrome total (Cr)		50		50		50	μg/L
	Cyanures (CN ⁻)		50		50		50	μg/L
	Hydrocarbures aromatiques polycycliques (HAP) : Somme des composés suivants : fluoranthène, benzo[b]fluoranthène, benzo[k]fluoranthène, benzo[a]pyrène, benzo[g,h,i]pérylène et indéno[1,2,3- cd]pyrène		0,2		0,2		1	μg/L
	Mercure (Hg)	0,5	1	0,5	1	0,5	1	μg/L
	Plomb (Pb)		10		50		50	μg/L
	Sélénium (Se)		10		10		10	μg/L
Pesticides	Par substances individuelles, y compris les métabolites		0,1 ^(1,2)		0,1 ^(1,2)		2	μg/L
	Total		0,5 (2)		0,5 (2)		5	μg/L
Paramètres	Bactéries coliformes	50		5 000		50 000		/100mL
microbiologiques	Entérocoques	20		1000		10 000		/100mL
	Escherichia coli	20		2 000		20 000		/100mL
	Salmonelles	Absent dans 5000 mL		Absent dans 1000 mL				
(1) Pour l'aldi	rine, la dieldrine, l'heptachlor	e et l'hepta	chlorepox	yde, la limit	e de quali	té est de o,o	ο3 μg/L.	libratian 11
(2) Ces valeurs ne cono peut être fait a	cernent que les eaux superfici Ippel à des eaux de qualités d	elles utilisé ifférentes, l	es directei e taux de	ment, sans dilution dev	allution pr vant être c	ealable. En alculé au ca	cas de c s par ca	s.

NB :

Article R1321-38

Modifié par Décret n°2007-49 du 11 janvier 2007 - art. 1 JORF 12 janvier 2007

Les eaux douces superficielles sont classées selon leur qualité dans les groupes A1, A2 et A3 en fonction des critères définis par arrêté du ministre chargé de la santé relatif aux limites de qualité des eaux douces superficielles utilisées pour la production d'eau destinée à la consommation humaine. Leur utilisation pour la consommation humaine est subordonnée pour les eaux classées en :

1° Groupe A1 : à un traitement physique simple et à une désinfection ;

2° Groupe A2 : à un traitement normal physique, chimique et à une désinfection ;

3° Groupe A3 : à un traitement physique et chimique poussé, à des opérations d'affinage et de désinfection.

Carte des tarages effectués de 1997 à 2013 sur et aux alentours du bassin d'alimentation du captage d'Yport.



Tableaux récapitulatifs des minimums, maximums et moyennes des données de turbidité, précipitation et hauteur d'eau dans le forage pour chaque mois des années allant de 2010 à 2014.

			Turbidité (NTU)										
		Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
	Minimum	1,5	0,0	4,3	3,0	2,2	1,6	1,0	1,5	0,9	0,9	1,5	0,0
2010	Maximum	45,7	154,1	135,0	22,0	22,7	5,7	5,1	10,5	9,8	8,4	245,4	134,9
2010 2011 2012	Moyenne	7,6	14,9	14,6	6,8	3,8	3,0	3,0	2,8	2,9	3,1	21,1	6,5
	Minimum	2,1	1,5	0,9	0,8	0,9	0,3	0,9	1,0	0,9	0,9	0,8	0,9
2011	Maximum	62,8	4,3	11,2	10,5	11,8	3,7	5,1	22,8	256,1	10,5	7,0	307,9
	Moyenne	6,7	2,3	1,9	2,0	2,0	1,9	1,8	2,3	9,8	2,0	1,6	37,2
	Minimum	2,2	1,5	1,5	0,8	0,8	0,8	0,9	0,8	0,2	0,8	5,6	3,5
2012	Maximum	48,6	9,7	2,2	5,7	5,0	3,6	5,7	5,7	3,0	167 , 0	17,8	262,9
2010 2011 2012 2013 2014	Moyenne	7,7	2,8	1,9	1,9	1,6	1,5	1,3	1,2	1,4	22,3	11,7	70 , 8
	Minimum	0,2	2,8	2,1	0,9	0,9	0,2	0,9	0,9	0,9	0,8	0,5	2,2
2013	Maximum	122,1	76 , 4	49,2	9,1	5,0	4,3	3,7	15,2	5,0	405,3	157,5	86,7
	Moyenne	15,1	12,5	8,3	3,1	2,5	2,1	1,9	1,8	1,7	20,3	23,7	12,2
	Minimum	0,2	2,9	2,2	1,5	0,9	2,0	0,9	1,0	1,0	0,8	0,2	0,8
2014	Maximum	51,9	12,2	11,8	8,4	3,6	2,9	9,8	48,4	11,8	11,8	3,6	53,9
	Moyenne	10,7	7,7	4,0	2,7	2,2	2,4	2,2	5,4	2,4	2,0	1,6	7,0

		Précipitation (mm)											
		Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
2010	Minimum	0	0	0	0	0	0	0	0	0	0	0	0
	Maximum	3,5	10,4	7,0	3,5	5,2	14,5	4,4	8,7	13,7	6,1	10,4	7,0
	Moyenne	0	0	0	0	0	0	0	0	0	0	0,9	0
2011	Minimum	0	0	0	0	0	0	0	0	0	0	0	0
	Maximum	7,8	4,4	7,8	5,2	2,6	8,7	11,3	17,1	11,3	4,4	5,2	8,7
	Moyenne	0	0	0	0	0	0	0	0	0	0	0	0,9
2012	Minimum	0	0	0	0	0	0	0	0	0	0	0	0
	Maximum	7,8	3,5	7,0	4,4	7 , 8	11,9	12,8	9,6	20,6	12,2	6,1	7,8
	Moyenne	0	0	0	0	0	0	0	0	0	0,9	0	0
2013	Minimum	0	0	0	0	0	0	0	0	0	0	0	0
	Maximum	3,5	4,4	6,1	4,4	5,2	6,1	24,1	7,0	9,6	12,8	8,7	11,3
	Moyenne	0	0	0	0	0	0	0	0	0	0	0	0
2014	Minimum	0	0	0	0	0	0	0	0	0	0	0	0
	Maximum	6,1	6,1	6,1	6,1	6,1	16,3	13,1	23,2	7,0	4,4	5,2	7,0
	Moyenne	0	0	0	0	0	0	0	0	0	0	0	0

		Hauteur d'eau dans le forage (m)											
		Janv.	Fév.	Mars	Avr.	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.
2010	Minimum	57	57	57	57	57	57	57	57	57	57	57	57
	Maximum	57	57	57	57	57	57	57	57	57	57	57	57
	Moyenne	57	57	57	57	57	57	57	57	57	57	57	57
2011	Minimum	57	57	57	57	57	57	57	57	56	56	56	56
	Maximum	57	57	57	57	57	57	57	57	57	56	56	56
	Moyenne	57	57	57	57	57	57	57	57	56	56	56	56
2012	Minimum	56	55	55	55	55	55	55	55	55	55	56	56
	Maximum	56	56	55	55	55	55	55	55	55	56	56	57
	Moyenne	56	55	55	55	55	55	55	55	55	55	56	56
2013	Minimum	57	57	57	58	58	58	58	58	57	57	57	57
	Maximum	57	57	58	58	58	58	58	58	58	58	58	57
	Moyenne	57	57	57	58	58	58	58	58	58	57	57	57
2014	Minimum	57	57	58	58	58	58	58	58	58	57	57	57
	Maximum	57	58	58	58	58	58	58	58	58	58	57	57
	Moyenne	57	57	58	58	58	58	58	58	58	57	57	57



Représentation des 22 évènements sélectionnés pour la modélisation.

Figure 83: Illustration de l'évènement 1.



Figure 84: Illustration de l'évènement 2



Figure 85: Illustration de l'évènement 3



Figure 86: Illustration de l'évènement 4



Figure 87: Illustration de l'évènement 5



Figure 88: Illustration de l'évènement 6



Figure 89: Illustration de l'évènement 7



Figure 90: Illustration de l'évènement 8



Figure 91: Illustration de l'évènement 9



Figure 92: Illustration de l'évènement 10



Figure 93: Illustration de l'évènement 11



Figure 94: Illustration de l'évènement 12



Figure 95: Illustration de l'évènement 13



Figure 96: Illustration de l'évènement 14



Figure 97: Illustration de l'évènement 15



Figure 98: Illustration de l'évènement 16



Figure 99: Illustration de l'évènement 17



Figure 100: Illustration de l'évènement 18



Figure 101: Illustration de l'évènement 19



Figure 102: Illustration de l'évènement 20



Figure 103: Illustration de l'évènement 21



Figure 104: Illustration de l'évènement 22

Présentation de l'algorithme de rétropropagation.

Afin de modifier les paramètres W du modèle pour chaque itération de l'apprentissage, le gradient de la fonction de coût doit être calculé. Pour ce faire, l'algorithme de rétropropagation est utilisé (Rumelhart, D. E., & *al.* (1986). Nous utiliserons les méthodes de second ordre qui sont plus rapides pour la minimisation de la fonction de coût. Ainsi, en se basant sur les travaux de Press, W.H., & *al.* (1996), le modèle se trouve en un point

 γ supposé proche d'un point Ω correspondent au minimum de la fonction de coût. Si l'on émet l'hypothèse

que la fonction de coût est quadratique autour du point Ω , il est possible en négligent les termes d'ordre supérieur de décomposer la fonction de coût en séries de Taylor jusqu'au deuxième ordre. Pour approcher le modèle du minimum Ω , l'incrémentation se fait selon δ_w :

$$j(\gamma + \delta w) = j(\gamma) + \nabla j(w) \cdot \delta w + \frac{1}{2} \delta w^{t} \cdot H(w) \cdot \delta w$$
 Équation 1

Avec j(c) la fonction de coût. t l'opérateur transposé et H la matrice hessienne. On retrouve également γ le point où se trouvent le modèle et Ω le point minimum de la fonction de coût.

En partant du principe que les dérivées sont nulles au niveau du minimum de la fonction de coût, on trouve que :

$$\frac{\delta j(\gamma + \delta w)}{\delta w} = 0 = \nabla j(w) + H(w).\delta w$$
 Équation 2

Mais également :

$$\delta w = -H^{-1}(w) \nabla j(w)$$
 Équation

Ainsi, l'application de ces règles permet au modèle de converger vers le minimum de la fonction de coût. Avec les informations précédentes, on peut organiser l'apprentissage du modèle en différentes étapes:

- 1. Initialisation aléatoire des paramètres w du modèle
- 2. Calcul de la sortie du modèle avec les paramètres précédemment fixés
- 3. Calcul de la fonction de coût j(w) et de son gradient $\nabla j(w)$
- 4. Incrémentation des paramètres w en fonction du gradient de la fonction de coût
- 5. Reprise de l'apprentissage à l'étape 2 avec les nouveaux paramètres $_{\mathcal{W}}$ fixés durant l'étape 4 jusqu'à atteindre le minimum Ω de la fonction de coût

C'est ainsi en répétant les étapes précédemment cité, en utilisant l'algorithme de Levendberg-Marquart et en diminuant la fonction de coût que l'ont obtient un modèle optimal. Cependant, comme nous allons le voir dans la partie suivante, l'apprentissage devra suivre certaines règles pour éviter que le modèle n'effectue un surapprentissage des données à sa disposition et ne soit ainsi plus capable de généraliser.

3



Figure 105 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 2



Figure 106 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 3



Figure 107 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 6



Figure 108 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 10



Figure 109 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 11



Figure 110 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 16



Figure 111 : Résultat de la décomposition en ondelette continue de la turbidité observée (graphique de gauche) et du résidu de prédiction à 12h du modèle à deux branches dirigé (graphique de droite). Evénement 22



Figure 112 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 2. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.



Figure 113 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 3. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.



Figure 114 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 6. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.


Figure 115 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 10. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.



Figure 116 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 11. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.



Figure 117 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 16. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.



Figure 118 : Analyse du résidu issu de la soustraction de la turbidité observée et de la turbidité prévue par le modèle dirigé à deux branches pour la prédiction à 12h de l'événement 22. (a) comparaison entre turbidité prédite et turbidité observée, (b) résidu de prévision et (c) assemblage des niveaux de décomposition M1, M2 et M3.

ANNEXE 11

Article soumis à la conférence Eurokarst 2018

Limits in Using Multiresolution Analysis to Forecast Turbidity by Neural Networks. Case study on the *Yport* Basin (*Normandie*-France)

Michaël Savary^{1,2}, Anne Johannet², Nicolas Massei¹, Jean Paul Dupont¹, Emmanuel Hauchard^{1,3}

¹Normandie University, UNIROUEN, UNICAEN, CNRS, M2C, 76000 Rouen, France

² LGEI, IMT Mines Alès, Univ Montpellier, 6 avenue de Clavières, 30 319 Alès Cedex, France

³ Communauté d'agglomération Havraise, 19 Rue Georges Braque, 76600 Le Havre, France

Presenting author: michaël.savary@mines-ales.fr

Corresponding author: anne.johannet@mines-ales.fr

Abstract Approximately 25% of the world population drinking water depends on karst aquifers. Nevertheless, due to their poor filtration properties, karst aquifers are very sensitive to pollutant transport and specifically to turbidity. As physical processes involved in solid transport (advection, diffusion, deposit...) are complicated and badly known in underground conditions, a black-box modelling approach using neural networks is promising. Despite the well-known ability of universal approximation of multilayer perceptron, it appears difficult to efficiently take into account hydrological conditions of the basin. Indeed these conditions depend both on the initial state of the basin (schematically wet or dry), and on the intensity of rainfalls. In this context, the present paper addresses the application of the multiresolution analysis to decompose the turbidity on several time scales in order to better consider various phenomena at various time scales. Because of "border effects" usually neglected by authors, a specific adaptation was shown as necessary that diminish the quality of results for real-time forecasting. Decomposing turbidity using multiresolution analysis adds thus questionable improvements.

I. Introduction

25% of the world population drinking water comes from karst aquifers. The comprehension and the protection of these aquifers appear thus as crucial. In *Normandie* (North-West of France), the highly karstified chalk aquifer is the principal exploited aquifer. For example, the *Yport* pumping well, fed by a karst aquifer, provides half of *Le Havre* conurbation (236,000 inhabitants) tap water. Due to the karstification, connections between surface water and underground water induce turbidity occurrences that decrease the water quality. It is thus very important for manager to be able to predict the occurrence of turbidity events. Processes involved in the karst aquifer make the turbidity difficult to model because of both the non-linearity of the rainfall/turbidity relation and the multiplicity of variables and phenomena. In this context, the aim of this work is to predict turbidity peaks in order to help managers to decrease the impact of turbidity on water treatment. Because of the lack of accurate physical description of the karst system (for example no rating curve is available at the place of pumping), neural networks models are chosen to achieve prediction using black-box models.

In previous works, the prediction of turbidity at 12h lead-time was studied using two kinds of neural networks architectures: the classical multilayer perceptron, and the so-called *two-branches* model. This previous work highlighted that the two-branches model was better to predict 100 NTU threshold surpassing.

In this context, the present chapter addresses the application of the multiresolution analysis to decompose the turbidity on several time-scales in order to better consider various phenomena at various time scales. A focus is done on limitations of the multiresolution analysis in the context of real-time prediction. The identified limitations are discussed in order to adapt the feed-forward model. For this reason it appears that results are not as much improved using multiresolution analysis as they are in the literature.

The chapter is organized as follows: after the introduction, the first part focuses on the estimation of turbidity by machine learning; in the second part the area of interest is presented: the *Yport* pumping well in *Normandie* (*France*). Then the Material and methods section presents neural networks modeling and multiresolution analysis; results are finally presented and discussed in section IV.

II Estimating turbidity by machine learning

Turbidity is the cloudiness of a fluid caused by suspended particles. The unit of turbidity is the Nephelometer Turbidity Unit (or NTU). Water is considered as drinkable when turbidity is inferior to 1 NTU.

At present, physical modeling of turbidity can't be successfully performed due to the lack of knowledge about physical properties of underground circulations. For this reason, statistical approaches in the framework of systemic modeling appear as valuable. Amongst them, one can note first the study of discharge-turbidity relation, called *sediment-rating* curve, thanks to several tools and strategies: SVM [1], multi-linear regression [2], correlation analysis [3], and finally neural networks [2]. Because of their flexibility, neural networks were applied to various kinds of relations: sediment-rating curve, chemistry-turbidity relation (conductivity, temperature, pH, ammonium concentration...). Neural networks were proved better than others methods by [4], [5]. When discharge measurements are not available, rainfall-turbidity relation could be investigated using a rainfall-runoff model [6]. Synthetically, it appears that modeling the direct relation between rainfall and turbidity is little published, and for the best of our knowledge, except [7], no work addresses the modeling of processes occurring at short time scales (*e.g.* 1hour).



Fig. 1. Yport alimentation basin and rain gauges localizations

III. Yport pumping well site of study

The Seine Maritime Regional context and Yport pumping well

The *Yport* basin is localized in *Normandie* in the *Pays de Caux* at the west of the *Seine Maritime* department, France. The yearly mean temperature of this region is between 10°C and 12°C. The yearly mean rainfall is between 600 mm and 1100 mm. The climate is both continental and oceanic. In 2012, the alimentation basin area was estimated at 320 km² [8].

The *Yport* pumping well is exploited by the *Communauté d'Agglomération Havraise* (CODAH) to supply *Le Havre* inhabitants. The implementation of the *Yport* water station was motivated by the *Yport* Fountain localized on the tidal zone of *Yport* beach. With a flow rate estimation of its outlet spring, between 1,5 m³/s [9] and 2.8 m³/s [10], the alimentation basin of the *Yport* Spring appears valuable for tap water exploitation. The *Yport* pumping well was thus implanted in 1994 directly in a major karst drain.

Turbidity, water production and uncertainty

At *Yport* Plant, turbidity is measured with a nephelometer considered as well calibrated. Thus the estimation of uncertainty is the one given by the manufacturer: 2% for turbidity between 0 NTU and 40 NTU and 5% for turbidity superior to 40 NTU. When turbidity exceeds the threshold of 100 NTU water must decant longer. This diminishes the output flow of 30% to 20%. Being able to anticipate the 100 NTU threshold thus would allow to anticipate by: *(i)* stocking more water, *(ii)* implementing the chain of quality control and treatment.

Two goals will be thus addressed in this study: prediction of turbidity peaks, and 100 NTU threshold overtaking. The necessary anticipation time should be at least 12 hours lead-time, or, if possible 24 hours lead-time.

Database

Event	ts without	t turbic	lity pea	ık	Events with turbidity peak(s)							
Even	Duration	n			Event	Duration	n					
Even	(hours)	Max	Min	Cumul	Even	(hours)	Max	Min	Cumul			
1	288	7	1	15.9	2	624	302	1.54	41.3			
4	384	10	0.9	14.1	3	1008	135	0	26.7			
5	336	8	1.5	17.5	6	720	245	1.5	42.0			
8	360	27	1	22.5	7	744	85	0.1	19.2			
9	384	9	1.0	20.2	10	576	256	0.9	24.9			
12	456	12	0.8	28.7	11	744	308	0.9	54.8			
13	576	13	0.9	30.8	16	648	405	0.8	53.8			
14	384	14	0.9	23.3	17	744	157	0.5	50.7			
15	600	15	0.9	31.7	18	744	87	2.2	42.8			
19	504	19	1.5	30.3	22	623	53	0.8	44.2			
20	576	20	0.9	40.8								
21	600	48	0.9	48.5								

Table 1. Simple statistics of events in the database

The database provided by the CODAH consists in hourly rainfalls coming from six rain gauges located in the alimentation basin since 2009, and hourly turbidity, measured in the karst drain of *Yport*, since 1993. Thanks to the database, an event is selected if the cumulative mean rainfall reaches more than 3.5 mm in a sliding time-window of 24h. Following this method, 22 events were selected: 12 events without turbidity peak and 10 events with turbidity peak. An event presents a turbidity peak when the maximum of the turbidity reaches 50 NTU. Selecting events thanks to the rainfall amount is better than selection them using only turbidity peaks because in this case, the model can learn the two configurations: (*i*) a rainfall event has a turbidity peak, and (*ii*) a rainfall event hasn't any turbidity peak.

Split of the database for Neural Network modelling

When designing a model using a neural network, it is necessary to split the database in several subsets: training set, stop set (usually called validation set in the literature) and test set. Depending on the test set, model design stage has its own splitting and consequently a different database and a different resulting model.

III Material and methods

III.1 Neural network modeling

Multilayer perceptron and two-branches model

The multilayer perceptron is the most known model in hydrology because of its properties of universal approximation [11] and parsimony [12]. Nevertheless in order to better manage the role of evapotranspiration that is predominant for turbidity, another model, based on multilayer perceptron, was preferred in this study: the twobranches model [13], [7]. It is shown in Fig. 2.



Fig. 2. Two-branches model, based on the multilayer perceptron.

Feed-forward vs. recurrent model

The feed-forward model is fed by only exogenous inputs. Specifically, added to exogenous variables (rainfall, temperature, evapotranspiration...), this model receives variables of the measured output, here the turbidity, at previous time steps (k-1, ...). In automatic control, this information can be considered as providing the state of the system. The feed-forward model can be mathematically explained as:

$$y^{k}(\mathbf{u}^{k},\mathbf{w}) = g_{NN}(y_{o}^{k-1},...,y_{o}^{k-r},\mathbf{u}^{k},...,\mathbf{u}^{k-m+1},\mathbf{w}),$$
(2)

where y^k is the estimated turbidity, g_{NN} is the non-linear function implemented by the neural network, k is the discrete time step, \mathbf{u}^k is the vector of exogenous variables (rainfalls, evapotranspiration, etc), r is the order of the model, m is the width of the sliding time-window of exogenous variables, w is the matrix of parameters, y_0^k is the measured (or observed) turbidity.

When turbidity measurements are corrupted by noise, the measured turbidity can be replaced by turbidity estimations calculated by the model itself, at previous times steps. The advantage of this model is that it takes better into account the dynamics of the system. Nevertheless, it is generally less effective because its feedback is corrupted by noise [14].

With the same notations, the recurrent model can be stated mathematically as:

$$\mathbf{y}^{k}(\mathbf{u}^{k},\mathbf{w}) = \mathbf{g}_{NN}(\mathbf{y}^{k-1},...,\mathbf{y}^{k-r},\mathbf{u}^{k-1},...,\mathbf{u}^{k-m+1},\mathbf{w})$$
(3)

Bias variance and regularization methods

Using neural networks, database is usually divided into three sets: a training set, a stop set, and a test set. The training set is used to calculate parameters through a training procedure that minimizes the mean quadratic error calculated on output neurons. In this study the Levenberg-Marquardt training rule was chosen [15]. The training is stopped thanks to the stop set, and the model quality is measured on the remaining part of the database: the test set, which is separate from the previous sets. The stop set is chosen as proposed in [16]. The model's ability to be efficient on the test set is called *generalisation*. Depending on the complexity of the model, overfitting can occur; this major issue in neural network modelling is called bias-variance trade-off [17]. This trap can be avoided using regularization methods [18] and rigorous model selection.

Model selection

[18], [19] showed that overfitting can be avoided thanks to a rigorous model selection. This consists to choose not only the number of neurons in the hidden layers but also the order of the model and the dimension of input variables vectors using cross-validation. By this way, numerous combinations of variables are tried, and the one minimizing the variance is chosen. Another hyper-parameter to choose is the initialization of the parameters. This can be done thanks to cross-validation in general case. Nevertheless it was shown by [20] that a more robust model could be designed using an ensemble strategy. Ten models are thus trained and the maximum of the output, at each time step, is taken. The model design is thus made as the following: first the hidden neurons number, then the number of input variables (selecting m), and lastly the order (r).

III.2 Multiresolution analysis

Multiresolution analysis belongs to wavelet analysis. It is a tool used for analyzing distribution of frequencies in a signal, localized in time. To be short, multiresolution analysis allows creating several signals, called components, each one describing the content of the original signal for a range of frequencies. In the present case, turbidity events contain 1024 samples and are decomposed in 10 components of 2014 samples. The component 1 contents high frequency and the component 10 low frequency information. The interest of this decomposition is that the behaviors can be separated in different signals depending on their temporal scale.

As the length of the chapter is short, we propose to the interested reader to refer to [21], [22] for a presentation of the technics.

Generally neural networks and multiresolution analysis are used jointly to provide prediction of discharges. Numerous papers present how the multiresolution, used as pretreatment before prediction help to improve significantly the results. Nevertheless three important issues are never been addressed: (*i*) the necessity to dispose of a really independent test set (not affected by training data during the low frequency decomposition), (*ii*) the necessity to design decomposition using a causal filter, and (iii) the consequence of border effects inherent to the decomposition: border effects implies a modification of the signal near its beginning and near its end. Thus, when the real time prediction is performed, the user is always at the end of the signal: the used value is thus different from the value used during training. This drawback is only concerning the feed-forward models: those that are well working in the literature. One of the goal of the chapter is thus to evaluate the influence of the border effect on the quality of prediction of both kind of models: feed-forward and recurrent.



Fig. 3. Prediction strategy

For this reason we proposed in this work to design several 2-branches predictors, each one receiving the full rainfalls and predicting one particular component of the turbidity, or the addition of several components. Two strategies of prediction are possible: (*i*) design 10 models, one for each detail, (*ii*) design only necessary models in order to diminish the complexity of the whole model and to maximize the quality of generalization. After an analysis of the information included in different components (by continuous wavelets), the global architecture shown in Fig. 3 was proved more efficient [23].

Quality criteria

In order to assess the performance of models, we propose to use the Nash-Sutcliff efficiency, or R^2 [24], is the most commonly used criterion in hydrology.

The nearest than 1 the R^2 is, the best the results are.

IV. Results

After a rigorous design in order to maximize the generalization, results of different strategies are evaluated on several test events. As shown in Table 2 and Fig. 3, it appears that adding multiresolution analysis to feed-forward model degrades the performances because of the border effects. This effect was never cited in the literature because the predictions were not realized in real-time, but on past recorded data, where the past and the future data were available. There was thus no border at the instant of prediction. It appears also that the multiresolution improves a little recurrent models.

	Tw	o-branches	Using multiresolution						
Event in	Dogurrant	Food Forward	Model with	all details	Model with grouped details				
test	Recurrent	reed Forward	Recurrent	Feed Forward	Recurrent	Feed Forward			
$R^2 T2$	0.93	0.63	0.84	0.52	0.94	0.75			
<i>R</i> ² T10	0.56	0.65	0.50	0.59	0.19	0.69			
<i>R</i> ² T11	0.91	0.46	0.85	0.24	0.86	0.39			
<i>R</i> ² T16	0.75	Х	0.47	0.42	0.45	0.34			
<i>R</i> ² T22	0.89	0.65	0.76	0.44	0.87	0,71			

Table 2. Nash criteria (R^2) for several events in test

V. Conclusion

Karst water is a valuable source of drinkable water, but karst aquifers are very sensitive to pollutant transport and specifically to turbidity. The goal of this study was thus to improve the prediction of turbidity karst groundwater. Neural networks were chosen because of their ability to predict efficiently using a database without having information about physical processes. Suggested by a wide range of papers in the hydrological literature, the multiresolution analysis was used in conjunction with neural modeling in order to better express the information over various time scale components. After having designed models and compared the performances of models using multiresolution and the classical ones, it appeared that multiresolution degraded the performance of the feed-forward models while it improved slightly the recurrent models. The explanation concerns the border effects that, for the best of our knowledge, are never cited in the literature. This negative results open research to find others strategies in order to counterbalance the border effects and provide improved models for karst water managers.

Acknowlegments

The authors would like to thank the CODAH for providing rainfall and turbidity data. The *Normandie* Region and *Seine-Normandie* Water Agency are thanked for the co-funding of the study. We are also very grateful to S. Lemarie and J. Ratiarson of for the very helpful discussions they helped organize. Our thanks are extended to D. Bertin in the design and implementation of the NN simulation tool.



Fig. 3. Illustration of 12h predictions on the event 11: (top) Classical Multilayer Perceptron, (middle) two-branches model, and (bottom) prediction using multiresolution as shown in Fig. 3.

References

- 1. Kisi, O., et al (2012). Suspended sediment modeling using genetic programming and soft computing techniques. Journal of Hydrology, 450, 48-58.
- Rajaee, T., et al (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Science of the total environment, 407(17), 4916-4927.
- 3. Massei, N., et al (2006). Investigating transport properties and turbidity dynamics of a karst aquifer using correlation, spectral, and wavelet analyses. J. of Hydrology, 329 1-2, 244-25.
- Nieto, P. G., et al. (2014). Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: A case study in Northern Spain. Ec. Engineering, 73, 192-200.
- Iglesias, C., et al. (2014). Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain. Water resources management, 28(2), 319-331.
- Beaudeau, P., et al. H. Q. (2001). Forecasting of turbid floods in a coastal, chalk karstic drain using an artificial neural network. Ground Water, 39(1), 109-118.
- Savary M., et al. (2017) Operational Turbidity Forecast Using Both Recurrent and Feed-Forward Based Multilayer Perceptrons. In: Rojas I., Pomares H., Valenzuela O. (eds) Advances in Time Series Analysis and Forecasting. ITISE 2016.
- SAFEGE (2012) Étude du Bassin d'Alimentation du forage d'Yport "Phase 1 :Délimitation du bassin d'alimentation et analyse de la vulnérabilité intrinsèque du basin". 112p.
- Hole, J. P., Roux. J. C., 1978. Vulnérabilité aux pollutions du bassin hydrogéologique des sources d'Yport (Seine-Maritime), Rapport BRGM. 81.
- 10. Gaillard, T., Hauchard, E., & Roux, J. C. (2012) Les Fontaines d'Yport (Seine-Maritime), émergences majeures du littoral normand crayeux: Exploitation et vulnérabilité de la ressource en eau. *Ressources et gestion des aquifères littoraux. Cassis 2012*. 235-244.
- 11. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366.
- 12. Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. Information Theory, IEEE Transactions on, 39(3), 930-945.
- Johannet, A., Vayssade, B., & Bertin, D. (2008). Neural networks: from black box towards transparent box. Application to evapotranspiration modeling. International Journal of Computational Intelligence, 4(3), 163-170.

Dreyfus, G. (2005). Neural networks: methodology and applications. Springer Science & Business Media, 497p.

- Artigue, G., et al. (2012). Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin. NHESS, 12(11), 3307-3324.
- 15. Dreyfus, G. (2005). Neural networks: methodology and applications. Springer Science & Business Media, 497p
- 16. Toukourou M., et al.. 2011. Rainfall-runoff Modeling of Flash Floods in the Absence of Rainfall Forecasts: the Case of "Cévenol Flash Floods", Applied Intelligence, 35 2,178-189.
- 17. Geman, S., Bienenstock E. and R. Doursat, (1992). Neural networks and the bias/variance dilemma, Neural Computation 4(1), 1-58.
- 18. Kong-A-Siou, L., et al. (2012). Optimization of the generalization capability for rainfall–runoff modeling by neural networks: the case of the Lez aquifer (southern France). Environmental Earth Sciences, 65(8), 2365-2375.
- 19. Kong-A-Siou, L., et al.. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez basin (southern France). Journal of Hydrology 403, 367–380.
- 20. Darras, T., et al. (2014). Influence of the Initialization of Multilayer Perceptron for Flash Floods Forecasting: How Designing a Robust Model, (ITISE 2014), 687-698
- D. Labat, R. Ababou, and A. Mangin "Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution," In Journal of Hydrology, Vol. 238, December 2000, pp. 149-178.
- Johannet, A., et al. Prediction of spring discharge by neural networks using orthogonal wavelet decomposition. In IJCNN 2012. 10.1109/IJCNN.2012.6252620.
- 23. Savary, M., PhD Dissertation (2018), to be defended.
- 24. Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. Journal of hydrology, 10(3), 282-290.

ANNEXE 12

Savary, M., Johannet, A., Massei, N., Dupont, J. P., & Hauchard, E. (2018). Operational Turbidity Forecast Using both Recurrent and Feed-Forward *based* Multilayer Perceptrons. Journal of hydrology.

Operational Turbidity Forecast Using both Recurrent and Feed-Forward *based* **Multilayer Perceptrons.**

Michaël Savary^{1,2}, Anne Johannet², Nicolas Massei¹, Jean-Paul Dupont¹, Emmanuel Hauchard³

¹Rouen University, M2C Laboratory, Place E. Blondel, 76821 Mont-Saint-Aignan, France {michael.savary1, nicolas.massei,jean-paul.dupont}@univ-rouen.fr ²Ecole des mines d'Alès, LGEI, 6 avenue de Clavières, 30 319 Alès Cedex, France

anne.johannet@mines-ales.fr

³ Communauté d'Agglomération Havraise, 19 Rue Georges Braque, 76600 Le Havre, France

emmanuel.hauchard@codah.fr

Abstract. Approximately 25% of the world population drinking water depends on karst aquifers. Nevertheless, due to their poor filtration properties, karst aquifers are very sensitive to pollutant transport and specifically to turbidity. As physical processes involved in solid transport (advection, diffusion, deposit...) are complicated and badly known in underground conditions, a black-box modelling approach using neural networks is promising. Despite the well-known ability of universal approximation of multilayer perceptron, it appears difficult to efficiently take into account hydrological conditions of the basin. Indeed these conditions depend both on the initial state of the basin (schematically wet or dry), and on the intensity of rainfalls. To this end, an original architecture has been proposed in previous works to take into account phenomenon at large temporal scale (moisture state), coupled with small temporal scale variations (rainfall). This architecture, called hereafter as "two-branches" multilayer perceptron is compared with the classical two layers perceptron for both kinds of modelling: recurrent and non-recurrent. Applied in this way to the *Yport* pumping well (*Normandie*, France) with 12 hours lag time, it appears that both models proved crucial information: amplitude and synchronization are better with "two-branches" feed forward model when thresholds surpassing prediction is better using classical feed forward perceptron.

Keywords. Neural networks, Recurrent, Feed-forward, turbidity, karst.

1 Introduction

Turbidity is crucial for water quality because it is generally the indicator of the contamination of the underground water by surface water, potentially polluted by phytosanitary products or biological organisms. When turbid water is pumped, complex and expensive treatments are engaged. Predicting turbid events allows thus optimising treatment processes in order to provide drinking water satisfying standards. Nevertheless, both complexity of the hydrologic system and difficulty to quantify physical behaviours, prevent designing operational physical models; statistical framework and specifically machine learning appear thus as a complementary solutions. In this context, the present study is one of the early studies devoted to prediction of rainfall-turbidity relation.

The paper is organized following six parts: after the introduction, turbidity and the state of the art are described. Presentation of neural networks follows in section 3, and the *Yport (Normandie –* France) watershed and database are presented. Section 5 presents results and discussion, and the conclusion shows in section 6 that an original coupling between both recurrent and non-recurrent models allows significant anticipation on the occurrence of turbid event.

2 Estimating turbidity by machine learning

2.1 Definition of the turbidity

Turbidity is the cloudiness of a fluid caused by suspended particles. The unit of turbidity is the Nephelometer Turbidity Unit (or NTU) and water is considered as potable when turbidity is inferior to 1 NTU. Various measurement methods develop analysis of the interactions between light beams and suspend mater. All of them need to be carefully calibrated in the proper range of: pH, temperature, conductivity... and suspended particles

(size, shape, color, number...). Due to complex suspended particles composition, a direct relation between NTU and suspended sediment mass is not possible. This complicates modeling of turbidity and makes it especially difficult to perform.

2.2 State of the art

At present, due to the lack of knowledge about physical properties of underground circulations, physical modeling of turbidity can't be successfully performed. For this reasons other strategies were developed using statistical approaches in the framework of systemic modeling. Amongst them, one can note first the exploration of the causal relation between velocity of water and turbidity. The relation between discharge and turbidity, called sediment-rating curve, is thus established using various tools and strategies: SVM [1], multi linear regression [2], correlation analysis [3], neural networks... Because of their flexibility, neural networks were applied to various kinds of relations: sediment-rating curve, chemistry-turbidity relation (conductivity, temperature, pH, ammonium concentration...). Neural networks were proved better than others methods by [4], [5]. When discharge measurements are not available, rainfall-turbidity relation could be investigated using a rainfall-runoff model [6]. Synthetically, it appears that modeling the direct relation between rainfall and turbidity is little published, for the best of our knowledge, due to the complexity of the relation.

2.3 Turbidity, uncertainty and water production

At *Yport* Plant, turbidity is measured with a nephelometer (which analyses the light scattered at 90° by the suspended particles). The nephelometer is considered as well calibrated, thus the estimation of uncertainty is the one given by the manufacturer: 2% for turbidity between 0 NTU and 40 NTU and 5% for turbidity superior to 40 NTU. Regarding the production process, when turbidity exceeds the threshold of 100 NTU it is necessary to make water longer decanting. This diminishes the output flow of 20% to 30%. Being able to anticipate the 100 NTU threshold thus would allow to anticipate by: (*i*) stock more water, (*ii*) assessing the quality of treatment chain.

3 Design of the model

3.1 Multilayer perceptron

The multilayer perceptron was chosen due to its properties of universal approximation [7] and of parsimony [8]. The model is shown in Fig. 1. It is fed by exogenous variables, in this study: rainfalls (\mathbf{u}_r), evapotranspiration (\mathbf{u}_e) and observed turbidity (y_o), and delivers, as output, the estimated variable of interest (y), k is the discrete time step. As this model is very famous it is not detailed herein; for more information on multilayer perceptron, the reader can refer to [9].



Fig. 1. Standard multilayer perceptron

3.2 Specific architectures

As the behavior of the rainfall-turbidity relation is dynamic, it is important to take into account information about the state of the basin, this can be done usually using two kinds of models: feed-forward and recurrent models [10].

- Feed-forward/recurrent models.

variables, w is the matrix of parameters.

The feed-forward model is a multilayer perceptron fed by only exogenous inputs. Specifically, added to exogenous variables (rainfall, temperature, evapotranspiration...), this model receives variables of the measured output, here the turbidity, at previous time steps $(k-1, \ldots k-r)$. In automatic control, this information can be considered as providing the state of the system (position, speed, acceleration). The feed-forward model can be mathematically explained as:

$$y^{k}(\mathbf{u}^{k},\mathbf{w}) = g_{NN}(y_{o}^{k-1},...,y_{o}^{k-r},\mathbf{u}^{k},...,\mathbf{u}^{k-m+1},\mathbf{w}),$$
(1)

where y^k is the estimated turbidity, g_{NN} is the non-linear function implemented by the neural network, k is the discrete time step, y_0^k is the measured (or observed) turbidity, \mathbf{u}^k is the vector of exogenous variables (rainfalls, evapotranspiration, etc), r is the order of the model, m is the width of the sliding time window of exogenous

When turbidity measurements are corrupted by noise, these data can be replaced by turbidity estimations calculated by the model, at previous times steps. The advantage of this model is that it takes better into account the dynamics of the system. Nevertheless, it is generally less effective for predicting the future as illustrated by [10].

With the same notations, the recurrent model can be stated mathematically as:

$$y^{k}(\mathbf{u}^{k},\mathbf{w}) = g_{NN}(y^{k-1},...,y^{k-r},\mathbf{u}^{k-1},...,\mathbf{u}^{k-m+1},\mathbf{w})$$
(2)

- One-branch/two-branches.

A specific *ad hoc* model was built in order to represent a conceptual hypothesis about the role of evapotranspiration and rainfalls on the hydrogeological basin [11].



Fig. 2. Two-branches multilayer perceptron

In this view, the process is split in, (*i*) the rainfall-turbidity relation and (*ii*) the evapotranspiration influence on the previous relation. The rainfall-turbidity relation is fast and controlled by recent rainfalls while the potential evapotranspiration (ETP) has slower dynamics. Because of the different dynamics it could be advantageous to calculate a nonlinear transformation for each of the processes (ETP or rainfalls) before taking them into account in a coupled model. The model presented in Fig. 2 aims at implement this strategy; it is composed of two branches: one for the rainfall-turbidity relation (upper branch), the other for the evapotranspiration (lower branch). Finally both branches are connected in a supplementary non-linear hidden layer. Hidden layers are composed of non-linear neurons based on arctg function.

3.3 Bias variance and regularization methods

Being statistical models, neural networks are designed in relation to a database. This database is usually divided into three sets: a training set, a stop set, and a test set. The training set is used to calculate parameters through a training procedure that minimizes the mean quadratic error calculated on output neurons. In this study the Levenberg-Marquardt training rule was chosen [9]. The training is stopped thanks to the stop set (usually called validation set), and the model quality is measured on the remaining part of the database: the test set, which is separate from the previous sets. The choice of the stop set is crucial as it influences the model in a very important way. For this reason we proposed in [12] to choose the stop set, for each model, in taking the set having the best

score in validation. This choice guarantees that there is a strong coherence between the training set and the stop set.

The model's ability to be efficient on the test set is called *generalisation*. One has to underline that the training error is not an efficient estimator of the generalisation error because the efficiency of the training algorithm makes the model specific to the training set. This specialisation of the model over the training set is called overtraining. Overtraining is exacerbated by large errors and uncertainties in field measurements; the model then learns the specific realization of noise in the training set. This major issue in neural network modelling is called bias-variance trade-off [13]. This trap can be avoided using regularization methods, particularly cross-validation [14], [15].

3.4 Model selection

[15], [16] showed that overfitting can be avoided thanks to a rigorous model selection. This consists to choose not only the number of neurons in the hidden layers but also the order of the model and the dimension of input variables vectors using cross-validation. By this way, numerous combinations of variables are tried, and the one minimizing the variance is chosen. Another hyper-parameter to choose is the initialization of the parameters. This can be done thanks to cross-validation in general case. Nevertheless it was shown by [17] that a more robust model could be designed using an ensemble strategy. Ten models are thus trained and the median of the output, at each time step, is taken. The model design is thus made as the following: first the hidden neurons number, then the number of input variables (selecting m), and lastly the order (r).

3.5 Quality criteria

In order to assess the performance of models, several quality criteria are used: R^2 , persistency and the percentage of pic discharge (P_{PD}).

The Nash-Sutcliff efficiency, or R^2 [18], is the most commonly used criterion in hydrology.

$$R^{2} = 1 - \frac{\sum_{k=1}^{n} (y_{o}^{k} - y^{k})^{2}}{\sum_{k=1}^{n} (y_{o}^{k} - \overline{y_{o}})^{2}}$$
(3)

The nearest than 1 the R^2 is, the best the results are. Nevertheless this criterion can reach good values even if the model proposes bad forecasts [19]. To avoid this problem, the persistency is used.

The Persistency: C_p [20], provides information on the prediction capability of the model compared to the naive forecast. The naive forecast postulates that the output of the process at time step k+l (where l is the lead-time) is the same than the value at time k. The nearest than 1 the persistence efficiency is, the best the results are. A positive result means that model prediction is better than the naive prediction.

$$C_{p} = 1 - \frac{\sum_{k=1}^{n} (y_{o}^{k+l} - y^{k+l})^{2}}{\sum_{k=1}^{n} (y_{o}^{k+l} - y_{o}^{k})^{2}}$$
(4)

The percentage of the Turbidity Peak: P_{TP} , inspired from [10], assesses performance of a model at the time of the peak. It calculates the ratio between the observed and forecast peak values. Calculation is visualized in Fig. 3; k_{max} is the instant of the peak. In Fig. 3, as there is two curves, there is consequently two different instants for the turbidity peak (one for observed, one for simulated peak).

$$P_{TP} = 100 \frac{y^{k_{max}}}{y_o^{k_{max}}} \tag{5}$$



Fig. 3. Definition of the *P*_{TP}

4 Site of study: *Yport* pumping well

4.1 Overview of the basin

Yport pumping well is situated in *Normandie* (North-West of France). Managed by the CODAH (*Communauté* d'*agglomération Havraise*), it delivers roughly half of *Le Havre* conurbation drinking water (236 000 inhabitants). The area of the alimentation basin is estimated to 320 km² and it is essentially devoted to agriculture. Rain falling on the basin is measured by six rain gauges (*Froberville, Annouville, Goderville, Anglesqueville, Manevillette* and *Etainhus*) as shown in Fig. 4. Drinking water is pumped in a well, dug in a natural underground conduit. The turbidity is recorded at the entry of the *Yport* treatment plant.



Fig. 4. Yport Basin: location of rain gauges.

4.2 Database

Rainfalls are measured by the six previously cited stations between 01/07/2009 and 28/04/2015. Turbidity was measured at *Yport* Plant between 23/10/1993 and 06/02/2015.

Database was hourly re-sampled from the original five minutes period for both turbidity and rainfall. Hourly rainfalls were obtained by addition, and turbidity by picking the maximum hourly value. Because of gaps in turbidity measurements, an event-based modelling approach was chosen. Events whose cumulative rainfalls exceeded 3.5 mm in 24 hours were extracted. This way of selection intended to avoid false positive (induced by a heavy rain without turbidity peaks). Finally, 22 events were extracted. The Table 1 presents these events. Amongst them, 10 events (events 2, 3, 6, 7, 10, 11, 16, 17, 18 and 22) present peaks of turbidity.

As explained in section 3, three sets were distinguished: test set (event 11), stop set (7, 10, 13, 16 and 17) and training set (the rest of the database). Event 11 is chosen for test set as it contains high and double peak of turbidity.

"Ev	vent" withou	ut turbidi	ty peak		Event with turbidity peak						
Event	Duration	Turbidity (NTU)		Rain (mm)	Ev	Duration	Turbidity (Rain (mm)			
	(nours)	Max	Min	Cumul	ent	(nours)	Max	Min	Cumul		
1	288	7.07	1	15.9	2	624	302.48	1.54	41.3		
4	384	9.82	0.91	14.1	3	1008	135.03	0	26.7		
5	336	7.71	1.52	17.5	6	720	245.38	1.53	42.0		
8	360	26.87	0.97	22.5	7	744	84.67	0.05	19.2		
9	384	9	1.00	20.2	10	576	256.15	0.92	24.9		
12	456	12	0.84	28.7	11	744	307.89	0.87	54.8		
13	576	13	0.86	30.8	16	648	405.25	0.81	53.8		
14	384	14	0.86	23.3	17	744	157.45	0.49	50.7		
15	600	15	0.85	31.7	18	744	86.67	2.18	42.8		
19	504	19	1.50	30.3	22	623	53.91	0.80	44.2		
20	576	20	0.89	40.8							
21	600	48.44	0.93	48.5							

 Table 1. Database composed of 22 events

5 Results

5.1 Selected architecture

Based on MLP, the four selected architectures are presented in Table 2. One can note that the two-branches feed-forward model is more parsimonious than the recurrent two-branches model, specifically regarding the number of hidden neurons.

		Re	current	Feed-f	orward
Parameters		MLP	Two branches	MLP	Two branches
	Rainfall layer	Х	15	Х	10
Hidden layer	Evapotranspiration layer	Х	1	Х	1
	Global layer	5	15	5	10
T	Rainfall	30	50	50	50
input windows	Evapotranspiration	3	3	3	3
widths	Order	1	1	1	10

Table 2. Models architectures

After selection of the best model and training, the test set was run (event 11), and forecasts for each model are shown in Fig. 5. They correspond to the prediction of an ensemble of ten models differing by random initialization of parameters. The grey line corresponds to the median of the output of the ensemble. The grey area around this line shows the "uncertainty" provided by the model (max and min of the prediction at each time step).

It appears on Fig. 5 that two-branches models seem working better that the standard multilayer perceptron and that the feed-forward models provide a good amplitude prediction of the maximum amplitude when the recurrent model delivers a good synchronization of the peaks. Quality criteria are provided in Table 3.

After this first step of validation, a kind of *cross-test* was performed in order to assess the quality of prediction on the whole database. The test is thus performed on each event of turbidity of the database each one its turn; "events" without turbidity, but high rainfall were not tried. We can note on Table 4 that, satisfactorily, the forecasting behavior is quite stable on the whole database. Moreover it appears clearly that the best model appears to be the feed-forward two branches model.

			P_{TP}	Nash	Persistency
Two bra	anches	Maximum	33.73	0.28	0.12
recurrent		Minimum	23.57	-0.17	-2.11
Two bra	anches	Maximum	103.40	0.81	0.31
feed-forw	ard	Minimum	70.37	-1.20	-1.20
MID		Maximum	27.71	0.37	0.37
MLP recurrent		Minimum	18.18	-0.23	-3.74
MLP	feed-	Maximum	70.40	0.85	0.34
forward		Minimum	50.17	-0.67	-0.66

Table 3. Quality criteria for the 10-ensemble model and the four architectures. Test on ev. 11.

Table 4. Models performance on the whole database for 12h lag time. The model named T_n is the model designed with the
event <i>n</i> in test. Best results are highlighted in bold. The median calculated over all events is shown in last row.

		Two branches feed- forward		Two re	branches	۱ feed-	ALP forward	MLP recurrent		
		P_{TP} Peak delay (h)		P _{TP}	Peak delay (h)	P_{TP}	Peak delay (h)	P_{TP}	Peak delay (h)	
T_2	Median	79.95	4	38.01	18	80.62	19	39.58	21	
T_3	Median	112.95	4	60.99	15	108.67	4	70.63	3	
T_6	Median	217.41 8		79.97	10	93.30	14	69.70	10.50	
T_7	Median	Stop set		75.06	2	82.89	14	75.44	3	
T_{10}	Median	86.04	7	40.92	4	50.38	50.38 14.50		Stop set	
T_{11}	Median	80.36	15	28.48	10	59.46	8.50	25.69	3	
T_{16}	Median	106.39	2		Stop set	65.52 14		44.16	10	
<i>T</i> ₁₇	Median	105.55	5	87.55	101	Stop set		50.21	9	
T_{18}	Median	98.17	97	152.86	3	89.84	102	79.53	0	
<i>T</i> ₂₂	Median	143.54	3	152.33	14	169.18	4	149.17	6	
Median		105	5	75	10	83	14	70	6	



Fig. 5. Measured (black) and forecast (grey) turbidity with a lag time of 12h. Test on the event 11. Uncertainty is shown in grey area.

As suggested in section 2, another way to assess the quality of the modeling approach for operational end users, is to focus on operational stakes. Regarding the *Yport* Plant it is important to be able to detect the occurrence of turbid events exceeding 100 NTU. Counting the number of false predictions can assess this. To this end, Table 5 presents the number of errors in warning: false positive and false negative. One can note that these false warnings are marginal for the threshold of 100 NTU. Regarding Table 5, it appears that the multilayer feed-forward and two-branches recurrent models are globally the best models to predict the thresholds surpassing.

Table 5. Prediction of the 100 NTU threshold surpassing. All events of the database are investigated, successively in test. The model designed with the event *n* in test is called T_n . F_p means *false positive* and F_n *false negative;* d_r is the delay for the rising part of the curve, and d_d for the decreasing part. Best values are highlighted in bold. The X means that no 100 NTU threshold surpassing in observed in simulated or observed data. In the last row, *M* means the average for F_p and F_n , and means the median for the delays d_r and d_d . Best results are highlighted in bold.

	Tw	o brancl forwa	nes feed urd	1-	Two	brancl	hes recurr	rent	Multi	ilayer _l fo	perceptron prward	feed-	Multilayer percept recurrent			otron
Т	$F_{\rm p}$	Fn	dr (h	d_d (ł	$F_{\rm p}$	Fn	$d_r(\mathbf{h})$	$d_d(h)$	$F_{\rm p}$	$F_{\rm n}$	$d_r(\mathbf{h})$	$d_d(\mathbf{h})$	$F_{\rm p}$	$F_{\rm n}$	$d_r(\mathbf{h})$	$d_d(h)$
2	0	0	6	11	0	0	6	-76	0	0	6	8	0	0	6	-53
3	1	0	0	3	1	0	-11	-4	1	0	7	4	0	0	0	-22
6	0	0	-4	10	0	0	-5	6	0	0	-4	9	0	0	-4	4
7		Sto	p set		X	X	X	X	X	X	X	X	X X X		X	
10	0	1	-2	10	0	0	-1	9	0	0	-9	11		Stop set		
11	0	1	3	6	0	1	-17	-31	0	0	-12 / 8	-3/4	0	2	Х	Х
16	0	0	14	18			Stop set		0	0	1	12	0	0	6	-19
17	1	1	Х	Х	0	1	Х	Х			Stop set		0	1	Х	Х
18	0	0	X	Х	0	0	Х	Х	0	0	Х	Х	1	0	Х	Х
22	1	0	Х	х	0	0	X	X	0	0	X	X	1	0	Х	Х
М	0.2	0.3	1.5	10	0.1	0.3	-5	-4	0.2	0	-1.5	8	0.3	0.4	3	-20

Synthetically, an operational tool based on a multi approach: multilayer feed-forward, two-branches feed-forward and recurrent models will be of great interest.

6 Conclusion

Due to the complex phenomena involved in turbidity, prediction is a very difficult task seldom investigated in bibliography. Nevertheless, as water policy imposes norms on turbidity, and because turbidity is usually associated with pollutants transport, end users must take this aspect into account. In this context this study aims to predict peaks of turbidity with 12 hours lag time.

Recurrent and feed-forward models were run and it was shown that, thanks to the design of a new architecture, taking into account explicitly the role of evapotranspiration, called two-branches network, and to a rigorous selection of the model, it will be possible to anticipate the instant and the amplitude of the peak of turbidity as well as 100 NTU thresholds surpassing. A synthesis of several MLP-based architectures will allow designing an efficient and operational tool for water managers. Future works will investigate longer horizon of prediction and the way to improve performances of the recurrent two branches model, which seems specifically promising.

Acknowlegments

The authors would like to thank the CODAH for providing rainfall and turbidity data. The *Normandie* Region and *Seine-Normandie* Water Agency are thanked for the co-funding of the study. We are also very grateful to S. Lemarie and J. Ratiarson of for the very helpful discussions they helped organize. Our thanks are extended to D. Bertin for his extremely fruitful collaboration in the design and implementation of the Neural Network simulation tool: RnfPro.

References

Kisi, O., Dailr, A. H., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. Journal of Hydrology, 450, 48-58.

Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M., & Nourani, V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Science of the total environment, 407(17), 4916-4927.

Massei, N., Dupont, J.P., Mahler, B.J., Laignel, B., Fournier, M., Valdes, D., Ogier, S., (2006). Investigating transport properties and turbidity dynamics of a karst aquifer using correlation, spectral, and wavelet analyses. J. of Hydrology, 329 1-2, 244-25.

Nieto, P. G., García-Gonzalo, E., Fernández, J. A., & Muñiz, C. D. (2014). Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: A case study in Northern Spain. Ecological Engineering, 73, 192-200.

Iglesias, C., Torres, J. M., Nieto, P. G., Fernández, J. A., Muñiz, C. D., Piñeiro, J. I., & Taboada, J. (2014). Turbidity prediction in a river basin by using artificial neural networks: a case study in northern Spain. Water resources management, 28(2), 319-331.

Beaudeau, P., Leboulanger, T., Lacroix, M., Hanneton, S., & Wang, H. Q. (2001). Forecasting of turbid floods in a coastal, chalk karstic drain using an artificial neural network. Ground Water, 39(1), 109-118.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366.

Barron, A. R. (1993). Universal approximation bounds for superpositions of a sigmoidal function. Information Theory, IEEE Transactions on, 39(3), 930-945.

Dreyfus, G. (2005). Neural networks: methodology and applications. Springer Science & Business Media, 497p. Artigue, G., Johannet, A., Borrell, V., & Pistre, S. (2012). Flash flood forecasting in poorly gauged basins using neural networks: case study of the Gardon de Mialet basin (southern France). Natural Hazards and Earth System Science, 12(11), 3307-3324.

Johannet, A., Vayssade, B., & Bertin, D. (2008). Neural networks: from black box towards transparent box. Application to evapotranspiration modeling. International Journal of Computational Intelligence, 4(3), 163-170.

Toukourou M., Johannet A., Dreyfus G., Ayral P.A. 2011. Rainfall-runoff Modeling of Flash Floods in the Absence of Rainfall Forecasts: the Case of "Cévenol Flash Floods", Applied Intelligence, 35 2,178-189.

Geman, S., Bienenstock E. and R. Doursat, (1992). Neural networks and the bias/variance dilemma, Neural Computation 4(1), 1-58.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society. Series B (Methodological), 111-147.

Kong-A-Siou, L., Johannet, A., Valérie, B. E., & Pistre, S. (2012). Optimization of the generalization capability for rainfall–runoff modeling by neural networks: the case of the Lez aquifer (southern France). Environmental Earth Sciences, 65(8), 2365-2375.

Kong-A-Siou, L., Johannet, A., Borrell, V., Pistre, S., 2011. Complexity selection of a neural network model for karst flood forecasting: The case of the Lez basin (southern France). Journal of Hydrology 403, 367–380.

Darras, T., Johannet, A., Vayssade, B., Kong-A-Siou, L. and Pistre, S. (2014). Influence of the Initialization of Multilayer Perceptron for Flash Floods Forecasting: How Designing a Robust Model, (ITISE 2014), Ruiz, IR, Garcia, GR Eds, 687-698

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I-A discussion of principles. Journal of hydrology, 10(3), 282-290.

Moussa, R., 2010. When monstrosity can be beautiful while normality can be ugly: assessing the performance of event-based flood models, Hydrological Sciences Journal, 55(6)2010, Special Issue: The court of miracles of hydrology, pp. 1074-1084.

Kitanidis, P. K., & Bras, R. L. (1980). Real-time forecasting with a conceptual hydrologic model: 2. Applications and results. Water Resources Research, 16(6), 1034-1044.

ANNEXE 13

Illustration de l'impact de la décomposition du signal sur les fonctions de transfert devant être interprétées par les modèles utilisant la multirésolution. Les données d'entrées et de sorties du système pour les différents modèles sont entourées en orange.



Prévision de la turbidité par apprentissage statistique : application au captage AEP d'Yport (Normandie)

<u>Résumé :</u>

Près de 25% de la population mondiale est alimentée par de l'eau en provenance d'aquifères karstiques. La compréhension et la protection de ces derniers apparait donc comme essentielle dans le cadre d'une augmentation des besoins en eau potable. De plus, une contamination des forages d'alimentation en eau potable par une eau turbide peut s'avérer fortement dommageable car entrainant une possible contamination des populations desservies.

Dans le cas de la Normandie, des coupures régulières son nécessaires afin de préserver la santé des habitants. La modélisation et la prédiction des augmentations de turbidité apparaissent comme un travail difficile du fait des nombreux phénomènes et paramètres régissant la turbidité ainsi que la non-linéarité de la réponse entre les précipitations et la turbidité.

Peu de modèles à l'heure actuelle ont été proposés pour représenter la relation liant la turbidité avec les précipitations. C'est ainsi, en s'intéressant au forage AEP d'Yport responsable de l'alimentation en eau potable de la ville du Havre, que nous proposons une application des réseaux de neurones pour la prévision de la turbidité.

Durant les travaux de thèse, nous avons mis en avant la nécessité d'effectuer des campagnes d'échantillonnages des produits phytosanitaires afin de permettre l'identification des éventuels proxies des produits phytosanitaires tel que la turbidité, les précipitations ou bien la conductivité. Par la suite, les travaux effectués dans cette thèse nous ont permis (i) de monter que les modèles par réseaux de neurones permettent de prévoir à 12h et 24h les variations de turbidité, (ii) de tester plusieurs voies d'amélioration de ces modèles, (iii) d'intégrer l'analyse multirésolution aux modèles par réseaux de neurones et pour finir (iiii) d'identifier un semi proxy des contaminations en produits phytosanitaires.

Mots clés: Turbidité, prévisions, réseaux de neurones, eau potable, produits phytosanitaires, POCIS, Yport.

Turbidity forecasting using neural network: case study of Yport drinking water pumping well (Normandie)

Abstract:

Approximately 25% of the world's population is supplied by water from karstic aquifers. The understanding and protection of these appears to be essential in the context of drinking water needs increasing. In addition, contamination of drinking water by turbid water can be highly damaging by resulting in possible contamination of the served populations.

In the case of Normandy, regular drinking water cut-off are necessary to preserve the health of the inhabitants. The modeling and prediction of turbidity event appears as a challenging work because of the number of phenomenon and parameters involves in turbidity variation as well as the non-linearity of the link between rainfall and turbidity.

Actually, few models have been proposed to represent the relationship between turbidity and rainfall. In this context, by focusing on Yport's pumping well which is responsible for Half of Le Havre city drinking water supply, we propose an application of neural networks for turbidity prediction.

During this thesis work, we emphasized the need to carry out sampling campaigns for phytosanitary products to enable the identification of possible phytosanitary product proxies such as turbidity, rainfall or conductivity. Subsequently, the work carried out in this thesis enabled us to (i) designed neural network models allow to predict at 12h and 24h the turbidity variations, (ii) test several ways to improve these models, (iii) integrate multiresolution analysis into neural networks models and finally (iiii) identify a semi proxy for phytosanitary product contamination.

Key words: Turbidity, forecasting, neural network, drinking water, phytosanitary product, POCIS, Yport.