

Faculté des bioingénieurs

Approche hybride pour la surveillance de la qualité de l'eau à grande échelle

Surveillance de la turbidité de l'eau de la
Meuse à l'aide de l'imagerie Sentinel-2

Auteur :	Axelle Dagnelie
Promoteurs :	Marnik Vanclooster (UCLouvain/ELI/ELIE) Alice Alonso (UCLouvain/ELI/ELIE)
Lecteurs :	Sandra Soares Frazao (UCLouvain/IMMC/GCE) Pierre Defourny (UCLouvain/ELI/ELIE)
Année Académique :	2019 - 2020
Bioingénieur :	Sciences et technologies de l'environnement

Remerciements

La réalisation de ce mémoire a été rendue possible grâce à plusieurs personnes que je voudrais remercier.

Je souhaite avant tout remercier mon promoteur de mémoire, Marnik Vanclooster, pour ses conseils avisés et pour le suivi dont j'avais besoin de la part d'un , ainsi que ma co-promotrice, Alice Alonso, à qui je dois une grande partie de ma motivation tout au long de cette année. Merci beaucoup à tous les deux, pour le temps qu'ils ont consacré à m'apporter les outils méthodologiques indispensables à la conduite de cette recherche, pour leur disponibilité sans égale et pour leurs encouragements, ainsi que pour la supervision éclairée tout au long de la rédaction de ce mémoire.

Je tiens également à dire merci aux personnes du SPW et de l'IRM pour l'accès aux données et leurs réponses à mes questions.

Un grand merci à mes amis, pour leur soutien et leurs encouragements lors de la réalisation de ce mémoire. Avec une mention spéciale à ma cokotteuse, Marie Pontoise, pour la relecture de ce mémoire et pour sa compagnie tout au long de cette dernière année confinée. Merci également aux autres, Bernard, Mathou DV et Fofie, pour leur soutien et leurs encouragements, pour leurs bons conseils et leur intérêt pour mon sujet, ainsi que pour les moments de détente.

Enfin, je voudrais remercier mes parents pour m'avoir soutenu dans mes démarches et mes choix, non pas seulement durant ce mémoire mais tout au long de mes études.

Merci à tous !

Table des matières

1	Introduction	2
2	État de l'art	4
2.1	Surveillance de la qualité de l'eau grâce à la télédétection	4
2.1.1	Principes de la télédétection de l'eau	4
2.1.2	Capteurs satellitaires utilisés pour la télédétection de l'eau	5
2.1.3	Capteurs optiques adaptés pour la caractérisation de la qualité des eaux intérieures	5
2.1.4	Sentinel-2	7
2.2	Techniques d'identification de la turbidité et signature spectrale	8
2.2.1	Turbidité	8
2.2.2	Signature spectrale de la turbidité	9
2.2.3	Méthodes d'extraction de la turbidité	9
2.3	Limitations	12
2.3.1	Rapport signal/bruit et résolution radiométrique	12
2.3.2	Complexité et diversité morphologique des eaux terrestres	13
2.3.3	Reflets du soleil	14
2.3.4	Effet de contiguïté	14
2.3.5	Corrections atmosphériques	15
3	Objectifs	17
4	Méthodologie	18
4.1	Approche générale	18
4.2	Zone d'étude	19
4.2.1	Situation géographique	19
4.2.2	Localisation et conditions environnementales des stations d'observation	21
4.3	Construction des bases de données	24
4.3.1	Collecte et prétraitement des données de terrain	24
4.3.2	Collecte et prétraitement des données satellites	26
4.4	Méthodes de modélisation	31
4.4.1	Définition des scénarios de modélisation	31
4.4.2	Analyse exploratoire et prétraitement des variables prédictives	32
4.4.3	Analyse en composantes principales : identification des variables pré- dictives	33
4.4.4	Développement des modèles prédictifs	35
5	Résultats	36
5.1	Descriptions des données	36
5.1.1	Comparaison des données de turbidité horaires et journalières	36

TABLE DES MATIÈRES

5.1.2	Résumé des données retenues pour analyse	37
5.2	Analyse univariée	38
5.2.1	Descriptions statistiques	38
5.2.2	Distribution de la turbidité sur la gamme de valeur	39
5.2.3	Distribution dans le temps	41
5.2.4	Normalisation	41
5.3	Analyse en composantes principales	42
5.3.1	Bandes spectrales sélectionnées	42
5.3.2	Construction des nouvelles variables prédictives	44
5.4	Évaluation des modèles prédictifs	47
6	Discussion	49
7	Conclusion	55
	Annexes	62
	Annexe A Bandes spectrales des satellites Sentinel-2	62
	Annexe B Script GEE pour l'extraction des réflectances de surface	63
	Annexe C Graphiques des transformations logarithmiques	66
	Annexe D Régressions linéaires associées aux modèles prédictifs	73

Liste des figures

1	Interaction des différents acteurs de la télédétection	4
2	Caractéristiques des bandes spectrales Sentinel-2	8
3	Schéma de la méthodologie suivie.	19
4	Bassin hydrographique de la Meuse en Belgique	20
5	Carte d'occupation du sol de la zone d'étude.	21
6	Localisation des stations de mesure.	22
7	Photo de la localisation de la station d'Hastière.	22
8	Photo de la localisation de la station d'Andenne.	23
9	Photo de la localisation de la station de Lixhe.	23
10	Fréquence de revisite de Sentinel-2	27
11	Illustrations des masques de nuages appliqués aux images	29
12	Zones de sélection de la fenêtre de 3x3 pixels	30
13	Comparaison des données de turbidité pour la station d'Hastière.	36
14	Comparaison des données de turbidité pour la station d'Andenne.	36
15	Comparaison des données de turbidité pour la station de Lixhe.	37
16	Histogrammes des données de turbidité.	40
17	Répartitions dans le temps des données des variables d'étalonnage des modèles.	41
18	Cercle de corrélation des bandes spectrales pour la station d'Hastière.	42
19	Cercle de corrélation des bandes spectrales pour la station d'Andenne.	43
20	Cercle de corrélation des bandes spectrales pour la station de Lixhe.	43
21	Cercle de corrélation des bandes spectrales pour les stations groupées.	44
22	Graphiques de la turbidité en fonction de la bande B4.	66
23	Graphiques du logarithme de la turbidité en fonction de la bande B4.	67
24	Graphiques de la turbidité en fonction du débit.	68
25	Graphiques du logarithme de la turbidité en fonction du débit.	69
26	Graphiques du logarithme de la turbidité en fonction du logarithme du débit.	70
27	Graphiques de la turbidité en fonction des précipitations.	71
28	Graphiques du logarithme de la turbidité en fonction des précipitations.	72
29	Régressions linéaires du scénario 1	73
30	Régressions linéaires du scénario 2	74
31	Régressions linéaires du scénario 3	75
32	Régressions linéaires du scénario 4	76
33	Régressions linéaires du scénario 5	77
34	Régressions linéaires du scénario 6	78

Liste des tableaux

1	Caractéristiques des capteurs de télédétection utilisés pour l'observation de la turbidité.	7
2	Sélection de méthodes empiriques appliquées pour l'extraction de la turbidité. Les équations contenues dans ce tableau présentent les équations telles que définies dans leurs publications originales.	11
3	Nombre d'images considérées dans l'analyse.	31
4	Descriptions statistiques des données de turbidité mesurées à 10 heures et les moyennes journalières.	37
5	Caractéristiques des jeux de données.	37
6	Descriptions statistiques des données pour la station d'Hastière.	38
7	Descriptions statistiques des données pour la station d'Andenne.	38
8	Descriptions statistiques des données pour la station de Lixhe.	39
9	Descriptions statistiques des données pour les trois stations groupées.	39
10	Facteur des composantes principales retenues pour le scénario 1.	45
11	Facteur des composantes principales retenues pour le scénario 5.	45
12	Facteur des composantes principales retenues pour le scénario 6.	46
13	Résultats des modèles de régression linéaire.	47
14	Détails des bandes spectrales des satellites Sentinel-2.	62

Liste des acronymes

ACP Analyse en Composantes Principales

BIC Critère d'Information Bayésien

CDOM Colored Dissolved Organic Mater

ESA European Space Agency

ETM+ Enhanced Thematic Mapper

GEE Google Earth Engine

IRM Institut Royal Météorologique

ISO Organisation internationale de normalisation

LaSRC Landsat Surface Reflectance Code

LiDAR Light Detection And Ranging

MAIN Modified Atmospheric correction for INland waters

MERIS MEdium Resolution Imaging Spectrometer

MODIS MOderate Resolution Imaging Spectroradiometer

MSI MultiSpectral Istrument

NASA National Aeronautics and Space Administration

NTU Nephelometric Turbidity Units

OLCI Ocean and Land Color Instrument

OLI Operational Land Imager

QA Quality Assessment

RMSE Root Mean Square Error

Rrs Remote Sensing Reflectance

S2 Sentinel-2

SAR Synthetic-Aperture Radar

SeaWiFS Sea-Viewing Wide Field-of-View Sensor

SNR Signal to Noise Ratio

SPW Service Public de Wallonie

SWIR Short-Wavelength InfraRed

TIRS Thermal InfraRed Sensor

LISTE DES ACRONYMES

UNEP United Nations Environment Programme

UNESCO United Nations Educational, Scientific and Cultural Organization

USGS United States Geological Survey

WLR World's Large Rivers

WLRI World's Large Rivers Initiative

1 Introduction

L'humanité profite quotidiennement des nombreux bénéfices et services que procurent les rivières à travers le monde. Leurs écosystèmes aquatiques ont des fonctions importantes pour l'environnement. Cependant, celles-ci sont menacées suite aux pressions environnementales qu'elles subissent, incluant la surexploitation, la pollution, les activités humaines mais aussi le changement climatique. Les grandes rivières du monde (« World's Large Rivers », WLRs) sont particulièrement exposées aux impacts de ces menaces et il devient donc urgent de développer des nouveaux outils de surveillance de la qualité de l'eau de ces rivières.

Aussi, l'UNESCO a mis en place une collaboration internationale et multidisciplinaire dans le but de recueillir, de partager et d'échanger des données sur la qualité de l'eau à grande échelle en vue d'obtenir une estimation globale de la situation actuelle des WLRs et de leur réponse aux futures changements climatiques ainsi qu'aux impacts anthropiques.

Cette chaire UNESCO fut donc créée en 2014 et est établie à l'université des ressources naturelles et des sciences de la vie à Vienne en Autriche. Elle permet le soutien de l'initiative sur les WLRs (« World's Large Rivers Initiative », WLRI) et le maintien des liens entre universités et instituts de recherche et la mise en commun des ressources de manière à créer une gestion efficiente, globale et durable des grandes rivières et de leurs bassins.

Ce mémoire s'inscrit donc dans le cadre de la WLRI, dont l'UCLouvain fait partie, et a pour but d'évaluer les potentialités d'une approche hybride de la surveillance de la qualité de l'eau à grande échelle.

À ce jour, plusieurs méthodes permettant de caractériser l'état de la qualité de l'eau existent que ce soit à petite échelle grâce à l'analyse physico-chimique de l'eau directement par échantillonnage ou à une échelle beaucoup plus grande grâce à la télédétection. Considérant que la qualité de l'eau est un processus dynamique à la fois dans le temps et l'espace, il est difficile de bien caractériser cette dynamique par les méthodes classiques de « grab sampling » et d'obtenir des informations robustes de la qualité de l'eau aux échelles d'extension spatiale et temporelle compatibles aux échelles opérationnelles de la gestion. C'est pourquoi depuis quelques décennies, les scientifiques se penchent, entre autres, sur le potentiel de la télédétection pour obtenir des informations complémentaires par rapport à la qualité de l'eau aux échelles spatiales et temporelles de la gestion (i.e. le bassin versant, la masse d'eau, la région, ...).

Dans ce contexte, l'objectif général de ce travail est de contribuer aux développements méthodologiques d'analyse de la qualité de l'eau à grande échelle sur base d'images satellites. Il est suggéré de réaliser ce progrès méthodologique par une meilleure interprétation et modélisation des signaux télédétectés en fonction de la qualité de l'eau. L'intégration de la télédétection peut conduire à une meilleure capacité de surveillance, et donc aussi de la gestion de la qualité des eaux. Pour concrétiser cet objectif général, ce mémoire se concentre sur la surveillance de la turbidité de l'eau des rivières. En effet, la turbidité affecte la qualité de l'eau en blo-

quant la transmission de la lumière, en altérant la concentration en nutriments et réduisant la concentration en oxygène. De façon plus poussée, la turbidité peut être utilisée pour le contrôle indirect de la présence de métaux lourds dans les masses d'eaux.

Afin de mener cela à bien, ce mémoire est scindé en cinq sections majeures. La première section est une étude bibliographique qui résume les avancées de la surveillance de la qualité de l'eau, et plus particulièrement la turbidité, grâce à la télédétection. Les concepts, les défis et les limitations de ce domaine sont passés en revue. La deuxième section correspond à la méthodologie et documente de manière approfondie les bases de données provenant de la Meuse et de Sentinel-2. Les analyses corrélatoires y sont également décrites ainsi que les outils utilisés. Les résultats sont repris dans la section trois. Dans la quatrième section, les résultats sont évalués par rapport aux autres résultats de la littérature et discutés en termes de qualité et performance. Finalement, ce mémoire se clôture avec une description des perspectives d'avenir et des conclusions.

2 État de l'art

2.1 Surveillance de la qualité de l'eau grâce à la télédétection

2.1.1 Principes de la télédétection de l'eau

Le principe de base de la télédétection de l'eau repose sur la mesure du rayonnement électromagnétique émis ou réfléchi depuis une masse d'eau, permettant ainsi d'en analyser les caractéristiques spectrales. Le rayonnement, après avoir traversé l'atmosphère où il est modifié une première fois par absorption et diffusion, est réfléchi ou passe la surface de l'eau pour se propager à travers la masse d'eau. Certains des constituants se trouvant dans l'eau, comme les microorganismes, les particules de matière en suspension ou la matière organique dissoute, interagissent avec le rayonnement et peuvent absorber une partie de ce dernier ou partiellement le réfléchir. Les rayonnements réfléchis par ces constituants peuvent alors ressortir de l'eau et être mesurés par les capteurs qui enregistrent leur intensité et leur longueur d'onde. Les mesures se traduisent en réflectance, c'est-à-dire le rapport entre le signal reçu par une surface et le signal renvoyé, et peut se mesurer à différentes longueurs d'onde du spectre électromagnétique. Cela permet de renvoyer une information sur l'état de l'eau et de ses constituants. Le fonctionnement de la télédétection est illustré à la figure 1 ci-dessous avec comme exemple de rayonnement la lumière incidente du soleil. Signalons que dans ce mémoire nous nous limitons à la télédétection électromagnétique. En effet, aucune application de l'observation de la qualité de l'eau par méthode de la télédétection non électromagnétique n'a été trouvée.

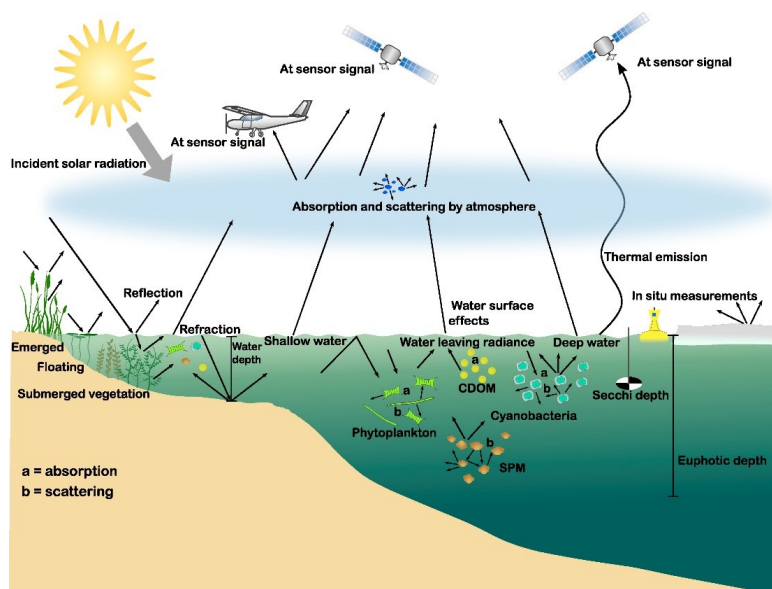


FIGURE 1 – Interaction entre le rayonnement, les indicateurs de télédétection de l'écologie des lacs et les capteurs satellites (DÖRNHÖFER et al., 2016)

2.1.2 Capteurs satellitaires utilisés pour la télédétection de l'eau

La télédétection permet donc d'acquérir des informations sur les propriétés de l'eau en utilisant le spectre électromagnétique. Il en existe deux sortes selon le type de capteur à bord du satellite : la télédétection passive et la télédétection active.

La majorité des capteurs satellitaires utilisés pour la surveillance de la qualité de l'eau sont des capteurs dit 'passifs', également appelés capteurs optiques, tels qu'une caméra ou un radiomètre. Ceux-ci utilisent l'énergie naturellement disponible, c'est-à-dire qu'ils détectent le rayonnement émis par réflexion des rayons du soleil.

Les capteurs dits 'actifs', tels que la détection et la télémétrie de la lumière ("Light Detection And Ranging", LiDAR), le radar à synthèse d'ouverture ("Synthetic-aperture radar", SAR), l'altimètre ou encore le sonar, envoient eux-mêmes leur propre source de rayonnement pour balayer un objet et utiliser l'énergie réfléchie ou la rétrodiffusion de l'objet pour obtenir la mesure souhaitée. Ils ne sont pas souvent exploités lors de l'étude de la qualité de l'eau malgré qu'ils soient connus pour être en mesure de fournir des informations sur les eaux intérieures, concernant notamment la couverture de glace, la bathymétrie et la quantité d'eau, ainsi que les mesures de la qualité de l'eau par fluorescence (GHOLIZADEH et al., 2016 ; PALMER et al., 2015).

Quatre grands types de résolutions définissent les capteurs satellitaires utilisés pour la télédétection : la résolution spatiale, la résolution spectrale, la résolution radiométrique et la résolution temporelle.

- Résolution spatiale : L'intervalle de distance séparant deux points contigus pour qu'ils soient correctement distincts. Pour un capteur satellite, elle peut être définie comme l'espacement entre les échantillons de pixels à la surface de la terre. Elle s'exprime généralement en mètres.
- Résolution temporelle : L'intervalle de temps entre lequel un capteur peut visualiser la même zone au sol. Autrement dit, c'est la fréquence de revisite du satellite à un endroit particulier. Elle s'exprime en jours.
- Résolution spectrale : L'intervalle entre les différentes longueurs d'onde qu'il est possible de capter. Plus elle est fine, plus la plage de longueurs d'onde est étroite pour une bande particulière. Elle s'exprime en nanomètres.
- Résolution radiométrique : L'intervalle d'intensité lumineuse ou de réflectance qu'un instrument peut capter. Plus la résolution est élevée, plus il sera possible de détecter les différences d'intensité ou de réflectance. Elle est exprimée en 'bit'.

2.1.3 Capteurs optiques adaptés pour la caractérisation de la qualité des eaux intérieures

Jusqu'à présent, la plupart des études réalisées portant sur l'observation de la qualité de l'eau grâce à la télédétection se concentrent sur les océans et zones côtières plutôt que sur les rivières et lacs (UNEP, 2016). La raison principale découle de l'importante complexité

du signal optique des eaux douces et des relativement petites dimensions des plans d'eau intérieurs et des réseaux fluviaux. En effet, les rivières ont souvent un niveau de turbidité bien plus élevé que les environnements marins ou côtiers, et les particules en suspension interfèrent spectralement avec la mesure d'autres paramètres.

Ces études exploitent le potentiel de capteurs spatiaux, tels que SeaWiFs, MERIS et MODIS, dont les caractéristiques techniques, notamment une large gamme spectrale d'énergie électromagnétique, conviennent à l'observation des masses d'eau océaniques. Plus récemment, les satellites Sentinel-3, lancés en 2016 et en 2018, ont à bord le capteur OLCI, possédant des caractéristiques spatiales, spectrales et radiométriques idéales et permettant la surveillance de systèmes aquatiques terrestres. Héritage de MERIS, les améliorations comprennent 21 bandes spectrales, un meilleur rapport signal/bruit, des caméras inclinées pour atténuer les reflets du soleil ainsi qu'une résolution spatiale globale de 300 m à pleine résolution (KRAVITZ et al., 2020). Malgré cela, leurs résolutions spatiales relativement grossières (≥ 300 m) convient moins aux applications sur des petits points d'eau ou des rivières plus étroites (DÖRNHÖFER et al., 2016).

Le lancement ultérieur de capteurs dotés de résolutions spatiale, spectrale et radiométrique suffisamment élevées a permis d'améliorer l'obtention des données sur de plus petits points d'eau. Parmi ceux-ci, on retrouve les satellites Sentinel-2, Worldview-2 et 3 et Landsat-8, toujours en fonction aujourd'hui. Malgré qu'ils soient avant tout destinés à des applications terrestres, ces capteurs montrent des hautes résolutions a priori adaptées à des applications pour les eaux intérieures et la récupération d'informations sur les constituants présent dans ces eaux (DÖRNHÖFER et al., 2016 ; UNEP, 2016).

Par ailleurs, Dörnhöfer and Oppelt (2016) suggèrent l'utilisation synergique de capteurs avec différentes résolutions temporelles, spatiales et spectrales telles que Landsat-8, Sentinel-2 et Sentinel-3 afin de surmonter les limites des systèmes uniques.

Le tableau 1 reprend les caractéristiques principales des différents capteurs cités plus haut.

TABEAU 1 – *Caractéristiques des capteurs de télédétection utilisés pour l'observation de la turbidité.*

Capteur	Satellite	Résolution temporelle (jours)	Résolution spatiale (m)	Nombre de bandes spectrales et leur gamme de longueur d'onde (nm)
MSI	Sentinel-2	5	10–20–60	13 (432–2290)
OLCI	Sentinel-3	1 à 2	300–1200	21 (400–1020)
ETM +	Landsat-7	16	30–15–60	5 (441–1750) - 1 Pan (520–900) 1 (2090–2350) - 1 (1040–1250)
OLI, TIRS	Landsat-8	16	30–15–100	5 (430–880) - 1 Pan (500–680) 2 (1570–2290) - 1 Cirrus (1360–1380) 2 TIRS (10 600–12 510)
SeaWiFS	SeaStar*	1	1100–4500	8 (402–885)
MODIS	Terra et Aqua	1 à 2	250–500–1000	2 (620–876) - 5 (459–2155) 29 (405–877 et thermique)
MERIS	Envisat*	1	300–1200	15 (390–1040)
WorldView-2	WorldView-2	1	1,85–0,46	8 (400–1040) - 1 Pan (450–800)
WorldView-3	WorldView-3	1 à 4	1,24–3,7–0,31	8 (400–1040) - 1 Pan (450–800) 8 SWIR (1195–2365)

* Ces satellites ne sont plus en fonction aujourd'hui.

2.1.4 Sentinel-2

Bien que non conçu pour les applications aquatiques, le satellite de l'Agence Spatiale Européenne (ESA) Sentinel-2 (S2) avec à son bord l'Instrument MultiSpectral (MSI) apporte une amélioration par rapport aux précédentes générations de capteurs utilisés pour la surveillance des constituants des eaux de surfaces. Lancés en 2016 pour Sentinel-2A et en 2018 pour Sentinel-2B, leurs performances ont été démontrées à plusieurs reprises lors d'études menées sur les lacs, les eaux côtières et certaines rivières. MSI offre des images à haute résolution spatiale (10 – 60 m) avec une fréquence de revisite de cinq jours à l'équateur et est muni de 13 bandes spectrales allant du visible au moyen infrarouge. Les bandes spectrales sont représentées dans la figure 2 avec leur résolution spatiale respective.

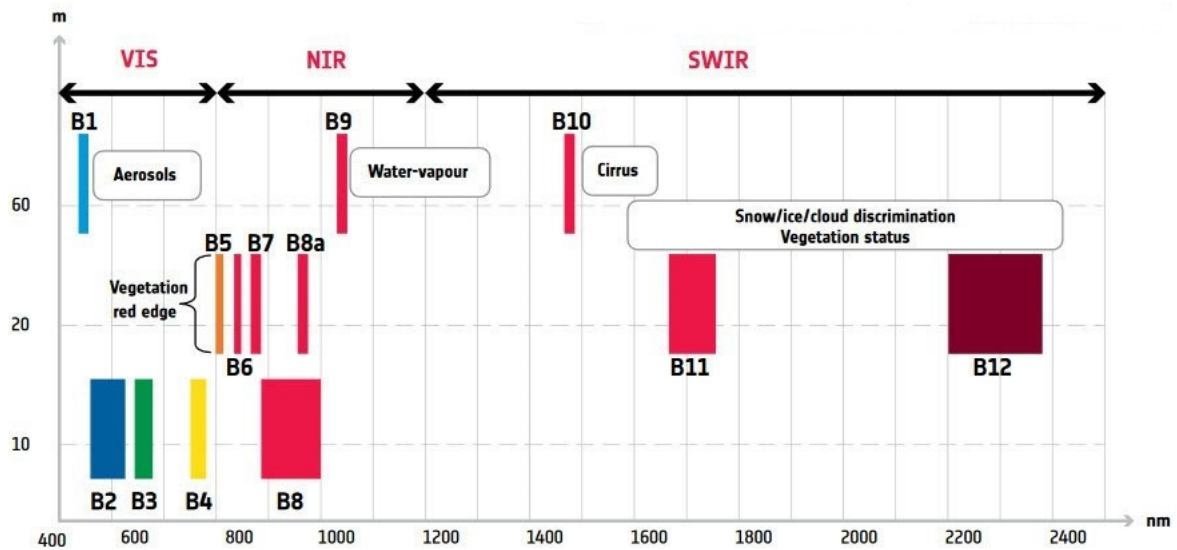


FIGURE 2 – Caractéristiques des bandes spectrales Sentinel-2 avec la résolution spatiale en ordonnée [m] et la longueur d'onde en abscisse [nm], VIS = visible, NIR = proche infrarouge et SWIR = infrarouge à courtes longueurs d'onde.

Source : <https://esamultimedia.esa.int/>

De plus, l'accès aux produits satellites S2 déjà traités et corrigés a été facilité grâce aux images et environnements de travail open-source permettant le traitement de ces données. Actuellement, les plateformes de 'cloud computing' Google Earth Engine et Amazon Web Services donnent accès aux données des satellites de la NASA, de l'USGS et de l'ESA et fournissent des ressources informatiques et outils puissants haut de gamme à un coût nul.

2.2 Techniques d'identification de la turbidité et signature spectrale

2.2.1 Turbidité

La qualité des eaux de surface est définie et affectée par les diverses substances qu'elle contient, leur quantité et leur effet sur les écosystèmes aquatiques et la santé humaine. Ces substances peuvent être, soit d'origine naturelle comme les minéraux et microorganismes, soit découler de la présence humaine (eaux usées) ou des activités industrielles et agricoles (substances toxiques, métaux, pesticides). C'est la concentration de ces différents éléments qui détermine la qualité d'une eau et influence donc ses propriétés physiques, chimiques ou biologiques.

La turbidité en fait partie et affecte ainsi les conditions de luminosité de ces eaux de surface par absorption et diffusion de la lumière incidente, ce qui a pour conséquence d'altérer l'état qualitatif de l'eau. La définition utilisée ici se base sur celle de l'Organisation Internationale de Normalisation ISO 7027 : la turbidité se caractérise par la réduction de la transparence de l'eau due à la présence de matières non dissoutes et/ou colloïdales et de petits organismes comme par exemple des bactéries, du phytoplancton ou du zooplancton. La transparence, ou

encore la visibilité dans l'eau, est indiquée par la perméabilité aux ondes électromagnétiques, de la lumière dans le cas présent (ORGANISATION INTERNATIONALE DE NORMALISATION, 2019). La turbidité est donc une propriété optique basée sur la quantité de lumière diffusée et absorbée par les particules dans l'eau. Plus il y a de particules, plus la lumière sera diffusée et par conséquent, plus la turbidité sera élevée. Les particules en suspension, telles que les sédiments, sont responsables de la majeure partie de la diffusion, tandis que l'absorption est contrôlée par la chlorophylle-a et les matières dissoutes ou particulaires colorées.

La turbidité se mesure en unités de turbidité néphélométrique ("Nephelometric Turbidity Units", NTU) et varie généralement dans une gamme allant de 0,05 à 400 NTU (ORGANISATION INTERNATIONALE DE NORMALISATION, 2016).

2.2.2 Signature spectrale de la turbidité

En général, une bonne corrélation entre la turbidité et la réflectance est trouvée dans les bandes satellites situées dans la partie rouge du spectre pour les valeurs de turbidité faibles à modérées voire des eaux très turbides (DOGLIOTTI et al., 2015).

L'utilisation d'une seule bande, choisie de manière appropriée, dans les techniques de télé-détection de la turbidité et des concentrations de particules en suspensions a montré des résultats satisfaisants (GHOLIZADEH et al., 2016). Cependant, l'eau contient des substances plus complexes ayant la capacité de modifier la réflectance du plan d'eau et provoquant alors des variations de couleur. Par conséquent, l'utilisation de différentes bandes spectrales pourrait être pertinent pour obtenir des résultats plus précis dans différentes concentrations de plans d'eau.

2.2.3 Méthodes d'extraction de la turbidité

De nombreuses approches ont été développées pour la récupération de la turbidité dans les eaux intérieures sur la base des interactions optiques des constituants de l'eau et du rayonnement électromagnétique dans les régions du visible et du proche infrarouge qui ne sont pas fortement touchées par l'absorption de l'eau. Ces algorithmes sont généralement divisés en deux grandes catégories : les méthodes empiriques et les méthodes analytiques/bio-optiques (SAGAN et al., 2020).

L'objectif des méthodes empiriques est d'identifier les relations entre les paramètres de qualité de l'eau et les valeurs de réflectance spectrale (bande individuelle ou combinaison de bandes) au moyen d'efforts de régression. Les approches empiriques sont entièrement ou semi-dirigées par les données et nécessitent des mesures in situ suffisantes de la qualité de l'eau qui sont utilisées pour établir une relation empirique entre la radiance de sortie de l'eau mesurée par le capteur (DÖRHNHÖFER et al., 2016) et les mesures de la qualité de l'eau. La plupart des approches empiriques utilisent la technique de modélisation de régression multivariée en raison de la complexité optique des eaux intérieures.

Les approches linéaires telles que la régression linéaire multiple, les moindres carrés partiels, la modélisation de mélange linéaire et les algorithmes génétiques sont parmi les méthodes les plus employées dans des études antérieures et ont donné des résultats prometteurs pour la plupart d'entre elles. Néanmoins, la présence de relations non linéaires entre les constituants de la qualité de l'eau a conduit les scientifiques à utiliser des approches plus avancées. Dans la littérature, les méthodes non linéaires telles que les approches basées sur la régression vectorielle de soutien et les réseaux neuronaux ont gagné en popularité en raison de leur capacité robuste à saisir des tendances statistiques complexes typiques des données de télédétection de la qualité de l'eau (SAGAN et al., 2020).

Une des limites des méthodes empiriques est la transférabilité géographique. Un modèle empirique développé pour une masse d'eau spécifique peut ne pas être entièrement transférable à d'autres masses d'eau car il est basé sur les caractéristiques uniques de la zone d'étude originale. Toutefois, comme ces techniques intègrent les caractéristiques uniques de la masse d'eau lors de l'établissement d'un modèle, elles génèrent généralement des résultats plus précis que les modèles spectraux et/ou bio-optiques plus généralisés. Un avantage unique de l'approche empirique est qu'il n'est pas nécessaire de comprendre au préalable les interactions souvent complexes entre l'eau et la lumière, ce qui rend l'approche empirique souhaitable pour faciliter la mise en œuvre (SAGAN et al., 2020).

Le tableau 2 synthétise quelques études ayant montré des résultats satisfaisants pour l'extraction de la turbidité grâce à des méthodes empiriques.

Plus récemment, et en raison de la grande complexité des eaux intérieures et côtières, les chercheurs ont donc développé un deuxième type de méthode d'extraction, la technique de modélisation bio-optique, ou analytique. Cette technique est principalement basée sur le transfert radiatif dans la colonne d'eau. En utilisant l'équation de transfert radiatif, les constituants optiquement actifs de la colonne d'eau peuvent être dérivés de l'eau quittant le rayonnement. Cependant, ces modèles nécessitent des informations spectrales détaillées des constituants optiquement actifs de l'eau dans la région cible. Sur la base des interactions fondamentales entre la lumière et l'eau, ces méthodes ont été proposées pour atténuer et surmonter les problèmes de transférabilité régionale hérités des méthodes empiriques (SAGAN et al., 2020).

TABLEAU 2 – Sélection de méthodes empiriques appliquées pour l'extraction de la turbidité. Les équations contenues dans ce tableau présentent les équations telles que définies dans leurs publications originales.

Capteur	Turbidité (NTU)	Type de masse d'eau et lieu	Modèle développé	Equations, bandes utilisées	Référence
TM ETM +	1,5 à 8	Rivière Guadalquivir, Espagne	Step-wise régression Modèles additifs généralisés	$\text{Log}(\text{NTU} + 0,01) = 1,195 + 14,45 \cdot B3$	(BUSTAMANTE et al., 2009)
(IRS-1A)	15 à 45	Réservoir Tawa, Inde	Corrélation entre les bandes et la turbidité	Bande 1, 2 et 3	(CHOUBEY, 1992)
SPOT-HRV	3 à 15	Réservoir Kansas, USA	Modèle de régression linéaire	$\text{Turbidité} = 1,713 \cdot B1 - 31,52$ $\text{Turbidité} = 2,463 \cdot B2 - 83,06$	(GOODIN et al., 1996)
MODIS	0,9 à 8	Baie de Tampa en Floride, USA	Algorithme semi-empirique à bande unique	$a = 1203,9$; $b = 1,087$ $\text{NT} = a \cdot \text{Rrs}(645)b$	(CHEN et al., 2019)
MODIS	3 à 19	Eaux côtières en Sicile, Italie	Algorithme semi-empirique à bande unique	$\text{NT} = a \cdot \text{Rrs}(645)b$ $a = 1203,9$; $b = 1,112, 1,399, 1,290$	(MALTESE et al., 2013)
MODIS	0,5 à 70	Baie de Biscay, France	Modèles de régressions linéaire et polynomiale	Bandes B1 et B2	(PETUS et al., 2010)
MERIS	0,2 à 25	Lagon en Nouvelle-Calédonie, baies à Fidji, lagon à Cuba	Modèle polynomial et modèle de puissance	R681 $R421/R620$; $R443/R670$; $R510/R681$ $R620 \cdot R861/R412$; $R620 \cdot R681/R510$	(OUILLOIN et al., 2008)
MERIS	1 à 60	Réservoir Alqueva, Portugal	Régression linéaire	$\text{Turbidité} = - 6,39 + 8,93 (B5/B1)$	(POTES et al., 2012)
ETM +	2 à 997	Lac en Alaska	Simple et multiple régressions linéaires	$\text{Log}(T) = -0,329 + 0,0513 \cdot B3$; $\text{Log}(T) = - 0,217 + 0,115 \cdot B4$; $\text{Log}(T) = -0,441 + 0,0340 \cdot B3 + 0,0507 \cdot B4$	(LIVERSEDGE, 2007)
ETM +	75 à 275	Lacs, Nouvelle-Zélande	Régression linéaire	$\text{Turbidité} = - 63.717 + 1,587.8 \cdot B4$	(HICKS et al., 2013)
Hyperspectral	3 à 40	Rivière Mississippi au Minnesota, USA	Step-wise régression	$\text{LN}(\text{turbidité}) = - 0.004 \cdot (705) + 4.88$	(OLMANSON et al., 2013)

2.3 Limitations

Les valeurs typiques de réflectance sortant de l'eau sont généralement de l'ordre de 2 à 10% tandis que les effets atmosphériques contribuent pour 80 à 90% du signal reçu par le capteur en haut de l'atmosphère. Ceux-ci sont principalement dû à la diffusion moléculaire et des aérosols dans l'atmosphère, qui diminue avec l'augmentation de la longueur d'onde. Les pourcents restants résultent des effets supplémentaires à la surface de l'eau, tels que les reflets solaires, la proximité des terres voisines, les effets du fond de l'eau, et doivent également être pris en compte. Pour cette raison, une élimination précise des effets dus à l'atmosphère et à la surface de l'eau est essentielle, les corrections atmosphériques représentent donc un réel défi. C'est également pourquoi il est important d'atteindre un rapport signal/bruit suffisant pour déduire les paramètres optiques et biophysiques du plan d'eau observé à partir du capteur.

Plusieurs de ces limitations pouvant être rencontrées dans le cas de ce travail sont décrites ci-après ainsi que les éventuelles solutions développées dans de précédentes études.

2.3.1 Rapport signal/bruit et résolution radiométrique

Le rapport signal/bruit ("Signal to Noise Ratio", SNR) est un indicateur de la qualité de la transmission d'une information. Il est quantifié comme une mesure du niveau du signal souhaité par rapport au bruit de fond qui vient interférer. Le niveau de bruit d'un capteur est inhérent à l'instrument et a une grande influence sur les estimations des constituants dans les eaux optiquement complexes. En effet, la correction atmosphérique précise et la récupération de la réflectance de surface nécessitent un SNR suffisant pour détecter et différencier le signal sortant de l'eau.

Les eaux côtières et intérieures turbides présentent des charges de sédiments élevées, contrairement aux océans. Elles contribuent donc à une réflectance sortant de l'eau plus prononcées dans la gamme spectrale verte et rouge. Dans ce cas, un SNR faible conduira souvent à une erreur relative élevée dans la réflectance à 443 nm (aérosols) à cause de la forte absorption par la chlorophylle-a et les matières organiques dissoutes colorées ('Colored Dissolved Organic Mater', CDOM). Selon l'approche de correction atmosphérique, un SNR faible dans les bandes du proche IR ou dans l'IR à courtes longueurs d'onde peut contribuer au bruit et à la polarisation systématique dans la gamme spectrale visible, en raison d'une erreur d'estimation de la contribution des aérosols au signal observé (FRANZ et al., 2015).

Il y a souvent un compromis entre le SNR et la résolution spatiale, typiquement lors de l'utilisation de la nouvelle génération de capteurs offrant des résolutions spatiales adaptées pour les corps d'eau à l'intérieur des terres, tels que OLI et MSI, respectivement à bord de Landsat-8 et Sentinel-2 (LI et al., 2015). Par exemple, MODIS possède un SNR variant de 516 à 1087 pour ses bandes utilisées pour la couleur de l'océan (405 à 877 nm) et le SNR de OLCI, à bord de Sentinel-3, varie de 232 à 2188 pour la même gamme de longueur d'onde, tandis que le SNR de Sentinel-2 ne dépasse pas 174 toutes bandes confondues.

De plus, étant donné que la taille des pixels du MSI est petite, la bande passante spectrale est large afin de recevoir suffisamment d'énergie lumineuse pour un SNR raisonnable tandis qu'en télédétection par satellite de la qualité de l'eau, des bandes passantes spectrales plus petites sont plus favorables. C'est pourquoi les capteurs MERIS et OLCI, spécialement conçus pour la surveillance de la qualité de l'eau, ont de nombreuses bandes étroites et des SNR plus élevés en raison de la faible résolution spatiale.

Cependant, plusieurs études ont démontré l'utilité de capteurs dédiées aux applications terrestres pour l'étude qualitative des eaux. Parmi celles-ci, l'article de Franz et al. (2015) démontre le potentiel du capteur 'Operational Land Imager' (OLI) à bord de Landsat-8 pour la radiométrie des couleurs de l'océan via SeaDAS, un logiciel open source pour le traitement, l'analyse et l'affichage des mesures de télédétection océanique. Il en conclut que OLI possède les bandes spectrales requises et les performances radiométriques suffisantes pour prendre en charge l'approche de correction atmosphérique standard utilisée pour les missions mondiales de couleur de l'océan, y compris la détermination et la suppression des contributions d'aérosols basées sur des modèles d'aérosols réalistes et des observations d'OLI dans le NIR et SWIR, et pour permettre la récupération quantitative des constituants de la colonne d'eau à partir des distributions spectrales de réflectance de sortie d'eau spectrale.

L'étude de Pahlevan et al. (2017) confirme les capacités d'OLI et démontre celles des images MSI implémentées elles aussi dans SeaDAS. Les validations in situ et les comparaisons des produits de réflectance de télédétection dérivés de MSI avec ceux dérivés de OLI indiquaient tous deux des produits raisonnables dans les eaux côtières et intérieures. Par rapport à l'OLI, le MSI offre encore plus de mesures spectrales dans la région proche infrarouge. Ces bandes spectrales supplémentaires sont avantageuses par rapport aux eaux côtières/intérieures optiquement complexes et fourniront des opportunités pour une récupération améliorée des paramètres de qualité de l'eau dans les eaux hypertrophiques, riches en CDOM ou extrêmement turbides.

2.3.2 Complexité et diversité morphologique des eaux terrestres

La collecte et l'interprétation des données sur les eaux à l'intérieur des terres est un grand défi en raison de la complexité plus élevée du signal optique provenant de celles-ci. Les rivières présentent souvent des niveaux de turbidité d'un ou plusieurs ordres de grandeur plus élevés que les environnements côtiers ou marins et ont une dynamique temporelle plus élevée (UNEP, 2016). De plus, dans les eaux profondes des océans ou des grands lacs, la surface et masse d'eau et les constituants de l'eau sont les principales sources de rayonnement. Au contraire, dans les eaux peu profondes, une partie du rayonnement incident atteint le fond de la masse d'eau et le rayonnement réfléchi contient alors des informations sur le substrat inférieur et la bathymétrie en plus des composants de l'eau (DÖRNHÖFER et al., 2016).

Par ailleurs, l'aspect des plans d'eau varie considérablement selon leur contenu et leur morphologie. La fluctuation des propriétés optiques inhérentes (c'est-à-dire celles qui dépendent

du milieu, l'eau, mais pas du champ lumineux ambiant dans le milieu) dépend des types de matières organiques et de leur contenu et du non-vivant. Par contre, les propriétés optiques apparentes (celles qui sont fonction du milieu ainsi que de la structure géométrique du champ lumineux ambiant) varient selon la morphologie du plan d'eau. Ces composites aux couleurs vraies montrent une diversité de l'eau, blanche, bleue, verte, noire et cyan (D'ANDRIMONT, 2017). Ces limitations ne devraient pas affecter les résultats du présent travail mais devraient être prises en compte dans une perspective d'analyse de diverses masses d'eau.

2.3.3 Reflets du soleil

Le rayonnement sortant de l'eau peut être difficile à observer en raison de la réflexion de la lumière directe du soleil sur la surface de l'eau dans la direction du champ de vision du satellite. C'est ce qu'on appelle en anglais le phénomène de « sunglint » et apparaît lorsque l'orientation de la surface de l'eau est telle que le soleil est réfléchi de manière spéculaire vers le capteur.

Il est à noter que la géométrie d'observation du satellite Sentinel-2 le rend vulnérable à la contamination par les reflets solaires, le capteur fonctionne à des angles de vision proche du nadir où les reflets solaires sont susceptibles de contribuer au signal reçu au niveau du satellite.

L'algorithme mis au point par HARMEL et al. (2018) sur les images Sentinel-2 utilise la partie infrarouge à courtes longueurs d'ondes du spectre pour corriger l'effet « sunglint » et améliore de 55% la corrélation entre les données satellites et les données in situ.

2.3.4 Effet de contiguïté

Le choix de la localisation pour l'analyse de la réflectance sortant de l'eau a toute son importance. En effet, la sélection du site vise, en outre, à réduire autant que possible les effets de contiguïté, définie ici comme l'éclaircissement des pixels de l'eau sombre en raison des réflexions des pixels terrestres voisins ayant des réflectances plus élevées. Cela a pour effet de réduire le contraste entre les pixels d'eau et de terre. En effet, la diffusion de la lumière parasite de la terre dans le champ de vision des capteurs est connue pour être une source importante d'erreur dans les bandes du proche infrarouge. Ces effets environnementaux entravent la récupération correcte de la qualité de l'eau et des paramètres des aérosols à partir d'images télédéteectées.

Néanmoins, KUHN et al. (2019), montrent que les techniques de correction terrestres, telles que le 'Landsat Surface Reflectance Code' (LaSRC), peuvent déterminer les réflectances de surface sur le fleuve Amazone de manière comparable aux méthodes aquatiques. Cela peut s'expliquer par les effets de contiguïté qui compromettent les approches de correction aquatique basées sur l'infrarouge à courte longueur d'onde ("Short-Wavelength InfraRed", SWIR) et ce, malgré le fait que les techniques terrestres ne corrigent pas les reflets du soleil. DE KEUKELAERE et al. (2018) montrent des résultats de réflectance améliorés en utilisant le cadre

de correction atmosphérique iCOR sur Landsat-8 et Sentinel-2 conçu pour fonctionner sur les eaux intérieures avec une correction de contiguïté explicite.

2.3.5 Corrections atmosphériques

Les effets atmosphériques contribuent pour 70 à 90% du signal reçu par le capteur en haut de l'atmosphère. Ceux-ci sont principalement dus à la diffusion moléculaire et des aérosols dans l'atmosphère, qui diminue avec l'augmentation de la longueur d'onde. Des chercheurs ont donc développé des algorithmes de correction atmosphériques permettant d'éliminer ces effets et n'obtenir que le signal souhaité.

L'étude de KUHN et al. (2019) évalue le potentiel de trois corrections atmosphériques (LaSCR, ACOLITE et SeaDAS) appliquées aux produits de réflectance de surface de niveau 1 de Landsat-8 et Sentinel-2 pour les fleuves Amazon, Columbia et Mississippi. La comparaison des produits satellites après les corrections atmosphériques avec les mesures de réflectance in situ des rivières montre globalement une meilleure correspondance pour les produits issus de LaSCR. Ce dernier, initialement prévu pour les applications terrestres, indique des valeurs de réflectance 36% plus élevées dans les bandes visibles par rapport aux méthodes aquatiques (ACOLITE et SeaDAS). Les différences les plus importantes entre produits terrestres et aquatiques sont observés chez Sentinel-2 avec une divergence de 62%. La différence avec LaSRC était plus grande pour ACOLITE mais pas pour SeaDAS.

PAGE et al. (2019) exploitent le potentiel d'une correction atmosphérique ('Modified Atmospheric correction for INland waters', MAIN) dans l'environnement Google Earth Engine (GEE) en l'appliquant à l'imagerie satellite de Sentinel-2 (S2) et Landsat-8 (L8) pour des eaux intérieures aux Etats-Unis. Les réflectances de S2 et L8 générées par MAIN sont alors comparées à celles générées par ACOLITE ainsi qu'aux produits de réflectance de surface déjà disponibles pour L8, ces dernières ayant été converties avec LaSCR par l'USGS. Dans l'ensemble, les méthodes MAIN et ACOLITE ont donné des valeurs de réflectances comparables aux produits de réflectance de surface de L8 de l'USGS pour S2 et L8 pour les 18 lacs échantillonnés. L'erreur la plus élevée dans les spectres MSI par rapport à OLI était attendue en raison des différences dans le SNR du capteur MSI. En outre, les valeurs de réflectance dans les bandes bleues, vertes et rouges (B2-B4) résultant à la fois de MAIN et d'ACOLITE ont donné l'erreur la plus basse sur des bandes comparables par rapport aux spectres de référence L8-USGS. Ceci est important car ces bandes contribuent le plus à l'estimation de l'écart-type pour les capteurs utilisés dans des recherches antérieures, et ont été considérées comme des variables cibles pour élaborer des modèles de clarté de l'eau.

Dans une autre étude menée par WARREN et al. (2019), six algorithmes de correction atmosphérique (ACOLITE, C2RCC, iCOR, SeaDAS, Polymer et Sen2Cor) ont été testés sur des données Sentinel-2/A provenant d'eaux côtières de la mer Baltique et d'eaux à l'intérieur des terres, ces dernières incluant 13 masses d'eau de cinq pays européens différents. Les produits de réflectance générés par les corrections atmosphériques comparés aux valeurs des mesures

de terrain ont été analysées statistiquement. Pour les eaux intérieures, Polymer et C2RCC ont réalisé les meilleurs résultats avec Polymer qui obtient le coefficient de corrélation le plus élevé (0,72) dans la bande de 560 nm (vert) dans l'ensemble des données intérieures. ACOLITE a des valeurs de coefficient de corrélation élevées dans le rouge et Proche InfraRouge (PIR), C2RCC a généralement les plus faibles erreurs et avec Polymer, ce sont les meilleurs dans la bande à 560 nm. Sen2Cor et iCOR ont obtenu de médiocres résultats pour les eaux côtières mais ont montré des performances améliorées pour les eaux intérieures, suggérant que ces algorithmes conviennent mieux pour ces dernières. Cependant, dans de nombreux cas, les bandes rouges à PIR ne montrent pas de bons résultats, avec des incertitudes beaucoup plus élevées, ce qui est une préoccupation pour la surveillance des eaux côtières et intérieures, car ces bandes sont nécessaires pour déterminer des paramètres clés tels que la concentration de chlorophylle-a et la turbidité. Ces incertitudes élevées dans les bandes PIR sont probablement dues en partie à la faible intensité du signal (absorption élevée) et au SNR inférieur du capteur MSI à ces longueurs d'onde.

Une source d'incertitude est l'effet d'adjacence lié à la distance au rivage. La propagation des mesures de la distance au rivage pour chaque ensemble de données implique que l'on s'attendrait à ce que les eaux intérieures soient plus affectées que les ensembles de données de la mer Baltique et du chenal W. L'erreur à 560 nm pour Polymer et C2RCC augmente d'environ 25 à 55% dans l'ensemble des données des eaux intérieures par rapport aux données des eaux côtières. Une partie de cette augmentation est probablement due à la proximité de la terre. En revanche, l'écart moyen à 560 nm pour iCOR et Sen2Cor a diminué de 10% et 25% respectivement dans l'ensemble des données intérieures par rapport au littoral, probablement en raison d'une meilleure approximation de l'aérosol à partir de la présence de plus de pixels terrestres.

3 Objectifs

L'état de l'art a prouvé qu'il était possible de mettre au point des modèles de prédiction des paramètres de la qualité de l'eau grâce à l'imagerie satellite.

L'objectif général de ce travail est d'évaluer le potentiel de l'imagerie Sentinel-2 pour la surveillance de la turbidité des rivières en Wallonie. Cet objectif se scinde en trois objectifs spécifiques.

Le premier sous-objectif concerne le potentiel des images issues des satellites Sentinel-2A et B. La mission Sentinel-2 présente une haute résolution spatiale, des bandes spectrales multiples, des temps de revisite courts et une politique d'ouverture des données, ce qui en a fait une riche archive de données satellitaires accessible au grand public. La paire de satellites Sentinel-2 présente une opportunité pour l'étude des écosystèmes aquatiques par la surveillance des constituants de la qualité des eaux intérieures. Toutefois, les produits dérivés de Sentinel-2 nécessitent des validations supplémentaires pour pallier les effets atmosphériques et garantir les performances dans les conditions environnementales différentes. Récemment, l'Agence Spatiale Européenne a mis au point un algorithme de correction atmosphérique et de terrain permettant d'obtenir les réflectances de surface directement utilisables sur le cloud. Ce premier objectif spécifique vise donc à examiner le potentiel d'utilisation des images Sentinel-2A et 2B de niveau 2A pour construire un modèle prédictif de la turbidité des rivières.

La turbidité est liée à plusieurs paramètres, que ce soit des paramètres hydrologiques comme le débit et les précipitations, ou des paramètres « optiques » comme leur manière d'interagir avec les ondes de la lumière. Ces paramètres ne sont pas tous facilement accessibles ni mesurables. Par ailleurs, ils ne sont pas tous liés explicitement à la turbidité. Le deuxième sous-objectif tend à déterminer la(les)quelle(s) des variables de précipitations, de débit et de la réflectance dans les différentes bandes spectrales mesurées par le capteur de Sentinel-2 sont utiles pour prédire la variabilité observée dans le signal de turbidité. Pour cela, les variables de prédiction seront implémentées au sein de modèles de régression linéaire. La qualité des modèles de chaque variable sera comparée afin de définir si les données Sentinel-2 améliorent le modèle de façon significative ou pas, et de comprendre quelle variable suffit pour l'estimation de la turbidité. De plus, il est nécessaire d'avoir un jeu de données de résolution temporelle adéquates pour construire un modèle de prédiction robuste, c'est-à-dire avec un nombre approprié de données dont la distribution soit répartie sur l'entièreté de la gamme de valeurs de turbidité et à travers les années et les saisons.

Beaucoup de facteurs comme par exemple l'occupation du sol ou les activités anthropiques influencent la variation de la turbidité au sein d'une même rivière. Le dernier sous-objectif teste la performance relative de plusieurs modèles de prédiction spécifiques à la localisation des stations de mesures de turbidité et celle d'un modèle généralisé à toute la rivière d'étude. Les diagnostics des différents modèles indiqueront la qualité des différentes approches de modélisation.

4 Méthodologie

4.1 Approche générale

La méthodologie générale de ce mémoire s'articule sur la construction étape par étape de modèles de régression linéaire multivariés dans le but de prédire au mieux la turbidité. Pour implémenter la méthodologie, la turbidité de la Meuse a été choisie comme étude de cas.

La méthodologie consiste dans un premier temps à compiler les bases de données de la variable réponse (la turbidité), et des variables prédictives potentielles (la réflectance de surface des images Sentinel-2, le débit dans la rivière d'étude et les précipitations). Le cas d'étude de ce mémoire est le segment wallon de la Meuse pour lequel des données de débit, de précipitations et de turbidité sont disponibles. Pour ce segment, trois stations de mesures sont analysées. Les images satellites Sentinel-2 de niveau 2A, corrigées atmosphériquement avec Sen2Cor, furent traitées sur l'environnement Google Earth Engine dans le but de sélectionner les images fiables et d'en extraire les réflectances de surface.

La sélection des variables prédictives potentielles a été faite sur base d'une Analyse en Composantes Principale (ACP). Après validation et prétraitement des variables prédictives potentielles, celles-ci furent implémentées dans une deuxième ACP afin de réduire la dimension du jeu de données en construisant de nouvelles variables prédictives. Ces dernières variables consistent en des combinaisons linéaires des variables originales, et ayant la propriété d'être indépendantes entre elles.

Dans l'étape finale, les nouvelles variables indépendantes furent employées pour la construction de modèles linéaires de prédiction de la turbidité relatifs aux trois stations de départ et également pour l'ensemble des données généralisé à la Meuse. Les modèles sont alors comparés entre eux en vue de déterminer laquelle des variables explique le mieux la turbidité.

Le schéma de la figure 3 illustre les étapes résumées précédemment.

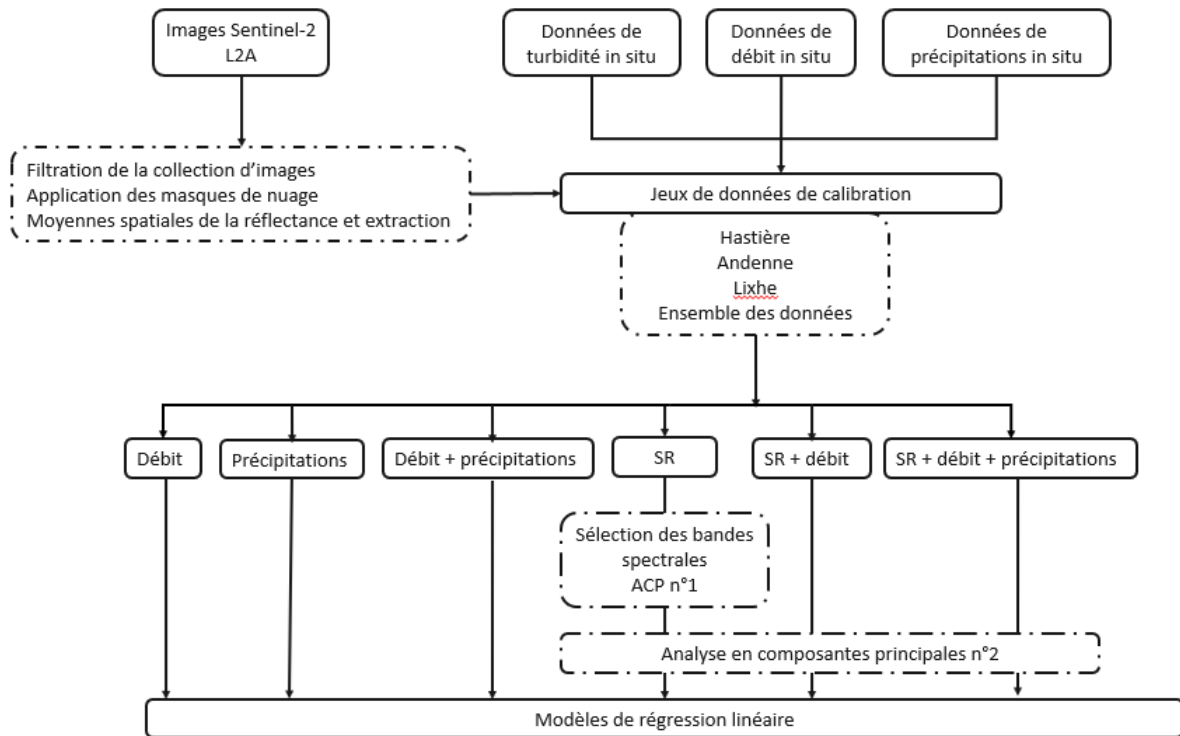


FIGURE 3 – Schéma de la méthodologie suivie.

4.2 Zone d'étude

4.2.1 Situation géographique

La Meuse est l'un des trois fleuves traversant la Belgique, elle prend sa source dans le nord de la France pour aller se jeter dans la mer du Nord des Pays-Bas. Longue de 950 km, seulement 192 traversent la Belgique. Le cours d'eau se situe dans le bassin hydrographique de la Meuse (figure 4) dont la Wallonie compte 12 230 km², soit un peu plus d'un tiers de sa superficie totale. Ce bassin comprend deux sous-bassins : Meuse amont et Meuse aval.



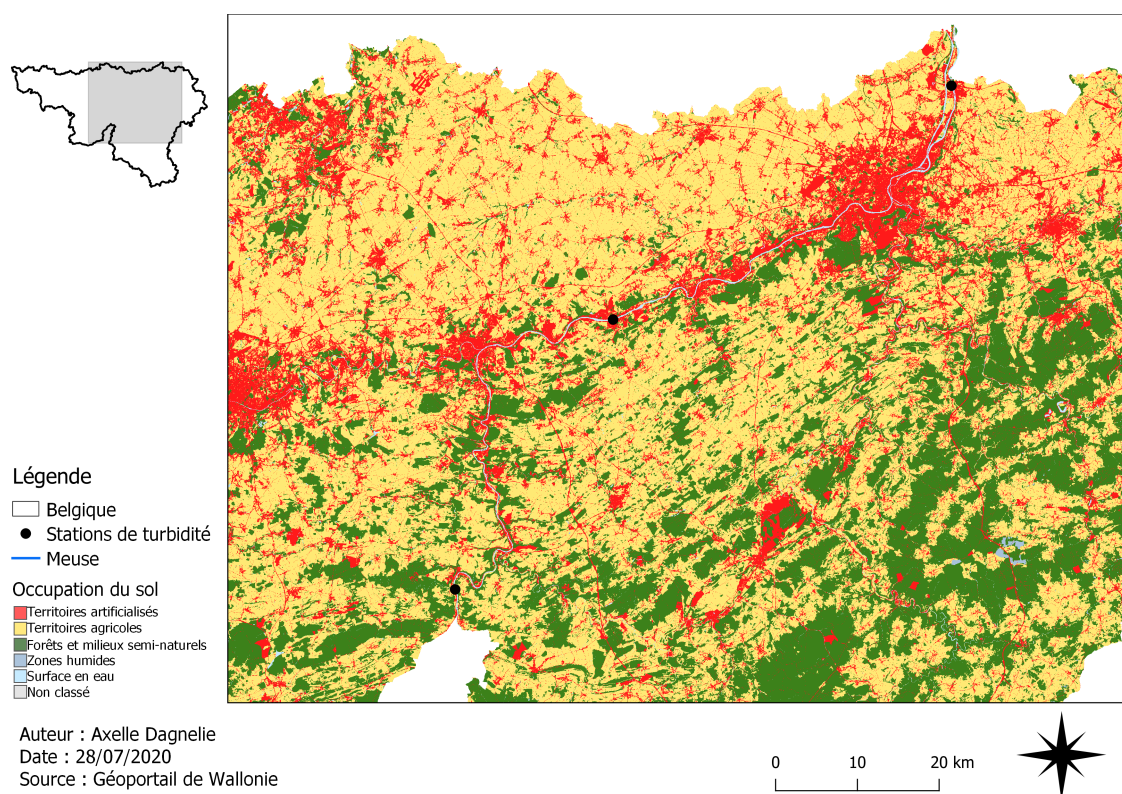
FIGURE 4 – *Bassin hydrographique de la Meuse en Belgique*
 (https://commons.wikimedia.org/wiki/File:Bassin_de_la_Meuse.svg, consulté le 28/07/2020)

Climatologie

Le climat tempéré de la Belgique se caractérise par des étés relativement frais et humides et des hivers relativement doux et pluvieux. La pluviométrie annuelle sur la région du bassin de la Meuse est d'environ 800 à 900 mm, moyenne calculée sur la période de 1981 à 2010 (IRM, 2020).

Occupation du sol

L'occupation est un paramètre important à prendre en compte lors d'étude sur la qualité de l'eau. Le type d'occupation du sol et les activités liées sont des facteurs influençant fortement les particules en suspension de l'eau. La carte ci-dessous (figure 5) représente l'occupation du sol en Wallonie.

FIGURE 5 – *Carte d'occupation du sol de la zone d'étude.*

La majorité du bassin de la Meuse situé en région wallonne est consacré à l'agriculture. La carte montre également une importante concentration d'habitats, représentées en rouges par les terrains artificialisés, aux abords de la Meuse et plus particulièrement au nord.

4.2.2 Localisation et conditions environnementales des stations d'observation

Ces données de terrain, comprenant la turbidité, le débit et les précipitations, furent collectées à différentes stations le long de la Meuse par le Service Public de Wallon (SPW) et l'Institut Royal Météorologique (IRM) belge.

Les stations de mesure ayant servi dans le cadre de ce travail sont représentées sur la figure 6.

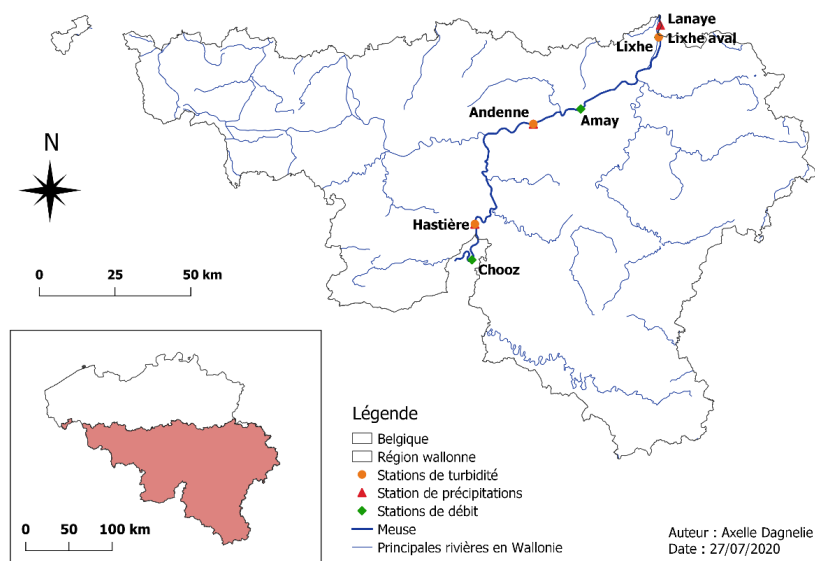


FIGURE 6 – Localisation des stations de mesure.

Stations de mesure de la turbidité

Située à proximité immédiate de la Meuse, une pompe, installée à l'intérieur de la station, aspire et rejette l'eau de la rivière en continu. L'eau est prélevée à proximité plus ou moins immédiate de la berge et à une profondeur variante entre un et deux mètres.

La station de Hastière (figure 7) se trouve dans le sous-bassin hydrographique Meuse amont dans le village de Hermeton-sur-Meuse. C'est dans la commune d'Hastière que la Meuse passe la frontière belge. La station portant le même nom est située en amont d'un barrage, où les alentours sont principalement boisés.



FIGURE 7 – Photo de la localisation de la station d'Hastière.

Le cours d'eau continue son trajet par la station d'Andenne (figure 8). La ville est localisée à 16 km à vol-d'oiseau de la ville de Namur. À 10 km en amont de la station, à proximité de la berge dans la localité de Sclayn, se trouve une carrière de dolomies toujours en exploitation aujourd'hui. Par ailleurs, un barrage se trouve en amont de la station.

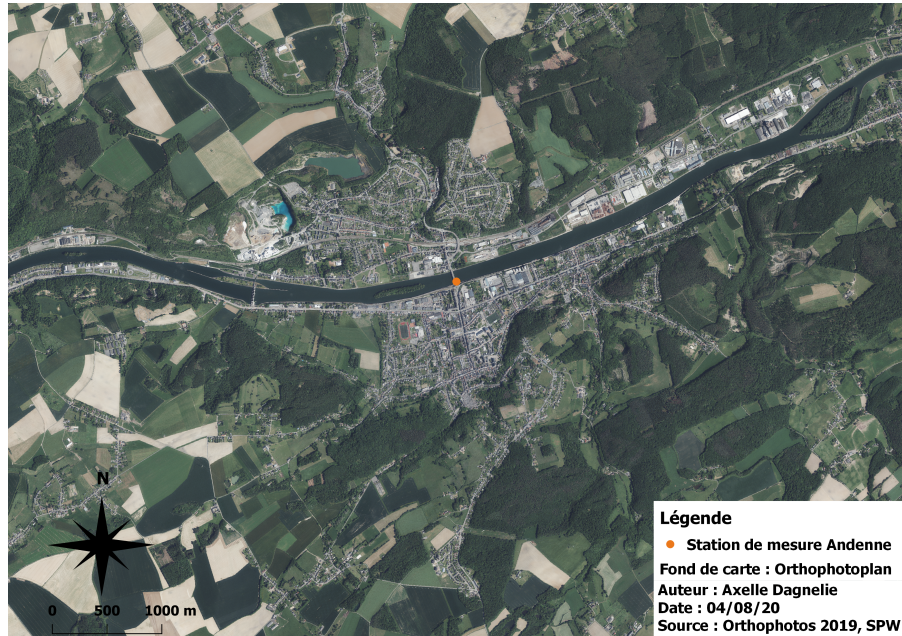


FIGURE 8 – Photo de la localisation de la station d'Andenne.

La station de Lixhe (figure 9) est localisée dans la ville de Visé, établie entre le canal Albert et la Meuse, à 15 km à vol-d'oiseau et en aval de la ville de Liège. La Meuse traverse ensuite la frontière belge au niveau de la province de Liège.



FIGURE 9 – Photo de la localisation de la station de Lixhe.

Stations de mesure du débit

Les trois stations de débit, Chooz (France), Amay et Lixhe (figure 6), ne sont pas installées aux mêmes endroits que les stations de mesure de turbidité. Toutefois, elles sont relativement proches que pour pouvoir les comparer avec environ 12 km entre Chooz et Hastière et 16 km entre Andenne et Amay.

Stations de mesure des précipitations

Les stations d'observation des précipitations (figure 6) sont localisées à Hastière, Andenne et Lanaye, à seulement 4 km de Lixhe.

4.3 Construction des bases de données

4.3.1 Collecte et prétraitement des données de terrain

Données de turbidité

Les données de turbidité furent récoltées aux stations de mesure appartenant au réseau Aquapol (<http://environnement.wallonie.be/aquapol/>) du Service Public de Wallonie (SPW). Le SPW est chargé de l'analyse en continu de la qualité des eaux de surface de la Région wallonne. Il s'occupe de l'archivage, de la validation et de la diffusion des données. La répartition des stations Aquapol sur le bassin hydrographique s'explique par la volonté de cibler les cours d'eau importants, les points transfrontaliers, l'impact de certains affluents ainsi que l'impact des zones urbaines et/ou industrielles.

Les données de turbidité sont récupérées toutes les deux heures à partir de minuit (00h00) et retransmises en unité de turbidité néphélogométrique (NTU). Elles sont mesurées grâce à un turbidimètre Hach Lange Ultraturb Sc plus (HACH, 2020) dont la gamme de mesures varie entre 0 et 1000 NTU. Cependant, la plage des données sélectionnées par le SPW ne dépasse pas 100 NTU. Le SPW justifie cela par le fait que la turbidité dépasse rarement 100 NTU et que la gamme de mesure 0 – 1000 NTU fait perdre en précision par rapport à la gamme réduite. Les données de turbidité demandées couvrent la période allant de 2015 à 2019. Cependant, la validation prenant du temps, les mesures pour l'année 2019 n'étaient pas disponibles à la date de réalisation de cette étude.

Le principe de fonctionnement du capteur de turbidité submersible repose sur un faisceau lumineux dirigé dans l'échantillon d'eau et mesurant la lumière qui se disperse ou est absorbée par les particules en suspension dans l'eau. La réponse du capteur est liée à la longueur d'onde de la lumière incidente et à la taille, la forme et la composition des particules dans l'eau (WAGNER et al., 2006).

Une fois les valeurs mesurées par le turbidimètre, elles sont transférées au système de télé-transmission qui envoie les données à l'application Aquapol.

La validation des données s'effectue de façon automatique dans l'application Aquapol. Il est toutefois nécessaire d'effectuer une seconde validation manuelle par le SPW sur ces données,

toujours dans cette même application Aquapol, afin de détecter et corriger les erreurs comme des mesures aberrantes ou des dérives qui ne seraient pas détectées lors de la première validation automatique.

Données de débit

La turbidité et le débit d'eau sont liés de façon causale (CHEN et al., 2019; KEMKER, 2014). Des débits élevés peuvent maintenir les particules en suspension au lieu de les laisser se déposer au fond. Ainsi, dans les rivières et autres environnements naturels à fort débit, la turbidité peut varier en fonction des débits. Dans ces zones, il est important de surveiller les débits et la turbidité d'une façon simultanée.

Trois types de facteurs sont responsables de la variabilité des débits des cours d'eau : les aléas climatiques tels que la durée et l'intensité des précipitations, les caractéristiques des rivières et de leurs bassins versants tels que la topographie, le type de sols et de sous-sols, et certaines activités humaines comme l'imperméabilisation, la navigation, le prélèvements en eau, les barrages, etc (SPW, 2020).

Les données de débit proviennent du réseau de mesure « Water CONtrol DAta system for Hydrology and water management » (WACONDAH) de la Direction générale opérationnelle de la Mobilité et des Voies hydrauliques du SPW.

Les données de débit comprennent les mesures journalières en m^3/s pour les stations de Chooz, Amay et Lixhe aval pour la période de 2015 à 2019. Le débit est mesuré grâce à un débitmètre à ultrasons. La qualité des mesures est garantie par le respect d'une série de procédures dont un contrôle quotidien des mesures télétransmises, la validation par bilan, la comparaison et l'estimation sur base des stations voisines, etc.

Données de précipitations

Par ailleurs, les conditions météorologiques, en particulier les fortes précipitations, affectent également le débit d'eau, qui à son tour affectent la turbidité. Les précipitations peuvent augmenter le volume des cours d'eau et donc le débit des cours d'eau, ce qui peut remettre en suspension les sédiments décantés et éroder les berges (KEMKER, 2014). Toutefois, cela peut également diminuer la concentration des particules en suspension par effet de dilution.

Les valeurs de turbidité sont souvent sujettes à une augmentation considérable pendant les tempêtes intenses en raison de la remobilisation et de la remise en suspension des sédiments du lit ainsi que des sédiments nouvellement érodés provenant de diverses sources dans le bassin versant. Selon la région, elles peuvent également augmenter après de petites pluies suivant une période sèche prolongée en raison de l'accumulation de sédiments sur la surface du sol qui, si elle est érodée, a un temps de suspension plus long que le sous-sol (CHEN et al., 2019).

Des stations pluviométriques sont positionnées à Hastière, Andenne et Lanaye et sont les propriétés de l'Institut Royal Météorologique (IRM). Les données de précipitations sont mesurées en mm/jour. Le réseau climatologique belge compte environ 200 stations réparties à travers

tout le pays et est constitué d'observateurs quasi bénévoles recrutés par l'IRM, et d'agents de l'Etat, des Régions ou d'entreprises diverses. Tous les matins vers 8 heures, les observateurs relèvent la quantité d'eau tombée en 24 heures dans un pluviomètre de l'IRM. Ces mesures sont complétées, selon le cas et la disponibilité des observateurs, par des observations supplémentaires en cours de journée, telles que la répartition des pluies au cours de la journée, la présence de brouillard, l'épaisseur de neige au sol ou le passage d'un orage au-dessus de la station. Quotidiennement ou mensuellement selon les cas, les observateurs transmettent à l'IRM le résultat de leurs relevés journaliers. Les informations sont centralisées dans la banque de données de l'IRM et la qualité des données est contrôlée.

4.3.2 Collecte et prétraitement des données satellites

Descriptions des produits satellites issus de Sentinel-2

Les produits satellitaires utilisés ici sont issus de l'imagerie de Sentinel-2, plus spécifiquement de la constellation de Sentinel-2A et Sentinel-2B, et ont été obtenues via la plateforme Google Earth Engine (GEE). Les images y sont elles-mêmes téléchargées depuis le Copernicus Open Access Hub. Les images sont accessibles à différents niveaux de traitement. Dans le cadre de ce mémoire, le choix s'est porté sur les images de niveau L2A contenant les données de réflectance de surface orthorectifiées avec un enregistrement multispectral sous-pixel. Les masques de nuages et terre/eau sont inclus dans le produit. Pour acquérir ce niveau, les images ont été traitées en exécutant la correction atmosphérique 'Sen2cor' (ESA, 2020). L'agence spatiale européenne (ESA) n'a pas produit de données L2A pour tous les actifs de niveau L1C et la couverture L2A antérieure n'est pas globale. Cependant, les données de l'Europe étant traitées en priorité, il est possible d'avoir un nombre d'images plus important pour la région de la Meuse étudiée.

Sen2cor est un algorithme permettant la génération et le formatage de produits Sentinel-2 niveau 2A. Il utilise les bandes 1, 9 et 10 à 60 mètres de résolution, respectivement sensibles aux aérosols, à la vapeur d'eau et aux cirrus, et effectue les corrections atmosphériques, de terrain et des cirrus des données d'entrée de niveau 1C du haut de l'atmosphère (TOA) des autres bandes. Sen2Cor crée des images de réflectance corrigées du fond de l'atmosphère (BOA) avec un format de produit de sortie équivalant au produit utilisateur de niveau 1C. Cette opération permet donc de réaliser la correction atmosphérique des bandes à 10 et 20 m de résolution utilisées pour la suite du travail (MAIN-KNORN et al., 2017).

Les actifs L2A contiennent 12 bandes spectrales représentant la réflectance à la surface de la Terre à une échelle de 10 000 et contrairement aux données 1C, il n'y a pas de bande 10 pour les images L2A disponibles sur GEE. Les bandes spectrales couvrent les domaines de longueurs d'onde allant du visible à l'infrarouge à ondes courtes avec des résolutions spatiales de 10, 20 ou 60 mètres. Les détails de ces bandes se trouvent en annexe A. De plus, trois bandes QA (Quality Assessment) sont présentes et font office de bandes de correction, dont une (QA60) est une bande de masque de bits avec des informations de masque de nuage (GEE, 2020). Il

existe également des produits supplémentaires pour le niveau 2A comme une carte d'épaisseur optique des aérosols (AOT), une carte de vapeur d'eau (WV) et une carte de classification des scènes (SCL) avec des indicateurs de qualité pour les probabilités de nuages et de neige.

Les produits atmosphériquement corrigés et orthorectifiés de niveaux 2A se trouvent sous forme de granules, également appelés tuiles, et sont constitués d'ortho-images carrées de 100 km par 100 km en projection UTM / WGS84. Il y a une tuile par bande spectrale (ESA, 2015). Chaque prise de données Sentinel-2 est coupée en morceaux pour être plus petite et plus facile à manipuler.

La mission Sentinel-2 comporte en réalité deux satellites Sentinel-2A et Sentinel-2B lancés respectivement le 23 juin 2015 et le 7 mars 2017. Ces satellites ayant chacun une couverture répétée de 10 jours à l'équateur offrent un temps de revisite réduit à 5 jours grâce à leur orbite décalée de 180 degrés (ESA, 2020b). En raison du chevauchement entre les fauchées des orbites adjacentes, la fréquence de revisite sera augmentée en fonction des différentes conditions d'observation (figure 10). Les images de la zone d'étude du bassin de la Meuse sont donc disponibles tous les deux jours.

Par ailleurs, le passage des Sentinel-2A et 2B au-dessus d'une zone se fait à 10h30 heure locale, choisie comme le meilleur compromis entre la minimisation de la couverture nuageuse et la garantie d'un éclairage solaire adéquat.

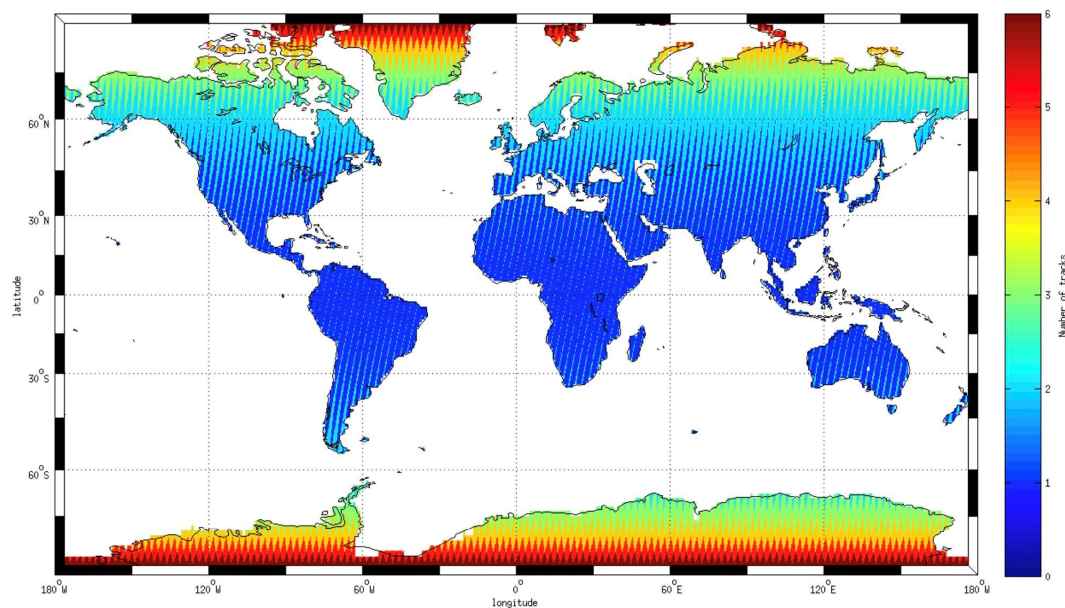


FIGURE 10 – *Fréquence de revisite géométrique due au chevauchement des orbites adjacentes de Sentinel-2.*

Traitement des images Sentinel-2

Une grande partie du traitement des données de télédétection de Sentinel-2 fut réalisée dans l'environnement GEE. Cette plateforme permet l'accès, le traitement et la visualisation rapide de collections d'images satellites via une interface facile d'utilisation, et le travail à distance sur le cloud. Les étapes sont décrites ci-après et le code se trouve en annexe B.

Filtration de la collection d'images

Les images importées dans l'environnement de travail furent sélectionnées selon différents critères :

- la période correspondant à celle de la disponibilité des mesures de turbidité, dans ce cas-ci, la période s'étend du 28 mars 2017, date à laquelle les images Sentinel-2 niveau 2A sont disponibles, jusque le 31 décembre 2018, date après laquelle les données de turbidité ne sont plus disponibles ;
- la zone d'étude, correspondant ici aux localisations des trois stations de mesure de turbidité.

Application des masques de nuage

À la suite de ce premier tri, deux masques ont été appliqués aux images pour masquer les pixels contaminés par des nuages.

Le premier masque est contenu dans la bande 'QA60' intégrée dans les données Sentinel-2 niveau 1C et 2A, où les nuages opaques et cirrus sont calculés sur base de critères spectraux.

Étant donné que la bande QA60 actuelle ne peut pas fournir de résultats de détection de nuages et d'ombres de nuages précis (YANG et al., 2020), un autre masque fut appliqué pour réduire le potentiel d'inclusion d'erreurs d'omission : le masque de classification des scènes (SCL), un des produits disponibles pour le niveau 2A. L'algorithme SCL permet de détecter les nuages, leurs ombres et la neige, et de générer une carte de classification. Cette dernière se compose de 11 classes : neige/glace, cirrus, eau, sols nus, végétation, ombre de nuages, pixels de zone sombre, pixels non classés/ faible probabilité de nuage, saturés ou défectueux et 2 classes de probabilités de nuages (moyenne et élevée). Dans le contexte de ce travail, seules les classes concernant les cirrus, les ombres de nuages, les pixels de zone sombre, les saturés ou défectueux et les classes de probabilité de nuages étaient utiles à la correction des images de la collection. Le fonctionnement des deux masques est illustré à la figure 11.

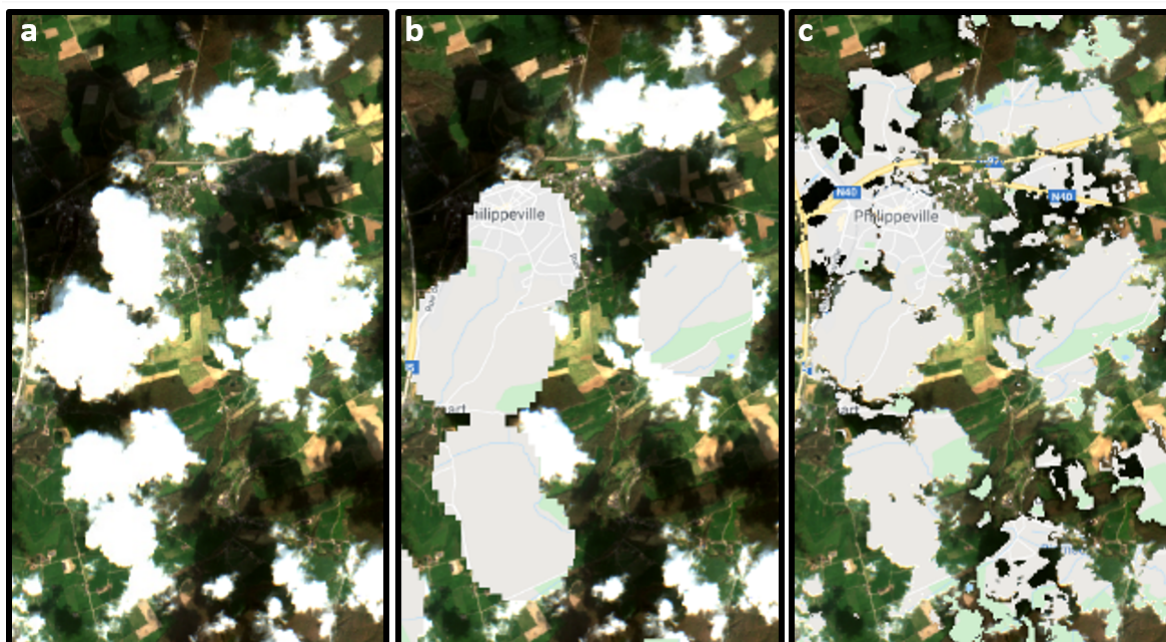


FIGURE 11 – Illustrations des masques de nuages appliqués aux images Sentinel-2 sur GEE, a) l'image brute, b) l'image avec le masque QA60, c) l'image avec le masque SLC, image datant du 03-04-2020.

Moyennes spatiales de la réflectance et export des données

Pour chaque image validée à l'étape précédente, les valeurs de réflectance de chaque bande furent extraites.

L'extraction des réflectances de surface s'est faite sur une fenêtre de 3x3 pixels située au milieu de la rivière et à moins de 100 mètres de la station de mesure pour chaque localisation. Une fenêtre d'un certain nombre de pixels (3x3, 5x5 ou 7x7) centrée sur l'emplacement de la mesure in situ permet de générer des statistiques simples, telles que la moyenne et l'écart-type, pour aider à l'évaluation de la stabilité spatiale, ou de l'homogénéité, au point de validation. En outre, l'utilisation d'une fenêtre multi-pixels augmente la possibilité qu'une mesure in situ soit disponible pour validation en augmentant les chances que la récupération satellite ait suffisamment de pixels clairs pour être utilisées. Ensuite, le pixel le plus proche de l'emplacement in situ n'a pas besoin d'être clair, tant que les pixels valides dans la boîte répondent à l'exigence d'homogénéité (BAILEY et al., 2006).

La résolution choisie pour le rééchantillonnage des pixels diffère d'une étude à l'autre et varie entre 10 et 60 m. Comme les pixels ayant une résolution de 10 mètres se trouvent dans les bandes du visibles qui sont les plus utilisées pour l'observation de la turbidité, la fenêtre fut donc réalisée grâce à l'implémentation d'un cadre de 30 sur 30 mètres permettant d'y inclure les 9 pixels de résolution de 10 m à prendre en compte (BONANSEA et al., 2019; HUOVINEN et al., 2019). La moyenne de ces 9 pixels fut calculée grâce à un réducteur pondéré (*ee.Reducer.Mean*). Les pixels sont inclus dans la moyenne si au moins approximativement 0,5% du pixel se trouve dans la région et si le masque de l'image est différent de 0 (DEVELOPERS, 2020).

$$\text{Moyenne pondérée} = \frac{\sum(\text{valeur} * \text{poids})}{\sum(\text{poids})} \quad (1)$$

Leur poids est le minimum du masque de l'image et la fraction approximative du pixel couvert par la région. Les photos de la figure 12 illustrent les fenêtres de 3x3 pixels aux trois localisations des stations de turbidité.



FIGURE 12 – Zones de sélection de la fenêtre de 3x3 pixels représentées sur les images Sentinel-2 niveau 2A disponibles sur GEE, de haut en bas : Hastière, Andenne, Lixhe.

Les données de réflectances furent ensuite téléchargées pour les 12 bandes et pour chaque station sous format 'CSV' et à une échelle de 1/10 000.

Une fois les données de réflectance organisées dans un fichier Microsoft Excel, elles furent vérifiées manuellement pour s'assurer que toutes les images restantes soient de bonne qualité et pour vérifier que les masques aient bien fonctionné.

La station de Andenne se situe au chevauchement de deux tuiles voisines de Sentinel-2, donnant ainsi deux mesures de réflectance différentes pour une même date. Pour que les séries chronologiques soient cohérentes, les pixels qui se chevauchent furent harmonisés. Une solution décente consiste à prendre toutes les valeurs de réflectance pour le même jour, générées à partir des mêmes pixels bruts, et à attribuer la moyenne de toutes les valeurs à ce jour. Il

en résulte une sortie propre avec une valeur de réflectance de surface par point par jour. Le nombre d'images pour Andenne passe donc de 89 à 45.

Disponibilité des images satellites valides pour la modélisation

L'application des filtres et masques de nuage a fortement réduit le nombre d'images disponibles. Le tableau 3 reprend le nombre d'images de départ pour la période de mars 2017 à décembre 2018 et le nombre d'images restantes pour chaque station pour la même période.

Une dernière vérification visuelle de chaque image, faisable grâce au nombre réduit d'images, a été effectuée pour confirmer leur qualité après application des masques. Suite à cela, encore 3, 5 et 6 images pour les stations de Hastière, Andenne et Lixhe respectivement, furent écartées.

TABLEAU 3 – *Nombre d'images considérées dans l'analyse.*

	Hastière	Andenne	Lixhe
Avant masques	194	388	195
Après masques	40	45	48
Après vérification visuelle	37	31	48

Cela montre qu'environ 20% des images était d'assez bonne qualité pour être utilisées dans les conditions de ce travail.

4.4 Méthodes de modélisation

4.4.1 Définition des scénarios de modélisation

Afin de déterminer laquelle des variables prédictives ou combinaison des variables est la plus efficace pour prédire la turbidité, plusieurs modèles sont testés avec les différentes variables prédictives prises individuellement ou combinées, ci-après référés à 'scénarios' :

1. La réflectance de surface associée aux 12 bandes spectrales
2. Le débit
3. Les précipitations
4. La réflectance de surface et le débit
5. Le débit et les précipitations
6. La réflectance de surface, le débit et les précipitations

Assignment des valeurs de turbidité aux variables prédictives

Les séries temporelles des différentes variables n'étant pas toutes continues, les ensembles de données furent réajustés afin que les dates de prises de mesures des différentes variables correspondent. Au vu du nombre restreint de données, c'est la réflectance de surface la variable limitante avec 37, 31 et 40 mesures pour Hastière, Andenne et Lixhe respectivement. Pour

chaque image valide, les valeurs des autres variables furent donc associées aux 12 bandes spectrales selon leur date.

La correspondance entre la turbidité et les données satellites s'est faite sur base des mesures horaires de turbidité. BAILEY et al. (2006) suggèrent que les correspondances entre les données satellitaires et in situ devraient idéalement se limiter aux mesures de terrain recueillies dans une fenêtre de trois heures au passage du satellite. D'autres auteurs ont démontré de bonnes correspondances en utilisant des fenêtres moins restrictives allant jusqu'à trois jours dans le cas de lacs sous conditions hydrologiques et atmosphériques stables. Pour la situation présente, l'heure de passage des satellites Sentinel-2 au-dessus de la Belgique est d'approximativement 10h40, avec un maximum de 10h57. Les données de turbidité étant disponibles toutes les deux heures, la turbidité mesurée à 10 heures a été utilisée, ce qui permet un décalage horaire avec l'heure de détection des images de maximum une heure pour l'ensemble des mesures de turbidité.

Les données de débit sont des données moyennes journalières, mesurées en m^3/s . Dans le but de garder une certaine cohérence entre les types de données utilisés, les données de turbidité mesurées à 10 heures furent comparées aux moyennes journalières de la turbidité. Ces moyennes sont également disponibles via le SPW et ont été calculées sur base des mesures horaires de turbidité. Les statistiques descriptives reprenant la moyenne, la médiane, l'écart-type, le minimum, le maximum et le nombre d'échantillons seront repris dans un tableau comparatif entre la turbidité mesurée à 10h et la turbidité moyenne journalière.

Afin de déterminer la relation la plus adéquate des précipitations avec la turbidité, plusieurs tests ont été réalisés. Dans un premier temps, l'exercice s'est porté sur le choix du type de données. Dans le cas présent, les résultats montraient une meilleure relation entre les précipitations cumulées et la turbidité, en outre confirmé par l'article de CHEN et al. (2019). Dans un deuxième temps, le nombre de jours de précipitations à cumuler fut décidé sur base de tests sur 2, 3 et 5 jours. Ce dernier montrant de meilleurs résultats, les précipitations cumulées sur 5 jours furent donc choisies pour ce cas d'étude. Par facilité, pour la suite du travail, les précipitations feront toujours allusion aux précipitations cumulées sauf si mention contraire.

Pour chaque scénario, trois modèles spécifiques aux stations et un modèle généralisé à l'ensemble de la rivière sont créés. Au total, 24 jeux de données de calibration sont donc exploités pour l'élaboration de 24 modèles respectifs.

Un tableau synthétique reprendra la description des jeux de données impliqués dans les différents niveaux cités précédemment.

4.4.2 Analyse exploratoire et prétraitement des variables prédictives

Analyse univariée des jeux de données

Afin de construire des modèles de régression linéaire robustes, il est nécessaire que les différentes variables aient des jeux de données adéquats, c'est-à-dire avoir un nombre approprié de

données dont la distribution soit équitablement répartie sur leur gamme de valeurs et à travers le temps.

Les histogrammes et l'analyse de la distribution dans le temps ont permis d'évaluer si les jeux de données sont suffisants pour développer un modèle de régression fiable.

Normalisation

Dans le cadre de ce mémoire, des modèles de régression linéaire sont établis pour prédire la turbidité en fonction des variables prédictives définies. Les méthodes de régression pour obtenir ces modèles impliquent que la normalité des variables de régression soit respectée. Certaines des variables, notamment le débit et la turbidité, ne respectent pas cette exigence. Sur base de la littérature, il a été décidé de tester la transformation logarithmique pour améliorer la normalité du débit et de la turbidité. En effet, plusieurs cas d'étude ont transformé la turbidité, ou les sédiments en suspension, en logarithme de base 10 (BUSTAMANTE et al., 2009 ; KIM et al., 2017 ; MONTANHER et al., 2014 ; RASMUSSEN et al., 2009) ainsi que le débit (CHEN et al., 2019 ; IONUȚ et al., 2014).

L'étude des graphiques de dispersions de points, des résidus et de la normalité permet de mettre en évidence la nécessité ou non d'une transformation des variables de départ (HELSEL et al., 2002). L'étude de la transformation éventuelle des variables a été faite uniquement sur l'ensemble des données et non sur les données de chaque station.

La validation de la transformation peut se faire grâce à l'analyse des graphiques de nuage de points, du graphique des résidus de la régression en fonction de la variable réponse et du graphique de la normalité des résidus.

La transformation acceptable des variables peut se justifier en comparant les tracés des résidus et les graphiques de probabilité en recherchant trois choses :

- une variance constante à travers la gamme de la variable prédite,
- la normalité des résidus,
- un motif linéaire, et non pas une courbure, centré verticalement autour de la ligne résiduelle zéro.

4.4.3 Analyse en composantes principales : identification des variables prédictives

L'Analyse en Composantes Principales (ACP) a été utilisée pour extraire et visualiser les informations importantes contenues dans les jeux de données. L'ACP est une méthode statistique qui permet d'explorer des données dites multivariées.

Dans la situation présente, deux ACP furent réalisées, la première afin de déterminer les bandes spectrales de Sentinel-2 les mieux corrélées avec la turbidité et la deuxième en vue de créer de nouvelles variables avec les bandes restantes, le débit et les précipitations pour les implémenter dans les modèles de régression linéaire.

Les ACP ont été implémentées dans l'environnement R via l'interface Jupyter Notebook.

Détermination des bandes spectrales corrélées avec la turbidité

L'ACP fut réalisée en incluant la turbidité et les 12 bandes spectrales pour chaque station séparément et pour l'ensemble des stations. Cette analyse permet de vérifier des corrélations entre variables brutes, c'est-à-dire ayant une distribution quelconque et n'ayant pas les mêmes unités. Dans l'analyse en composantes principales, les variables sont souvent standardisées (écart type égal à un et moyenne égale à zéro) ce qui permet d'analyser des variables avec des unités différentes.

Les résultats ont été interprétés sur base de l'étude visuelle des graphiques de corrélation montrant les relations entre les variables et les axes, ainsi qu'entre les variables elles-mêmes. La proximité entre deux variables donne une approximation de leur corrélation et peut être interprétée comme suit :

- les variables proches sont corrélées positivement,
- les variables qui s'opposent, positionnées sur les côtés opposés de l'origine du graphique (quadrants opposés), sont corrélées négativement,
- les variables orthogonales sont non corrélées.

Construction des nouvelles variables

Suite à la première ACP, une deuxième fut réalisée cette fois-ci pour construire de nouvelles variables. L'ACP consiste à transformer des variables corrélées entre elles en nouvelles variables décorréliées les unes des autres, appelées composantes principales (CP). Ces nouvelles variables correspondent à une combinaison linéaire des variables originelles. Cela permet de réduire le nombre de variables et de rendre l'information moins redondante. Ces CPs ont l'avantage d'être indépendantes entre elles.

Cette deuxième ACP fut réalisée uniquement sur les scénarios 1 (réflectance de surface), 5 (débit et réflectance de surface) et 6 (réflectance de surface, débit et précipitations). Les autres scénarios (2, 3 et 4) ne comprennent qu'une ou deux variables prédictives au maximum, il n'est donc pas nécessaire d'effectuer une ACP.

La construction des CP s'est faite sur base des deux étapes suivantes.

La première étape consiste à déterminer le nombre de CP à inclure au modèle pour obtenir un résumé suffisamment précis de l'information. La quantité de variance expliquée par chaque composante principale est mesurée par la valeur propre de la matrice de variance-covariance des données. Plusieurs options sont alors possibles pour choisir le nombre d'axes selon les valeurs propres obtenues. Dans le cas de ce travail, la variance expliquée par les composantes principales devait être supérieur à 95%, critère déjà utilisé lors de précédentes études (DOERFFER et al., 1989 ; FLINK et al., 2001).

La deuxième étape construit les nouvelles CP. Grâce aux valeurs calculées par l'ACP, il est possible d'élaborer les nouvelles CP en fonction des variables retenues. Les CP seront alors sous forme d'équation comme par exemple :

$$CP1 = x_1a + x_2b + x_3c + x_4d + \dots \quad (2)$$

où x_i sont les facteurs et a, b, c et d les différentes variables prédictives.

4.4.4 Développement des modèles prédictifs

Implémentation des nouvelles variables dans des modèles de régression linéaire

Les ACP correspondantes ont ensuite été insérées dans une équation de régression empirique.

Finalement, pour chaque scénario et chaque station, les variables prédictives furent ajustées dans un modèle de régression linéaire à la variable réponse, la turbidité [NTU] transformée en logarithme de base 10. Les variables prédictives utilisées furent donc les nouvelles composantes principales pour les scénarios 1, 5 et 6 comprenant les bandes spectrales, le débit également transformé en logarithme de base 10 et les précipitations cumulées; le logarithme du débit et les précipitations cumulées pour les scénarios 2, 3 et 4 n'ayant pas nécessités d'analyse en composantes principales. Les modèles de régression linéaires furent construits dans le logiciel R.

Évaluation des régressions grâce aux indicateurs de performance

La qualité des 24 modèles finaux fut comparée sur base des indicateurs suivants.

Le coefficient de détermination ajusté pour le nombre de variables prédictives (R^2a) indique la fraction de la variance de la variable réponse qui est expliquée par le modèle. La racine carrée de l'erreur quadratique moyenne ("Root Mean Square Error", RMSE) est une mesure de la variance entre les valeurs calculées par la régression et les valeurs observées.

Malgré que ces métriques soient largement utilisées, elles doivent être employées avec précaution et en combinaison avec d'autres métriques plus robustes.

Le Critère d'Information Bayésien (BIC) est un indicateur du surajustement du modèle, un terme de pénalité est introduit dans le BIC pour le nombre de paramètres dans le modèle. Le modèle ayant le BIC le plus bas est préféré. Il est défini par l'équation suivante.

$$BIC = k \ln(n) - 2 \ln(L) \quad (3)$$

où k est le nombre de paramètres estimés par le modèle statistique, n le nombre d'observations et L est la valeur maximisée de la fonction de vraisemblance du modèle estimé.

5 Résultats

5.1 Descriptions des données

5.1.1 Comparaison des données de turbidité horaires et journalières

Les courbes de la turbidité mesurée à 10 heures suivent quasiment parfaitement les courbes des moyennes journalières de turbidité. Les seuls écarts sont dus à des manques de données (figures 13, 14 et 15). Ce qui est confirmé par le tableau 4 des descriptions des statistiques des deux ensembles de données. Les données horaires peuvent donc être utilisées pour représenter les données journalières et ainsi être manipulées en association avec les moyennes journalières des débits.

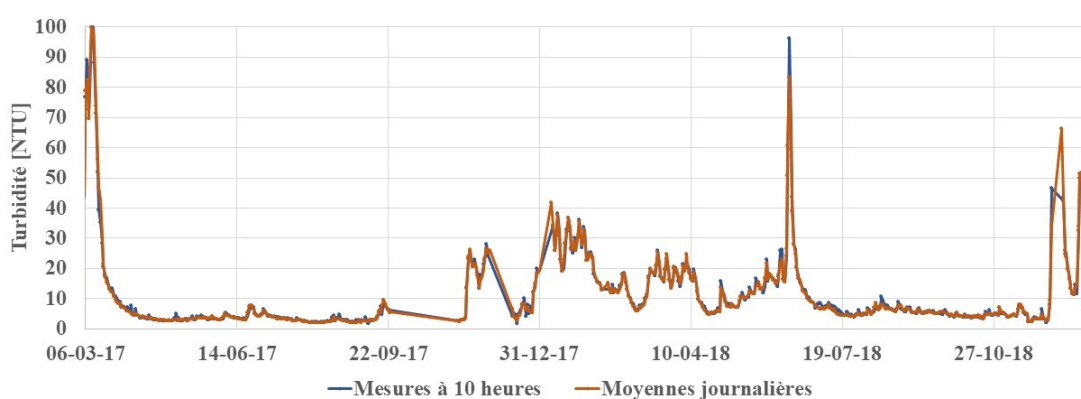


FIGURE 13 – *Comparaison des données de turbidité pour la station d'Hastière.*

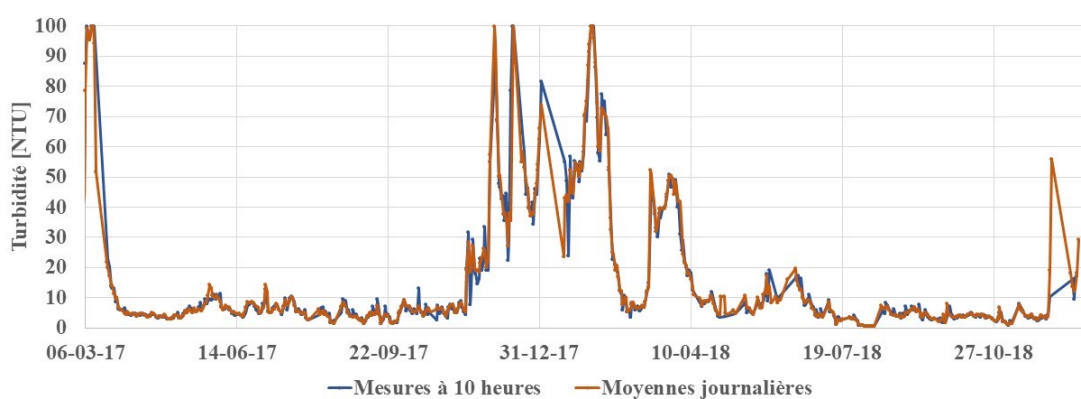


FIGURE 14 – *Comparaison des données de turbidité pour la station d'Andenne.*

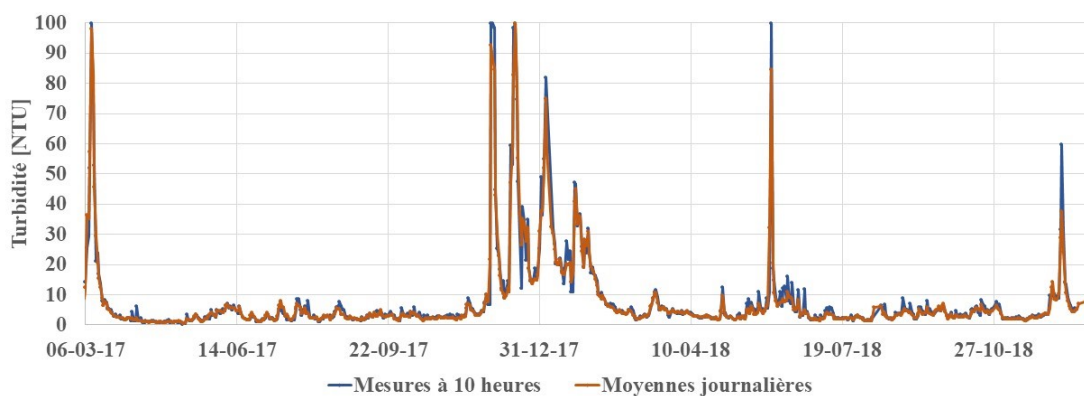


FIGURE 15 – Comparaison des données de turbidité pour la station de Lixhe.

TABLEAU 4 – Descriptions statistiques des données de turbidité mesurées à 10 heures et les moyennes journalières.

	Turbidité mesurée à 10 heures			Turbidité moyenne journalière		
	Hastière	Andenne	Lixhe	Hastière	Andenne	Lixhe
Moyenne	12,31	15,78	6,39	12,17	16,27	6,16
Médiane	6,52	7,43	4,22	6,07	7,56	3,83
Écart-type	14,36	20,95	10,19	14,27	21,26	9,82
Minimum	1,71	0,35	0,23	1,90	0,44	0,34
Maximum	99,77	99,84	99,79	99,77	99,84	99,78
Nombre d'échantillons	1353	1070	1274	1365	1129	1286

5.1.2 Résumé des données retenues pour analyse

Le tableau 5 récapitule les caractéristiques des jeux de données pour développer les modèles prédictifs de la turbidité.

TABLEAU 5 – Caractéristiques des jeux de données.

	Turbidité	Réflectance	Débit	Précipitations
Fréquence	Mesure à 10 heures	Mesures à +/- 10h30	Moyennes journalières	Cumulées sur 5 jours antérieurs
Unités	NTU	0,0001 sr-1	m ³ /s	mm
Période	2015 – 2018	28/03/2017 – 31/12/2018	2015 – 2019	2015 – 2018

5.2 Analyse univariée

5.2.1 Descriptions statistiques

Les tableaux 6 à 9 reprennent les descriptions statistiques des différents jeux de données.

Les séries complètes représentent les ensembles de données fournies par le SPW et l'IRM, tandis que les séries échantillonnées représentent les données qui ont servi à l'étalonnage des modèles prédictifs.

T_h = Turbidité horaire [NTU]

Q = Débit moyen journalier [m^3/s]

P = Précipitations cumulées sur les 5 jours antérieurs [mm]

TABLEAU 6 – Descriptions statistiques des données pour la station d'Hastière.

Hastière	Série complète			Série échantillonnée		
	T_h	Q	P	T_h	Q	P
Nombre	16 261	1826	1461	37	37	37
Minimum	0,8	17,5	0	2,9	18,9	0
Maximum	99,8	775,1	64	21,2	249,3	50,8
Médiane	5,9	81	6,7	6,8	36,6	1,8
Moyenne	12,1	134,6	11,1	8,5	72,5	6,1
Écart-type	14,5	139,3	12,2	5,1	69,4	10,2

TABLEAU 7 – Descriptions statistiques des données pour la station d'Andenne.

Andenne	Série complète			Série échantillonnée		
	T_h	Q	P	T_h	Q	P
Nombre	13 001	1826	1461	31	31	31
Minimum	0,3	18,6	0	0,4	33,5	0
Maximum	100,8	1123,3	75,8	99,8	699	26,1
Médiane	7,5	113	6,4	5,8	60,6	1,5
Moyenne	16	192,4	10,4	10,4	118	5,5
Écart-type	21,2	193	11,5	18,2	136,4	7,5

TABLEAU 8 – Descriptions statistiques des données pour la station de Lixhe.

Lixhe	Série complète			Série échantillonnée		
	T_h	Q	P	T_h	Q	P
Nombre	15 310	1826	1461	40	40	40
Minimum	0,2	6,2	0	0,7	13,7	0
Maximum	99,8	1405	72,5	48,8	994,8	24,2
Médiane	3,8	107,6	5,6	2,7	54,7	1,2
Moyenne	6,1	203,1	9,1	5,1	147,6	4,5
Écart-type	10	224,2	10,3	8,5	213,1	7

TABLEAU 9 – Descriptions statistiques des données pour les trois stations groupées.

Stations groupées	Série complète			Série échantillonnée		
	T_h	Q	P	T_h	Q	P
Nombre	42 263	5478	4383	108	108	108
Minimum	0,2	6,2	0	0,4	13,7	0
Maximum	100,8	1405	75,8	99,8	994,8	50,8
Médiane	5,6	99,1	6,3	4,8	51,8	1,4
Moyenne	11,6	176,7	10,2	7,8	113,4	5,4
Écart-type	16,4	191,1	129,9	11,5	156,3	8,3

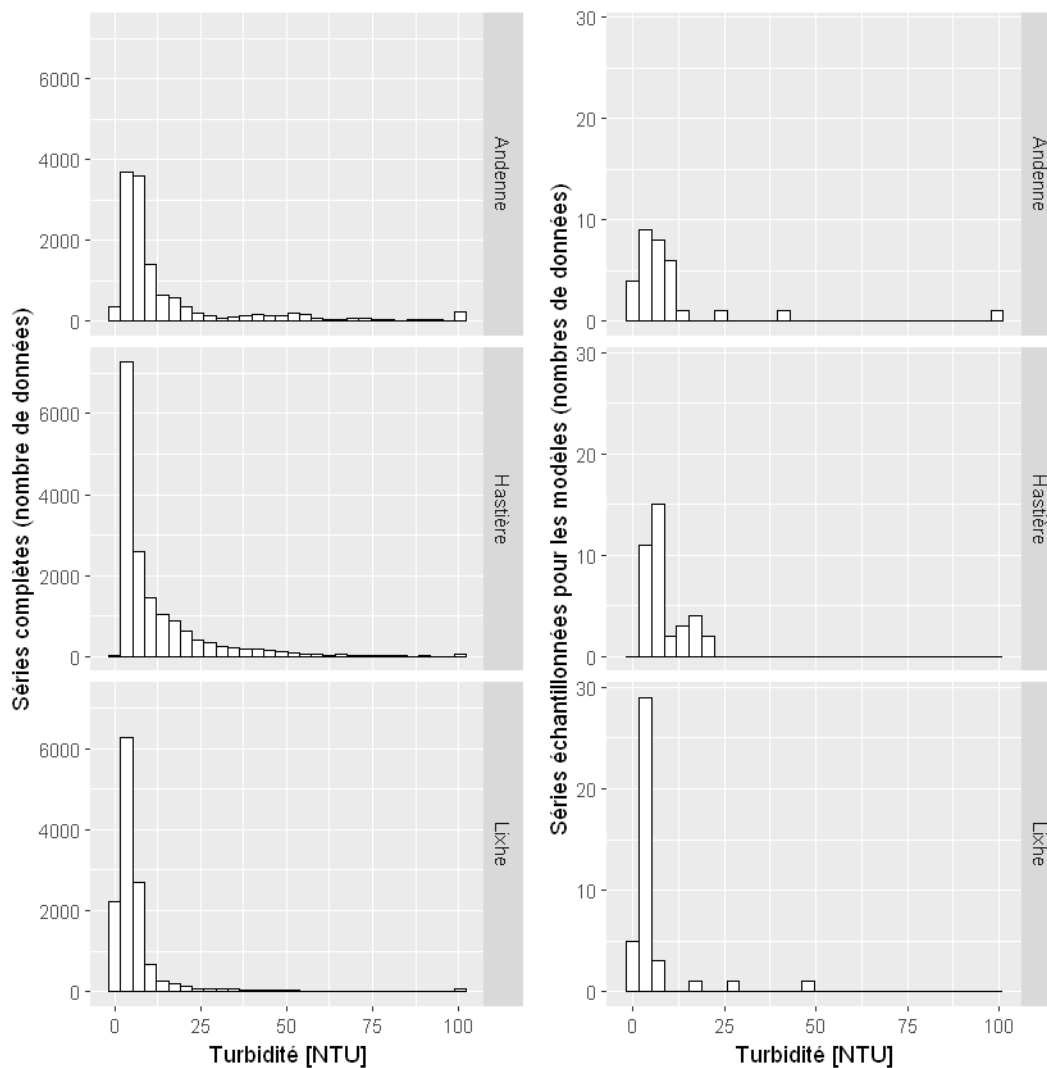
Étant donné la disponibilité très réduite des images satellites et le manque des données de turbidité pour l'année 2019, le nombre de valeur pour les jeux de données s'en voit considérablement limité.

En général, les gammes de valeurs de la turbidité, du débit et des précipitations semblent avoir été nettement raccourcies pour toutes les stations. Il en découle que la plupart des valeurs élevées furent écartées des jeux de données d'étalonnage des modèles. Cela se ressent un peu moins pour les données groupées des trois stations.

Les médianes ont également plus ou moins fortement diminué, sauf pour la turbidité de la station d'Hastière, indiquant par là qu'une plus grosse partie des valeurs se trouvent alors dans des plages de turbidité, débit et précipitations plus faibles.

5.2.2 Distribution de la turbidité sur la gamme de valeur

Pour plus de précision concernant la répartition des valeurs de la turbidité et pour avoir un soutien visuel aux tableaux, des histogrammes de la turbidité ont été construits à partir de l'ensemble des données horaires de turbidité initiales (histogrammes de gauche) et des données de turbidité mesurée à 10 heures servant à la calibration (histogrammes de droite) sur la figure 16.

FIGURE 16 – *Histogrammes des données de turbidité.*

Les histogrammes confirment que les valeurs élevées de turbidité furent écartées des jeux de données d'étalonnage du modèle. Malgré cela, la répartition des données de turbidité mesurées à 10 heures adoptent le même schéma que les moyennes journalières où la plupart des données se regroupent dans la gamme de valeur 0 – 20 NTU, voire 15 NTU pour les turbidités des séries échantillonnées.

5.2.3 Distribution dans le temps

Pour vérifier la distribution des données dans le temps, celles-ci ont été tracées sur une ligne du temps selon leur disponibilité (figure 17).

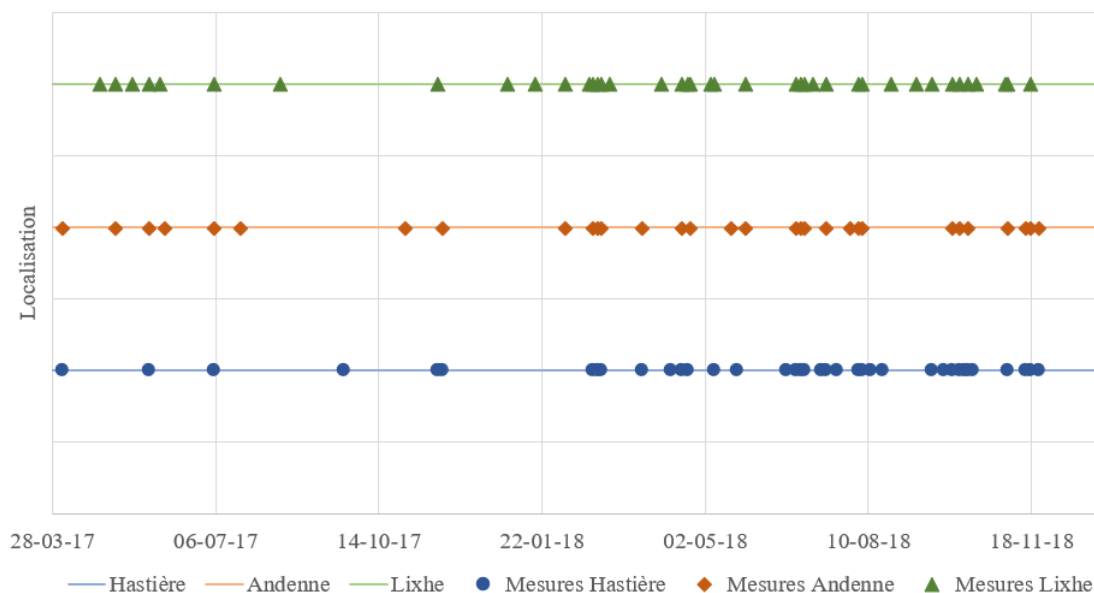


FIGURE 17 – Répartitions dans le temps des données des variables d'étalonnage des modèles.

Le graphique montre en effet une répartition non uniforme sur la période de disponibilité des images satellites. Cependant, une différence s'observe pour l'année 2017 où il y a plusieurs coupures pour les périodes de juillet à octobre, tandis que les images sont réparties de façon plus uniforme pour l'année 2018 mais forment des amas. Les saisons devraient être plus ou moins bien représentées pour l'année 2018. De plus, la répartition des données est cohérente entre les différentes stations.

5.2.4 Normalisation

Les graphiques des résidus et de la normalité se trouvant en annexe C indiquent une nette amélioration après transformation logarithmique de la turbidité que ce soit pour la bande B4 qui représente la réflectance, pour le débit ou pour les précipitations. Par ailleurs, la transformation supplémentaire du débit en logarithme indique aussi de meilleurs résultats que sans la transformation.

5.3 Analyse en composantes principales

5.3.1 Bandes spectrales sélectionnées

Les figures 18, 19, 20 et 21 montrent une certaine tendance générale à s'organiser en sous-groupes de bandes plus corrélées entre elles. Les bandes du visible 2, 3 et 4 forment un sous-groupe très soudé et sont relativement bien corrélées avec les bandes 1 et 5 étant également très corrélées entre elles. Ces cinq bandes sont les plus proches de la turbidité. Les bandes 9, 11 et 12 apparaissent quasiment orthogonales à la turbidité, effet légèrement moins visible pour la station de Lixhe cependant, indiquant l'absence de corrélation avec la turbidité.

Le groupe des bandes 6 à 8A quant à elles ont des tendances plus nuancées mais forment quasiment un angle droit avec la turbidité.

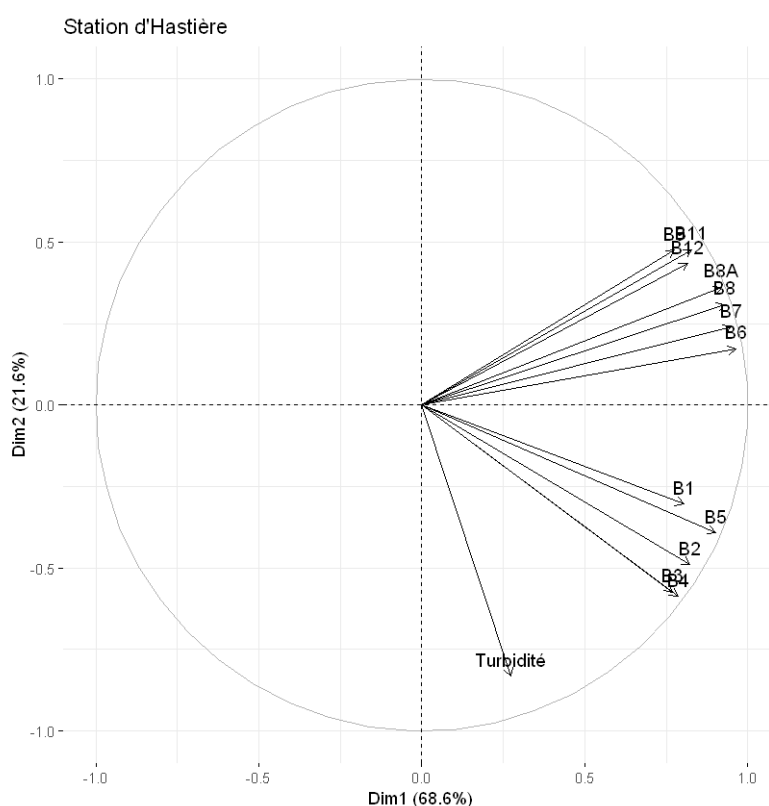


FIGURE 18 – Cercle de corrélation des bandes spectrales pour la station d'Hastière.

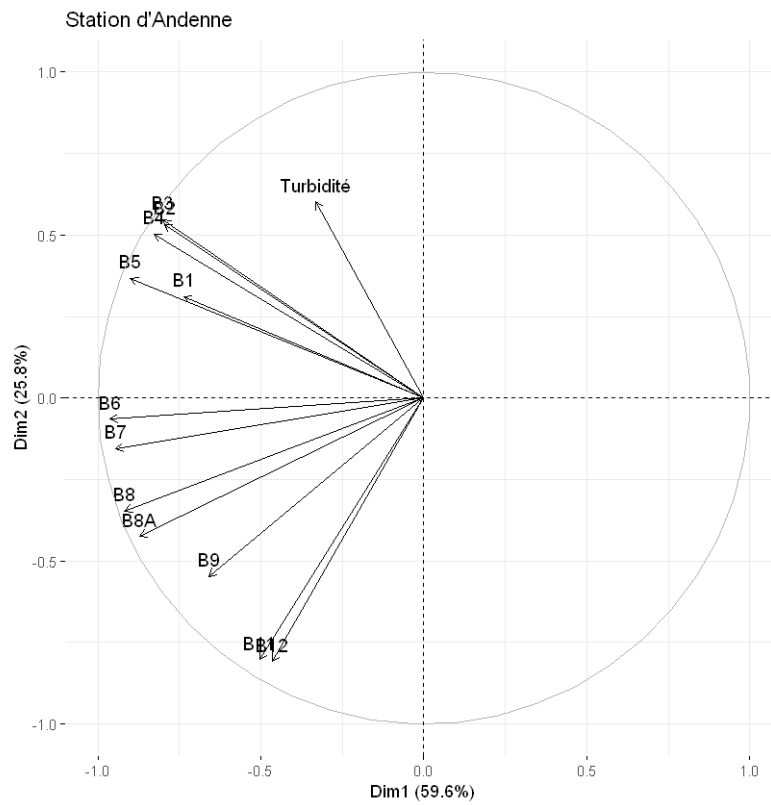


FIGURE 19 – Cercle de corrélation des bandes spectrales pour la station d'Andenne.

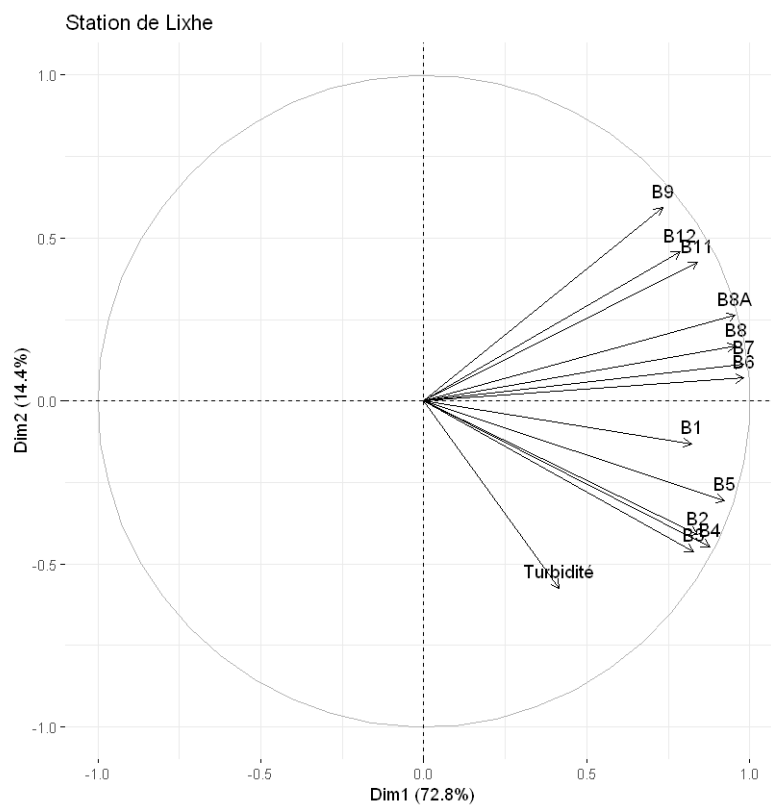


FIGURE 20 – Cercle de corrélation des bandes spectrales pour la station de Lixhe.

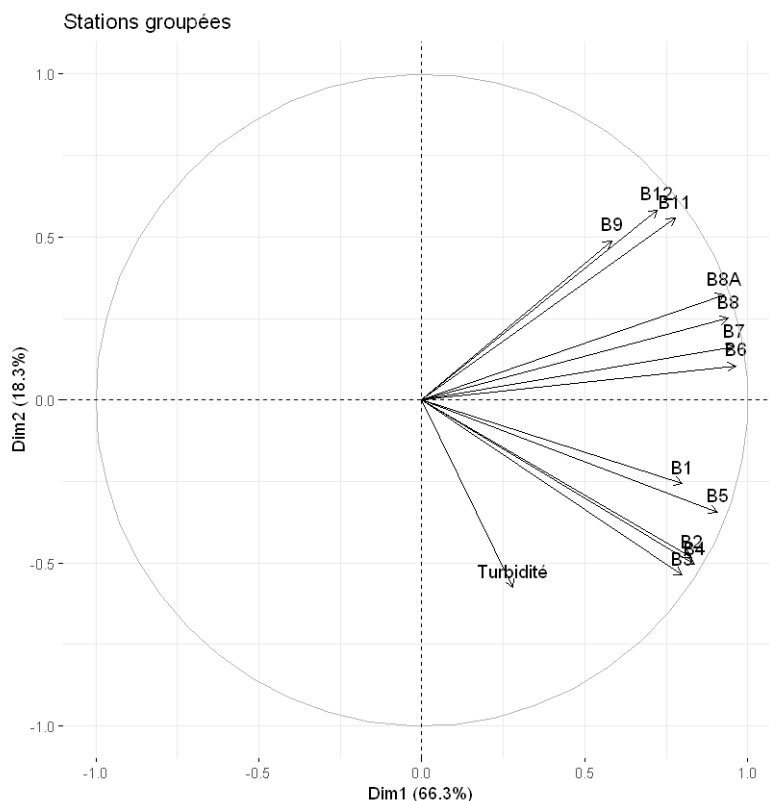


FIGURE 21 – Cercle de corrélation des bandes spectrales pour les stations groupées.

Étant donné l'allure des graphiques et avec appui de la littérature attestant de l'utilisation des bandes du visible pour des valeurs de turbidité variant entre 0 et 15 NTU (CHAWLA et al., 2020), les bandes 1 à 5 sont donc gardées.

Par ailleurs, on constate que les bandes 1 à 5 faisant partie de la région visible du spectre (443 à 705 nm) sont très corrélées entre elles. La deuxième ACP permettra de créer de nouvelles variables indépendantes entre elles.

5.3.2 Construction des nouvelles variables prédictives

Les tableaux 10, 11 et 12 ci-dessous reprennent les différents facteurs des composantes principales provenant de la deuxième ACP pour les scénarios 1, 5 et 6. Pour rappel, le choix du nombre de composantes principales à garder s'est basé sur les valeurs propres qui, combinées, contribuent à au moins 95% de la variance.

Pour le scénario 1, c'est en prenant en compte la CP1 et CP2 pour chaque localisation que les 95% étaient dépassés, avec 96,7%, 97,1%, 96,2% et 96,3% respectivement pour Hastière, Andenne, Lixhe et les trois stations groupées.

Pour le scénario 5, les trois premières CP's pour chaque cas contribuaient pour 97,5% pour Hastière, 98% pour Andenne, 97,6% pour Lixhe et 97% pour toutes les stations.

Enfin, les pourcentages étaient de 97,9%, 98,4%, 98% et 97,4% respectivement pour Hastière, Andenne, Lixhe et les stations groupées pour le scénario 6.

Scénario 1

TABLEAU 10 – *Facteur des composantes principales retenues pour le scénario 1.*

Scénario 1	Hastière		Andenne	
Variables	CP1	CP2	CP1	CP2
Eigenvalue [%]	90	6,7	89	8,1
B1	0,4	0,89	-0,39	0,86
B2	0,46	-0,06	-0,46	0,08
B3	0,46	-0,32	-0,47	-0,16
B4	0,46	-0,29	-0,46	-0,34
B5	0,46	-0,11	-0,45	-0,33
	Lixhe		Trois stations	
Variables	CP1	CP2	CP1	CP2
Eigenvalue [%]	88,8	7,4	89,7	6,6
B1	-0,41	0,80	0,41	0,85
B2	-0,46	0,18	0,46	0,08
B3	-0,46	-0,12	0,46	-0,14
B4	-0,46	-0,39	0,46	-0,37
B5	-0,45	-0,40	0,45	-0,35

Scénario 5

TABLEAU 11 – *Facteur des composantes principales retenues pour le scénario 5.*

Scénario 5	Hastière			Andenne		
Variables	CP1	CP2	CP3	CP1	CP2	CP3
Eigenvalue [%]	81,3	11,4	4,8	83,8	10,6	3,6
B1	0,38	0,40	-0,82	-0,35	0,71	0,44
B2	0,44	0,20	0,19	-0,43	0,15	0,12
B3	0,44	-0,01	0,32	-0,44	-0,04	0,00
B4	0,44	-0,04	0,26	-0,44	-0,11	-0,34
B5	0,43	0,10	0,15	-0,43	-0,02	-0,60
Log débit	0,30	-0,89	-0,32	-0,35	-0,68	0,56
	Lixhe			Trois stations		
Variables	CP1	CP2	CP3	CP1	CP2	CP3
Eigenvalue [%]	82,3	10,3	5	79,9	13,2	3,9
B1	-0,37	0,59	0,50	0,38	0,47	-0,68
B2	-0,44	0,15	0,23	0,44	0,13	-0,10
B3	-0,44	-0,03	0,01	0,44	0,03	0,08
B4	-0,44	-0,77	-0,38	0,45	-0,06	0,35
B5	-0,42	0,07	-0,62	0,43	0,03	0,51
Log débit	-0,34	-0,79	0,40	0,28	-0,87	-0,38

Scénario 6

TABLEAU 12 – *Facteur des composantes principales retenues pour le scénario 6.*

Scénario 6	Hastière				Andenne			
Variables	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Eigenvalue [%]	69,7	16,3	7,8	4,1	71,9	17	7,1	2,4
B1	0,38	0,23	-0,29	-0,83	-0,35	0,38	0,52	-0,59
B2	0,44	0,10	-0,18	0,19	-0,43	0,10	0,07	-0,04
B3	0,44	0,01	0,01	0,33	-0,44	-0,02	-0,05	0,05
B4	0,44	-0,03	0,00	0,25	-0,44	-0,14	0,03	0,30
B5	0,44	0,04	-0,11	0,14	-0,42	-0,13	0,16	0,53
Précipitations	-0,02	0,87	0,48	0,06	0,04	-0,87	0,41	-0,26
Log débit	0,30	-0,41	0,80	-0,28	-0,35	-0,24	-0,72	-0,45

	Lixhe				Trois stations			
Variables	CP1	CP2	CP3	CP4	CP1	CP2	CP3	CP4
Eigenvalue [%]	71,6	13,4	8,8	4,2	68,5	14,5	11,1	3,3
B1	-0,37	0,05	0,59	-0,51	0,38	0,08	-0,47	0,68
B2	-0,43	0,07	0,15	-0,22	0,44	0,00	-0,14	0,10
B3	-0,43	0,14	-0,03	0,01	0,44	-0,03	-0,04	-0,09
B4	-0,43	-0,01	-0,08	0,38	0,46	0,01	0,07	-0,34
B5	-0,42	0,00	0,07	0,62	0,43	0,03	-0,02	-0,50
Précipitations	-0,13	-0,98	-0,02	-0,07	0,04	0,97	0,25	0,06
Log débit	-0,33	0,07	-0,79	-0,41	0,28	-0,25	0,83	0,39

Pour chaque scénario, les CP1 sont en général la somme pondérée des cinq bandes spectrales dont le signe diffère entre les stations : positif pour Hastière et les stations groupées, négatif pour Andenne et Lixhe.

Dans le scénario 1, la deuxième CP est principalement expliquée par la bande B1 à 443 nm avec des facteurs supérieurs à 0,80.

Pour le scénario 5, les représentations des CP2 sont plus confuses mais le débit y contribue pour une grande partie suivi par la bande B1 et exceptionnellement la bande B4 pour la station de Lixhe. Les CP3 sont un mixe de toutes les variables.

Pour ce qui est du scénario 6, la première CP représente majoritairement les bandes spectrales, comme pour les autres scénarios, ainsi que le débit. Tandis que les CP2 représentent les précipitations quasiment à elles seules. Les CP3 sont bien représentée par le débit. Comme pour le scénario 5, les CP4 sont un mixe des variables.

5.4 Évaluation des modèles prédictifs

Le tableau 12 reprend les résultats des modèles de régression linéaire appliqués aux six scénarios et pour chaque station séparément et réunies. Les graphiques des régressions linéaires associées à ces résultats se trouvent en annexe D.

TABLEAU 13 – Résultats des modèles de régression linéaire.

Scénario 1	Hastière	Andenne	Lixhe	Tout	Scénario 2	Hastière	Andenne	Lixhe	Tout
R_a^2	0,38	0,38	0,59	0,34	R_a^2	0,68	0,49	0,31	0,26
n	34	28	37	105	n	35	29	38	106
RMSE	0,18	0,36	0,21	0,31	RMSE	0,30	0,34	0,28	0,33
BIC	-7,78	38,87	4,55	74,14	BIC	-34,63	30,50	22,48	83,69
Scénario 3	Hastière	Andenne	Lixhe	Tout	Scénario 4	Hastière	Andenne	Lixhe	Tout
R_a^2	-0,03	0,13	0,11	0,05	R_a^2	0,69	0,61	0,37	0,32
n	35	29	38	106	n	34	28	37	105
RMSE	0,23	0,44	0,32	0,38	RMSE	0,13	0,29	0,26	0,32
BIC	8,36	47,23	32,40	110,02	BIC	-33,62	24,36	21,33	78,25
Scénario 5	Hastière	Andenne	Lixhe	Tout	Scénario 6	Hastière	Andenne	Lixhe	Tout
R_a^2	0,59	0,40	0,59	0,35	R_a^2	0,58	0,46	0,61	0,40
n	33	27	36	104	n	32	26	35	103
RMSE	0,14	0,35	0,21	0,31	RMSE	0,14	0,33	0,20	0,30
BIC	-20,68	40,45	6,56	76,68	BIC	-17,24	39,20	7,17	71,67

n = nombre de degré de liberté du modèle.

Pour le scénario 1 où les bandes 1 à 5 ont été implémentées dans une ACP, le meilleur résultat vient de Lixhe avec un R_a^2 de 0,56 et une RMSE de 0,21. Les autres R_a^2 ne dépassent pas 0,38 pour des RMSE plus élevées.

Pour le scénario 2 ne comprenant que le logarithme du débit, c'est la station d'Hastière qui montre les meilleurs résultats avec un R_a^2 de 0,68. Cependant, les RMSE sont globalement plus élevés que pour le scénario 1.

Ce sont les précipitations cumulées sur cinq jours antérieurs à la date de prise de mesure qui obtiennent les plus mauvais résultats avec des R_a^2 ne dépassant pas 0,13 et même un R_a^2 de -0,03 pour la station d'Hastière.

Le scénario 4 incluant le débit en plus des précipitations fournit des résultats prometteurs pour les stations d'Hastière et Andenne avec une RMSE de seulement 0,13 pour Hastière. De plus, la combinaison des précipitations et du débit a permis de légèrement améliorer les résultats des modèles ne prenant en compte que l'une ou l'autre, bien que pour Hastière l'effet est très très faible.

Les bandes spectrales 1 à 5 et le logarithme du débit du scénario 5 avaient été implémentés dans une ACP. Les résultats montrent une très légère amélioration des performances des modèles par rapport à l'utilisation des bandes seules, sauf pour la station de Lixhe où aucun changement

n'est observé. A l'inverse, l'ACP a diminué les valeurs des R_a^2 en comparaison aux scénarios 2 et 4 pour les stations d'Hastière et Andenne sans pour autant diminuer les RMSE.

Le scénario 6 impliquant l'ensemble des variables prédictives rapporte les meilleurs résultats pour la station de Lixhe et pour le modèle généralisé avec des R_a^2 de 0,61 et 0,40 respectivement. Les résultats d'Hastière et Andenne ont cependant diminués par rapport aux scénarios incluant moins de variables comme le scénario 4, mais ont augmenté par rapport au scénario 5 ayant également implémenté l'ACP.

En ce qui concerne les BIC, ce sont les scénarios 2 et 4 qui présentent les meilleurs résultats pour stations d'Hastière et Andenne avec les BIC les plus faibles. Lixhe a son BIC le plus bas dans les scénarios 1 et 5, tandis que pour le modèle généralisé, le meilleur BIC se situe au niveau du scénario 6.

En résumé, les variables prédisant le mieux la turbidité sont le débit et les précipitations pour les stations d'Hastière et d'Andenne, et la surface de réflectance pour la station de Lixhe et pour les trois stations réunies. Les modèles montrent des résultats très variables entre les différents scénarios mais également pour les stations entre elles. Globalement, ce sont les modèles incluant le débit qui obtiennent les meilleurs résultats de l'ensemble des scénarios. De façon générale, les modèles généralisés à l'ensemble de la rivière montrent des résultats médiocres par rapport aux modèles spécifiques aux stations.

6 Discussion

Potentiel d'utilisation des images Sentinel-2 de niveau 2A pour construire un modèle prédictif de la turbidité des rivières

Le premier objectif de ce mémoire était d'évaluer les performances des images Sentinel-2 de niveau 2A.

Les corrections des masques de nuages QA60 et SLC appliqués sur les images déjà corrigées par Sen2Cor ont un fonctionnement satisfaisant lorsqu'ils sont utilisés conjointement. L'usage seul du masque QA60 ne pourrait être recommandé en regard du manque de détection des ombres et des nuages (voir figure 11 de la méthodologie). Son utilisation nécessite donc une attention particulière (YANG et al., 2020). Toutefois, le masque SLC devrait se suffire à lui-même, ses performances furent déjà prouvées dans l'article de MAIN-KNORN et al. (2017).

Contrairement aux attentes basées sur la littérature pour les cas d'étude mettant en jeux les images Sentinel-2, les modèles se basant uniquement sur la réflectance de surface (scénario 1) n'ont pas révélé de bons résultats, avec des R^2 entre 0,34 et 0,59. Bien que quelques études ont démontré le potentiel des images Sentinel-2 de niveau 1C pour l'observation des paramètres de la qualité de l'eau (BONANSEA et al., 2019 ; HUOVINEN et al., 2019 ; KUHN et al., 2019 ; SÒRIA-PERPINYÀ et al., 2020), peu d'entre elles ont évalué le potentiel des images de niveau 2A corrigées atmosphériquement par Sen2Cor.

La première raison qui expliquerait les résultats tangents de cette étude pourrait provenir de la reconstruction du spectre à partir des composantes principales. L'un des inconvénients de l'ACP est la perte de contrôle sur l'endroit où vont les informations intéressantes, c'est-à-dire dans quelles composantes principales. La pondération de l'information de certaines bandes dans les composantes principales pourrait également influencer les résultats et réduire l'information importante. Ceci est par exemple le cas pour la bande rouge B4 que l'on sait sensible aux variations de la turbidité.

Une autre raison serait la résolution radiométrique du capteur MSI, c'est-à-dire un rapport signal/bruit (SNR) trop faible pour l'observation de la turbidité dans les rivières. Les eaux intérieures turbides présentent des charges sédimentaires qui contribuent à la réflectance dans le vert et le rouge. Un faible SNR entraînera donc souvent une erreur relative élevée dans les récupérations de la réflectance aux longueurs d'onde courtes, comme la bande B1 à 443 nm. En effet, une absorption élevée par la chlorophylle et les matières organiques dissoutes font baisser le signal de sortie de l'eau dans le bleu. De plus, Sen2Cor n'a pas été conçu pour les applications aquatiques. Les effets des interférences atmosphériques et le manque de précision de ses corrections peuvent alors avoir un impact significatif sur l'estimation des constituants de l'eau. En outre, les pixels terrestres plus brillants adjacents la zone d'étude produisent une lumière parasite augmentant la radiance du capteur jusqu'à 50 % dans les bandes du proche infrarouge. Cela contribue alors à la sur-correction des effets atmosphériques pour les bandes visibles au-dessus de petites masses d'eau intérieures.

Par ailleurs, l'article de MONTANHER et al. (2014) explique les sources d'erreurs potentielles pouvant provenir des différences dans la composition des sédiments et la distribution de la taille des grains. Ces sources d'erreurs s'appliquant entre les images prises à différentes dates. Ces caractéristiques influencent fortement les propriétés optiques du milieu aquatique, notamment la forme et l'intensité du spectre de réflectance. De plus, la variation saisonnière du type de sédiments peut diminuer la capacité de prédiction des modèles empiriques, même lorsqu'ils sont développés au niveau régional, en fonction de l'environnement du bassin de drainage. En découle également une variation de l'aspect des masses d'eau, leur contenu influençant la variation de couleur observée sur les images Sentinel-2.

Variable(s) utile(s) pour prédire la variabilité observée dans le signal de turbidité

Le deuxième objectif de cette étude visait à déterminer laquelle des variables entre la réflectance de surface, le débit et les précipitations prédisait le mieux la turbidité.

Représentativité des jeux de données

Le nombre de valeurs pour les jeux de données s'est vu nettement réduit. Cela n'est pas un problème majeur dans le cas où l'ensemble de données est adéquat pour identifier un modèle. En d'autres termes, les jeux de données doivent consister en un nombre approprié d'échantillons de turbidité et de variables explicatives collectées dans toute la gamme des conditions hydrologiques observées pour la période d'enregistrement. Un modèle de régression élaboré à partir de 15 échantillons plus ou moins uniformément répartis sur les saisons et la plage de turbidité d'un site peut aboutir à un modèle plus représentatif qu'un modèle élaboré à partir d'un ensemble de données de 50 échantillons où la distribution des valeurs ne définit qu'une période ou une plage limitée des conditions sédimentologiques (RASMUSSEN et al., 2009).

Dans ce cas-ci, la plupart des valeurs de turbidité se trouvait dans une gamme de 0 à 20 NTU. Cela réduit nettement la gamme de valeurs mesurée par le turbidimètre (0-100 NTU) ainsi que celle de la turbidité pouvant aller jusqu'à 400 NTU (ORGANISATION INTERNATIONALE DE NORMALISATION, 2016). Par conséquent, les jeux de données ne deviennent plus représentatifs de toute la réalité. De plus, suite à la distribution inégale dans le temps pour 2017, celle-ci devrait être moins représentative de la variation saisonnière. En outre, les valeurs élevées de turbidité, de débit et des précipitations furent écartées suite à la réduction des jeux de données. Elles ne sont donc pas représentées ici. Il est donc possible qu'il y ait un lien entre les valeurs élevées de ces mesures in situ indiquant que les précipitations seraient alors plus fortement corrélées avec la turbidité lors de fortes précipitations. Cette hypothèse reste cependant à vérifier.

Performances des variables de prédiction

Globalement, les meilleurs résultats sont obtenus dans les modèles employant le débit. Cela fait de ce dernier la meilleure variable prédictive de la turbidité dans le cas d'étude de la Meuse.

Parmi les scénarios comprenant un seul type de variable prédictive, celui avec le débit (scénario 3) montre les meilleurs résultats. Ceci a été mis en évidence surtout pour la station de Lixhe et lorsque les trois stations sont groupées. Même en étant employé seul, le débit permet d'avoir des résultats aussi bons que lorsqu'il est employé dans une ACP avec la réflectance (scénario 5) ou les précipitations (scénario 4). Cela pourrait s'expliquer par le fait que le débit est une mesure in situ, contrairement aux mesures de réflectance, et donc que le débit suivrait les variations de turbidité de façon beaucoup plus rapprochée que pour la réflectance. Cette hypothèse devra également être vérifiée. Par ailleurs, une diminution de la qualité de représentation est observée lorsque le débit est employé dans une ACP. Cela pourrait provenir du fait que lors d'une ACP, l'entièreté de la variance expliquée par le débit n'y est pas représentée, ce dernier étant pondéré par un facteur dans la combinaison linéaire des CP's.

Outre cela, le débit n'ayant pas été mesuré aux mêmes endroits pour les mesures de turbidité à Hastière et Andenne, les résultats sont malgré tout meilleurs que pour la station de Lixhe. Le débit ne doit donc pas être mesuré à proximité immédiate des stations de turbidité et peut être généralisé à une portion plus grande que son lieu de mesure.

À l'inverse, les modèles se basant sur les précipitations seules ont des performances médiocres (scénario 3). Les précipitations n'améliorent que très peu la qualité de prédiction des modèles où elles sont employées en plus d'une autre variable (scénario 4 et 6). Cela pourrait s'expliquer par le fait que les séries échantillonnées pour l'étalonnage des modèles ne couvrent pas une gamme suffisante de valeurs des précipitations élevées. Pour une grande partie des données d'étalonnage des modèles, les précipitations cumulées sont faibles, voire nulles, étant donné qu'elles correspondent aux images Sentinel-2 prises par temps sec et dégagé. Les précipitations pourraient être une variable prédictive adéquate de la turbidité, à condition d'avoir des mesures qui couvrent l'ensemble de la gamme de valeurs de ces précipitations, et pour une période plus longue. En effet, plusieurs études ont prouvé que le cumul des précipitations était bien corrélé avec les sédiments en suspension (CHEN et al., 2019; IONUȚ et al., 2014; ZHANG et al., 2004), mais ces corrélations n'étaient pas génériques. Aussi, il est possible que la relation entre précipitations et turbidité doive être adaptée suivant le cas d'étude.

En ce qui concerne la réflectance de surface, les résultats sont relativement insatisfaisants par rapport à ceux trouvés dans la littérature (voir tableau 2). Néanmoins, bien qu'employée seule, la réflectance permet d'obtenir de meilleurs résultats que le débit seul pour les stations d'Andenne et Hastière.

Par ailleurs, l'analyse des graphiques des résidus et de la normalité (figures en annexe C) ont montré des résultats favorables de la transformation logarithmique de la turbidité, tout comme pour le débit. Il serait cependant intéressant d'essayer d'autres transformations, ainsi que pour les autres variables, ou alors de justifier plus rigoureusement que le logarithme est la transformation la plus appropriée pour ce genre de cas d'étude.

Performances relatives des modèles de prédiction

On remarque une très forte variation entre les résultats pour les différentes stations. De manière globale, les modèles généralisés aux trois stations obtiennent les moins bons résultats. Il n'était pas possible d'avoir un bon modèle généralisé à l'ensemble de la rivière au vu des performances des modèles spécifiques. Il est donc difficile de trouver une corrélation entre les stations ; ce qui indique que les processus contrôlant la turbidité pourraient être locaux plutôt que régionaux. La longueur d'onde dominante réfléchie change avec les constituants présents dans l'eau. Cependant, le fait d'avoir sélectionné les bandes 1 à 5 devrait permettre que les variations dans les constituants de la turbidité n'impliquent pas trop la variation du signal détecté pour les différents endroits.

Comme déjà mentionné par d'autres études (SAGAN et al., 2020), les modèles empiriques sont généralement considérés comme spécifiques à un site. Très peu d'études démontrent que les modèles peuvent être transférables à l'échelle régionale.

On constate également que suivant la station, l'amélioration de la qualité de représentation du modèle ne dépend pas toujours de la même variable. Les stations d'Hastière et d'Andenne voient leurs résultats améliorés en utilisant le débit comme variable prédictive. Pour la station de Lixhe et les stations groupées, les résultats sont améliorés grâce à la réflectance de surface (scénario 1 vs scénarios 2 et 5). La turbidité étant l'association de plusieurs types de particules, il est possible que ces associations varient d'une station à une autre selon le type d'occupation du sol ou des activités anthropiques à proximité de la rivière. En effet, Lixhe se trouve dans un environnement très artificialisé comparé à Andenne et Hastière.

Analyses en composantes principales

La première ACP a montré que les bandes les mieux corrélées à la turbidité étaient les bandes 1 à 5, incluant les longueurs d'onde de 442 à 704 nm. Ceci confirme les résultats de CHOUBEY (1992) stipulant que les courtes longueurs d'ondes, plus particulièrement la bande allant de 620 à 680 nm, étaient plus utiles pour quantifier la turbidité que la partie du spectre avec des longueurs d'ondes plus grandes. RITCHIE et al. (2003) ont également montré que l'eau à faible concentration en sédiments a une réflexion maximale de 550 nm, contre 600 nm pour l'eau à forte concentration en sédiments. Au-delà du domaine visible, la plus grande partie de l'énergie est absorbée par la masse d'eau.

En regardant les facteurs des CP's pour la deuxième ACP, on remarque qu'il n'est pas possible de réduire trop leur nombre. L'article de DOERFFER et al. (1989) suggérait que la première CP puisse refléter le comportement d'une variable telle que la turbidité ou la matière inorganique en suspension. Ces derniers réfléchissent les longueurs d'onde dans une gamme plus générale qu'un constituant particulier seul. Or, on constate ici, que seule la CP1 n'est pas suffisante pour expliquer la majeure partie de la variance.

D'après les études de FLINK et al. (2001) et DOERFFER et al. (1989), la façon la plus courante d'utiliser les principales composantes est probablement d'examiner les vecteurs propres afin

d'y associer des interprétations physiques. En général, les premières CP's de chaque scénario pourraient être liées à la turbidité, les CP's représentant une pondération des cinq bandes. Inversement, la CP2 est principalement représentée dans la bande B1 (scénario 1), le débit (scénario 2) ou les précipitations (scénario 3). Une étude plus poussée de ces CP's pourrait mettre en évidence le contraste entre ceux-ci.

On peut par ailleurs noter que les facteurs des vecteurs propres des différentes stations pour un même scénario sont assez similaires, ce qui indique que l'ACP ne peut être basée que sur une seule station, et que les vecteurs propres peuvent toujours être utilisés pour approximer les spectres des autres stations.

La qualité de représentation des CP's par rapport aux variables originelles reste à prouver. De plus, dans ce cas d'étude, son utilisation semble diminuer les performances des modèles. Toutefois, un avantage de l'ACP est la construction de variables non corrélées entre elles. Celles-ci sont alors mieux adaptées à l'analyse de régression en ce sens qu'elles donnent des estimations de paramètres plus stables. En effet, en basant la régression sur les CP's plutôt que sur les bandes spectrales, on pourrait construire des modèles plus fiables. Il est d'ailleurs souvent recommandé dans la littérature statistique d'utiliser des régressions sur les variables non corrélées (FLINK et al., 2001).

Perspectives

Ce travail constitue un début en termes de surveillance de la turbidité grâce aux images Sentinel-2 de niveau 2A. D'autres analyses plus poussées permettront de peaufiner les relations entre les différentes variables de prédiction et la turbidité.

L'ajout des mesures de réflectance pour l'année 2019 permettrait d'étoffer les bases de données et la représentativité de celles-ci. Des jeux de données avec un nombre plus important de mesures pourraient aussi améliorer la compréhension des processus entre les différentes variables explicatives. Il serait aussi intéressant de tester différents ratios de bandes spectrales en tant que variables prédictives. Plusieurs études ont prouvé leur utilité en tant que correction du bruit dû aux effets atmosphériques et aux capteurs des satellites. En outre, la combinaison des données de plusieurs satellites permettrait d'améliorer les lacunes en termes de disponibilité des données. Par exemple Landsat-8, dont les données sont également disponibles sur GEE, comporte des capteurs ayant des caractéristiques techniques similaires à de ceux de Sentinel-2.

Malgré que l'étude de WARREN et al. (2019) montre des résultats prometteurs pour Sen2Cor, il serait approprié d'essayer d'autres routines de correction dans le but de tester leurs performances en comparaison avec Sen2Cor.

La relation entre la réflectance et la turbidité de la Meuse n'étant pas encore très précise, il pourrait être pertinent d'analyser de façon plus poussée les constituants de la turbidité aux différentes stations. Cela dans le but de déceler une éventuelle différence de composition des constituants la turbidité. En effet, Hastière et Andenne se situent dans des milieux plus ruraux par rapport à Lixhe qui se trouve dans un milieu très artificialisé. De plus, cela pourrait

également confirmer que l'exclusion des bandes spectrales dans la gamme du proche infrarouge était un bon choix.

Ensuite, il serait intéressant d'approfondir les relations entre les différentes variables (précipitations/turbidité, débit/turbidité) et d'explorer d'autres pistes de transformation. Il est également suggéré de tester différents modèles de prédictions de la turbidité, linéaires et non linéaires, comme dans l'étude de (CHOUBEY, 1992). L'article de SAGAN et al. (2020) propose également d'autres méthodes que les modèles empiriques pour l'observation des paramètres de la qualité de l'eau.

Enfin, une dernière perspective serait d'élargir les recherches à d'autres paramètres de la qualité de l'eau. Il est aujourd'hui possible d'estimer indirectement des constituants de l'eau non-optiquement actifs.

7 Conclusion

L'objectif général de ce travail était d'évaluer le potentiel de l'imagerie Sentinel-2, individuellement ou en combinaison avec le débit et les précipitations, pour la surveillance de la turbidité de l'eau des rivières. Le cas d'étude était le segment wallon de la Meuse le long duquel les données ont été recueillies à trois stations de mesure différentes. La construction de modèles empiriques, spécifiques aux stations et généralisés à l'ensemble de la rivière, ont permis d'évaluer les différents jeux de données, aussi bien au niveau des données in situ et satellites, qu'au niveau de la localisation des stations.

La première partie, consacrée à l'élaboration des jeux de données, a prouvé qu'il était relativement facile d'extraire l'information d'images satellites, et ce, grâce au traitement des images et l'extraction des données de réflectance réalisé sur Google Earth Engine. Ce dernier fournit des outils puissants pour les tâches de traitement des images et autres données de télédétection. Malgré cela, il est dommage de constater que suite à la couverture nuageuse au-dessus du site d'étude, peu d'images étaient en réalité utilisables. Pour la période d'échantillonnage des années 2017 et 2018, seulement une trentaine d'images par station étaient d'assez bonne qualité par rapport aux 180 initiales.

Les résultats des modèles pour les différentes stations ainsi que pour l'ensemble des données ne sont pas très concluants. Force est de constater qu'un modèle généralisé à l'ensemble de la rivière n'est pas envisageable avec la méthode employée. Les méthodes empiriques ont cependant l'avantage d'être faciles à mettre en oeuvre et simples à traiter. Toutefois, comme les modèles sont dérivés de données qui ne peuvent pas représenter pleinement les variations naturelles, leurs performances sont intrinsèquement soumises aux conditions de l'eau et aux plages de variables à partir desquelles les données ont été obtenues pour le développement des algorithmes. L'ensemble de données utilisé pour la formation des modèles empiriques est d'une importance capitale et doit être soigneusement examiné en fonction de l'application.

Le débit, la profondeur et le transport des constituants de l'eau des systèmes fluviaux ont une incidence sur la réflectance des capteurs. Par conséquent, il est difficile de généraliser les modèles pour ces environnements. Il est cependant possible de généraliser les algorithmes pour les systèmes fluviaux en tenant compte des variations saisonnières et annuelles de la vitesse et du volume de l'écoulement. Certaines études sont actuellement en cours pour généraliser les modèles pour divers environnements aquatiques grâce à l'utilisation des données spatiales et temporelles à haute résolution provenant de satellites, notamment Landsat-8 et Sentinel-2.

Par ailleurs, au vu de la gamme de valeurs de la turbidité, les modèles étalonnés sur la turbidité de la Meuse ne pourront pas être appliqués à des cas d'étude dont la turbidité dépasse 20 NTU.

L'utilisation de l'analyse en composantes principales dans ce cas d'étude n'a pas montré des résultats satisfaisants. L'un des inconvénients de l'ACP est la perte de contrôle sur l'endroit où vont les informations intéressantes. Malgré cela, l'ACP construit des variables indépendantes,

qui sont mieux adaptées à l'analyse de régression en ce sens qu'elles donnent des estimations de paramètres plus stables.

Ce travail constitue un bon début en termes d'utilisation des images Sentinel-2 pour des applications aquatiques. L'un des avantages est son faible temps de revisite permettant de surveiller efficacement des événements spécifiques plutôt que de détecter des changements fréquents. Malgré tout, la couverture nuageuse limite sérieusement la couverture des produits satellites. D'autre part, l'absence de bandes étroites chez Sentinel-2 et son faible rapport signal/bruit posent encore quelques difficultés lors de l'observation des paramètres de la qualité des eaux des rivières.

Les satellites à résolution modérée comme Sentinel-2 sont souvent choisis pour les eaux intérieures mais ont une résolution spectrale qui leur font défaut. La capacité à observer les rivières depuis l'espace nécessite une résolution beaucoup plus élevée que celle des lacs. Finalement, et comme indiqué dans d'autres études, la correction atmosphérique au-dessus des eaux intérieures n'est également pas encore résolue et peut générer des erreurs importantes.

Références

- BAILEY, S. W. & WERDELL, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, 102(1), 12-23. doi :10.1016/j.rse.2006.01.015
- BONANSEA, M., LEDESMA, M., BAZÁN, R., FERRAL, A., GERMAN, A., O'MILL, P., RODRIGUEZ, C. & PINOTTI, L. (2019). Evaluating the feasibility of using Sentinel-2 imagery for water clarity assessment in a reservoir. *Journal of South American Earth Sciences*, 95, 102265. doi :10.1016/j.jsames.2019.102265
- BUSTAMANTE, J., PACIOS, F., DÍAZ-DELGADO, R. & ARAGONÉS, D. (2009). Predictive models of turbidity and water depth in the Doñana marshes using Landsat TM and ETM+ images. *Journal of Environmental Management. The GlobWetland Symposium : Looking at wetlands from space*, 90(7), 2219-2225. doi :10.1016/j.jenvman.2007.08.021
- CHAWLA, I., KARTHIKEYAN, L. & MISHRA, A. K. (2020). A review of remote sensing applications for water security : Quantity, quality, and extremes. *Journal of Hydrology*, 585, 124826. doi :10.1016/j.jhydrol.2020.124826
- CHEN, J. & CHANG, H. (2019). Dynamics of wet-season turbidity in relation to precipitation, discharge, and land cover in three urbanizing watersheds, Oregon. *River Research and Applications*, 35(7), 892-904. doi :10.1002/rra.3487
- CHOUBEY, V. (1992). Correlation of turbidity with Indian remote sensing satellite-1a data. *Hydrological Sciences Journal*, 37(2), 129-140. doi :10.1080/02626669209492573
- D'ANDRIMONT, R. (2017). *Towards consistent inland water body mapping across space and time from optical Earth observation systems* (thèse de doct., UCL - Université Catholique de Louvain). Récupérée 19 mars 2020, à partir de <https://dial.uclouvain.be/pr/boreal/object/boreal:186332>
- DE KEUKELAERE, L., STERCKX, S., ADRIAENSEN, S., KNAEPS, E., REUSEN, I., GIARDINO, C., BRESCIANI, M., HUNTER, P., NEIL, C., VAN DER ZANDE, D. & VAICIUTE, D. (2018). Atmospheric correction of Landsat-8/OLI and Sentinel-2/MSI data using iCOR algorithm : validation for coastal and inland waters. *European Journal of Remote Sensing*, 51(1), 525-542. doi :10.1080/22797254.2018.1457937
- DEVELOPERS, G. (2020). Statistics of an Image Region | Google Earth Engine. Library Catalog : [developers.google.com](https://developers.google.com/earth-engine/reducers_reduce_region). Récupérée 4 août 2020, à partir de https://developers.google.com/earth-engine/reducers_reduce_region
- DOERFFER, R., FISCHER, J., STÖSSEL, M., BROCKMANN, C. & GRASSL, H. (1989). Analysis of thematic mapper data for studying the suspended matter distribution in the coastal area of the German Bight (North Sea). *Remote Sensing of Environment*, 28, 61-73. doi :10.1016/0034-4257(89)90105-3
- DOGLIOTTI, A. I., RUDDICK, K. G., NECHAD, B., DOXARAN, D. & KNAEPS, E. (2015). A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and

- estuarine waters. *Remote Sensing of Environment*, 156, 157-168. doi :10.1016/j.rse.2014.09.020
- DÖRNHÖFER, K. & OPPELT, N. (2016). Remote sensing for lake research and monitoring – Recent advances. *Ecological Indicators*, 64, 105-122. doi :10.1016/j.ecolind.2015.12.009
- ESA. (2020). Sen2Cor. Récupérée 15 août 2020, à partir de <http://step.esa.int/main/third-party-plugins-2/sen2cor/>
- FLINK, P., LINDELL, L. T. & ÖSTLUND, C. (2001). Statistical analysis of hyperspectral data from two Swedish lakes. *Science of The Total Environment*. Lake water monitoring in, 268(1), 155-169. doi :10.1016/S0048-9697(00)00686-0
- FRANZ, B. A., BAILEY, S. W., KURING, N. & WERDELL, P. J. (2015). Ocean color measurements with the Operational Land Imager on Landsat-8 : implementation and evaluation in SeaDAS. *Journal of Applied Remote Sensing*, 9(1), 096070. Publisher : International Society for Optics and Photonics. doi :10.1117/1.JRS.9.096070
- GEE. (2020). Sentinel-2 MSI : MultiSpectral Instrument, Level-2A. Library Catalog : developers.google.com. Récupérée 31 juillet 2020, à partir de https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR
- GHOZIZADEH, M. H., MELESSE, A. M. & REDDI, L. (2016). A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques. *Sensors (Basel, Switzerland)*, 16(8). doi :10.3390/s16081298
- GOODIN, D., HARRINGTON, J., Jr., NELLIS, M. & RUNDQUIST, D. (1996). Mapping reservoir turbidity patterns using spot-hrv data. *Geocarto International*, 11(4), 71-78. doi :10.1080/10106049609354563
- HACH. (2020). Turbidimètre Ultraturb sc | Hach France - Aperçu | Hach. Récupérée 31 juillet 2020, à partir de <https://fr.hach.com/analyseurs-de-turbidite/turbidimetre-ultraturb-sc/family?productCategoryId=25113990775>
- HARMEL, T., CHAMI, M., TORMOS, T., REYNAUD, N. & DANIS, P.-A. (2018). Sun glint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. *Remote Sensing of Environment*, 204, 308-321. doi :10.1016/j.rse.2017.10.022
- HELSEL, D. R., HIRSCH, R. M., RYBERG, K. R., ARCHFIELD, S. A. & GILROY, E. J. (2002). *Statistical methods in water resources* (USGS Numbered Series N° 4-A3). U.S. Geological Survey. Code Number : 4-A3 Code : Statistical methods in water resources Publication Title : Statistical methods in water resources Reporter : Statistical methods in water resources Series : Techniques and Methods IP-089727. doi :10.3133/tm4A3
- HICKS, B. J., STICHBURY, G. A., BRABYN, L. K., ALLAN, M. G. & ASHRAF, S. (2013). Hindcasting water clarity from Landsat satellite images of unmonitored shallow lakes in the Waikato region, New Zealand. *Environmental Monitoring and Assessment*, 185(9), 7245-7261. doi :10.1007/s10661-013-3098-2
- HUOVINEN, P., RAMÍREZ, J., CAPUTO, L. & GÓMEZ, I. (2019). Mapping of spatial and temporal variation of water characteristics through satellite remote sensing in Lake Pangui-

- pulli, Chile. *Science of The Total Environment*, 679, 196-208. doi :10.1016/j.scitotenv.2019.04.367
- IONUȚ, M., DRAGOȘ, N. & BOGDAN, B. (2014). Preliminary Assessments of the Relation : Precipitations – Water – Sediments with Regard to some Anthropic Accumulations in the River Basin of Bahlui. *Present Environment and Sustainable Development*, 8(1), 119-126. doi :10.2478/pesd-2014-0010
- IRM. (2020). Atlas climatique - Précipitations annuelles. Library Catalog : www.meteo.be. Récupérée 31 juillet 2020, à partir de <https://www.meteo.be/fr/climat/atlas-climatique/cartes-climatiques/precipitations/quantites-de-precipitations/annuel>
- KEMKER, C. (2014). Turbidity, Total Suspended Solids & Water Clarity. Library Catalog : www.fondriest.com. Récupérée 29 mai 2020, à partir de <https://www.fondriest.com/environmental-measurements/parameters/water-quality/turbidity-total-suspended-solids-water-clarity/>
- KIM, H.-C., SON, S., KIM, Y. H., KHIM, J. S., NAM, J., CHANG, W. K., LEE, J.-H., LEE, C.-H. & RYU, J. (2017). Remote sensing and water quality indicators in the Korean West coast : Spatio-temporal structures of MODIS-derived chlorophyll-a and total suspended solids. *Marine Pollution Bulletin*, 121(1-2), 425-434. doi :10.1016/j.marpolbul.2017.05.026
- KRAVITZ, J., MATTHEWS, M., BERNARD, S. & GRIFFITH, D. (2020). Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets : Successes and challenges. *Remote Sensing of Environment*, 237, 111562. doi :10.1016/j.rse.2019.111562
- KUHN, C., de MATOS VALERIO, A., WARD, N., LOKEN, L., SAWAKUCHI, H. O., KAMPEL, M., RICHEY, J., STADLER, P., CRAWFORD, J., STRIEGL, R., VERMOTE, E., PAH-LEVAN, N. & BUTMAN, D. (2019). Performance of Landsat-8 and Sentinel-2 surface reflectance products for river remote sensing retrievals of chlorophyll-a and turbidity. *Remote Sensing of Environment*, 224, 104-118. doi :10.1016/j.rse.2019.01.023
- LI, J., CHEN, X., TIAN, L., HUANG, J. & FENG, L. (2015). Improved capabilities of the Chinese high-resolution remote sensing satellite GF-1 for monitoring suspended particulate matter (SPM) in inland waters : Radiometric and spatial considerations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 106, 145-156. doi :10.1016/j.isprsjprs.2015.05.009
- LIVERSEDGE, L. (2007). *Turbidity Mapping and Prediction in Ice Marginal Lakes at the Bering Glacier System, Alaska* (thèse de doct., University of Michigan).
- MAIN-KNORN, M., PFLUG, B., LOUIS, J., DEBAECKER, V., MÜLLER-WILM, U. & GASCON, F. (2017). Sen2Cor for Sentinel-2. In *Image and Signal Processing for Remote Sensing XXIII* (T. 10427, p. 1042704). doi :10.1117/12.2278218
- MALTESE, A., CAPODICCI, F., CIRAIOLO, G. & la LOGGIA, G. (2013). Coastal zone water quality : Calibration of a water-turbidity equation for MODIS data. *European Journal of Remote Sensing*, 46(1), 333-347. doi :10.5721/EuJRS20134619

- MONTANHER, O. C., NOVO, E. M. L. M., BARBOSA, C. C. F., RENNÓ, C. D. & SILVA, T. S. F. (2014). Empirical models for estimating the suspended sediment concentration in Amazonian white water rivers using Landsat 5/TM. *International Journal of Applied Earth Observation and Geoinformation*, 29, 67-77. doi :10.1016/j.jag.2014.01.001
- OLMANSON, L. G., BREZONIK, P. L. & BAUER, M. E. (2013). Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers : The Mississippi River and its tributaries in Minnesota. *Remote Sensing of Environment*, 130, 254-265. doi :10.1016/j.rse.2012.11.023
- ORGANISATION INTERNATIONALE DE NORMALISATION. (2016). ISO 7027-1 :2016. NTU. Récupérée 20 février 2020, à partir de <https://www.iso.org/cms/render/live/fr/sites/isoorg/contents/data/standard/06/28/62801.html>
- ORGANISATION INTERNATIONALE DE NORMALISATION. (2019). ISO 7027-2 :2019. Récupérée 20 février 2020, à partir de <https://www.iso.org/cms/render/live/fr/sites/isoorg/contents/data/standard/06/95/69545.html>
- OUILLO, S., DOUILLET, P., PETRENKO, A., NEVEUX, J., DUPOUY, C., FROIDEFOND, J.-M., ANDRÉFOUËT, S. & MUÑOZ-CARAVACA, A. (2008). Optical algorithms at satellite wavelengths for total suspended matter in tropical coastal waters. *Sensors*, 8(7), 4165-4185. doi :10.3390/s8074165
- PAGE, B. P., OLMANSON, L. G. & MISHRA, D. R. (2019). A harmonized image processing workflow using Sentinel-2/MSI and Landsat-8/OLI for mapping water clarity in optically variable lake systems. *Remote Sensing of Environment*, 231, 111284. doi :10.1016/j.rse.2019.111284
- PALMER, S. C. J., KUTSER, T. & HUNTER, P. D. (2015). Remote sensing of inland waters : Challenges, progress and future directions. *Remote Sensing of Environment*. Special Issue : Remote Sensing of Inland Waters, 157, 1-8. doi :10.1016/j.rse.2014.09.021
- PETUS, C., CHUST, G., GOHIN, F., DOXARAN, D., FROIDEFOND, J.-M. & SAGARMINAGA, Y. (2010). Estimating turbidity and total suspended matter in the Adour River plume (South Bay of Biscay) using MODIS 250-m imagery. *Continental Shelf Research*, 30(5), 379-392. doi :10.1016/j.csr.2009.12.007
- POTES, M., COSTA, M. & SALGADO, R. (2012). Satellite remote sensing of water turbidity in Alqueva reservoir and implications on lake modelling. *Hydrology and Earth System Sciences*, 16(6), 1623-1633. doi :10.5194/hess-16-1623-2012
- RASMUSSEN, P. P., JOHN R. GRAY, G. DOUGLAS GLYSSON & ANDREW C. ZIEGLER. (2009). *Guidelines and procedures for computing time-series suspended-sediment concentrations and loads from in-stream turbidity-sensor and streamflow data*. Techniques and methods; U.S. Geological Survey,
- RITCHIE, J. C., ZIMBA, P. V. & EVERITT, J. H. (2003). Remote Sensing Techniques to Assess Water Quality. *Photogrammetric Engineering & Remote Sensing*, 69(6), 695-704. doi :10.14358/PERS.69.6.695

- SAGAN, V., PETERSON, K. T., MAIMAITIJIANG, M., SIDIKE, P., SLOAN, J., GREELING, B. A., MAALOUF, S. & ADAMS, C. (2020). Monitoring inland water quality using remote sensing : potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. *Earth-Science Reviews*, 205, 103187. doi :10.1016/j.earscirev.2020.103187
- SÒRIA-PERPINYÀ, X., VICENTE, E., URREGO, P., PEREIRA-SANDOVAL, M., RUÍZ-VERDÚ, A., DELEGIDO, J., SORIA, J. M. & MORENO, J. (2020). Remote sensing of cyanobacterial blooms in a hypertrophic lagoon (Albufera of València, Eastern Iberian Peninsula) using multitemporal Sentinel-2 images. *Science of The Total Environment*, 698, 134305. doi :10.1016/j.scitotenv.2019.134305
- UNEP. (2016). A snapshot of the World's Water Quality : Towards a Global Assessment. *United Nations Environment Programme, Nairobi, Kenya*.
- WAGNER, R., BOULGER, R., OBLINGER, C. & SMITH, B. (2006). *Guidelines and standard procedures for continuous water-quality monitors : Station operation, record computation, and data reporting Techniques and Methods 1-D3*.
- WARREN, M. A., SIMIS, S. G. H., MARTINEZ-VICENTE, V., POSER, K., BRESCIANI, M., ALIKAS, K., SPYRAKOS, E., GIARDINO, C. & ANSPER, A. (2019). Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. *Remote Sensing of Environment*, 225, 267-289. doi :10.1016/j.rse.2019.03.018
- YANG, X., QIN, Q., YÉSOU, H., LEDAUPHIN, T., KOEHL, M., GRUSSENMEYER, P. & ZHU, Z. (2020). Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. *Remote Sensing of Environment*, 244, 111803. doi :10.1016/j.rse.2020.111803
- ZHANG, J.-c., ZHANG, W., LI, J.-h., SHI, Z.-g. & PU, S.-y. (2004). Relation between precipitation and sediment transport in the Dasha River Watershed. *Chinese Geographical Science*, 14(2), 129-134. doi :10.1007/s11769-004-0020-4

Annexes

A Bandes spectrales des satellites Sentinel-2

TABLEAU 14 – *Détails des bandes spectrales des satellites Sentinel-2.*

Bande	Sentinel-2A		Sentinel-2B		Résolution spatiale (m)
	Longueur d'onde centrale (nm)	Largeur (nm)	Longueur d'onde centrale (nm)	Largeur (nm)	
1	442,7	21	442,3	21	60
2	492,4	66	492,1	66	10
3	559,8	36	559	36	10
4	664,6	31	665	31	10
5	704,1	15	703,8	16	20
6	740,5	15	739,1	15	20
7	782,8	20	779,7	20	20
8	832,8	106	833	106	10
8a	864,7	21	864	22	20
9	945,1	20	943,2	21	60
10	1373,5	31	1376,9	30	60
11	1613,7	91	1610,4	94	20
12	2202,4	175	2185,7	185	20

B Script GEE pour l'extraction des réflectances de surface

```
// Point d'intérêt de 9 pixels , exemple avec Hastière
var poi = ee.Geometry.Point(4.820, 50.198).buffer(15).bounds();
var zoom = 14;
Map.centerObject(poi, zoom);
```

```
// Période
var startDate = '2017-03-28';
var endDate = '2018-12-31';
```

```
// FONCTIONS PERMETTANT D'ÉLIMINER LES NUAGES //
```

```
// avec la bande QA60
var maskCloudsQA60 = function(image){

  var qa = image.select('QA60');
  var cloudBitMask = Math.pow(2, 10);
  var cirrusBitMask = Math.pow(2, 11);
  var clear = qa.bitwiseAnd(cloudBitMask).eq(0)
               .and(qa.bitwiseAnd(cirrusBitMask).eq(0));
  var mask1 = clear.eq(1);
  return image.updateMask(mask1);};
```

```
// avec la carte de classification des scènes (SLC)
var maskCloudsSCL = function(image){
```

```
  var scl = image.select('SCL');
```

```
  var shadow = scl.eq(3); // ombre de nuage
  var LowClouds = scl.eq(7); // nuages à probabilité faible
  var MedClouds = scl.eq(8); // nuages à probabilité moyenne
  var HighClouds = scl.eq(9); // nuages à probabilité haute
  var cirrus = scl.eq(10); // cirrus
```

```
  var mask = (cirrus).neq(1).and((shadow).neq(1)).and
              ((LowClouds).neq(1)).and((MedClouds).neq(1))
```

```

        .and((HighClouds).neq(1));

return image.updateMask(mask);});

// FIN FONCTIONS //

// Charger les données de Sentinel-2 SR
var s2Col = ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(startDate, endDate)
    .filterBounds(poi)
    .select(['B1', 'B2', 'B3', 'B4', 'B5', 'B6',
        'B7', 'B8', 'B8A', 'B9', 'B11', 'B12',
        'AOT', 'WVP', 'SCL', 'TCI_R', 'TCI_G',
        'TCI_B', 'QA10', 'QA20', 'QA60'])
    .map(maskCloudsQA60)
    .map(maskCloudsSCL);

// Nombre d'images dans la collection
var count = s2Col.size();
print('Number of images in the collection ', count);
Map.addLayer(s2Col, {bands: ['B4', 'B3', 'B2'],
    min: 0, max: 2000},
    'S2 Collection ', true);

// GRAPHIQUE DES SÉRIES CHRONOLOGIQUES //

// Charger l'imagerie Sentinel-2 SR et sélectionner les bandes
var S2collection = s2Col
    .select(['B1', 'B2', 'B3', 'B4', 'B5', 'B6',
        'B7', 'B8', 'B8A', 'B9', 'B11', 'B12']);

// Options du graphique
var opGraf = {
    title : 'Surface réflectance moyenne (3x3) – Hastière',
    vAxis: {title: 'Réflectance de surface [0.0001/sr]'},
    lineWidth: 1,
    pointSize: 3,};

// Afficher le graphique
print('Hastière',

```

```

        ui.Chart.image.series(S2collection, poi, ee.Reducer.mean())
            .setChartType('LineChart')
            .setOptions(opGraf));

// Compter les pixels dans la zone sélectionnée
// sur la première image de la collection non masquée
var pixstats = ee.Image(ee.ImageCollection('COPERNICUS/S2_SR')
    .filterDate(startDate, endDate)
    .filterBounds(poi)
    .first()).reduceRegion({
    reducer: ee.Reducer.count(),
    geometry: roi,
    scale: 10});

print('Pixels number', pixstats);

// Ajouter la localisation sur la carte
Map.addLayer(roi, {color: 'FF0000'}, "Hastière");

```

C Graphiques des transformations logarithmiques

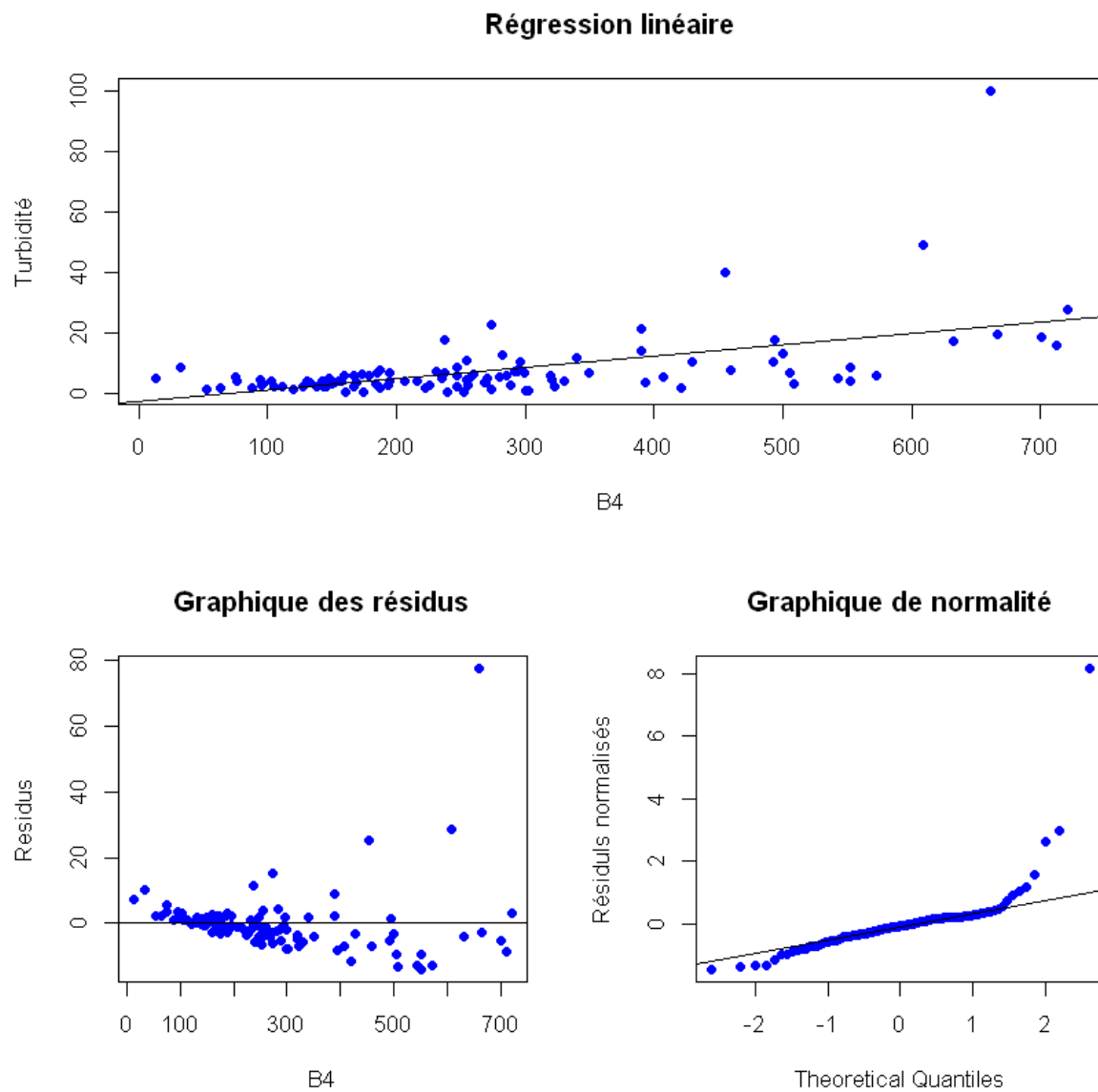


FIGURE 22 – *Graphiques de la turbidité en fonction de la bande B4.*

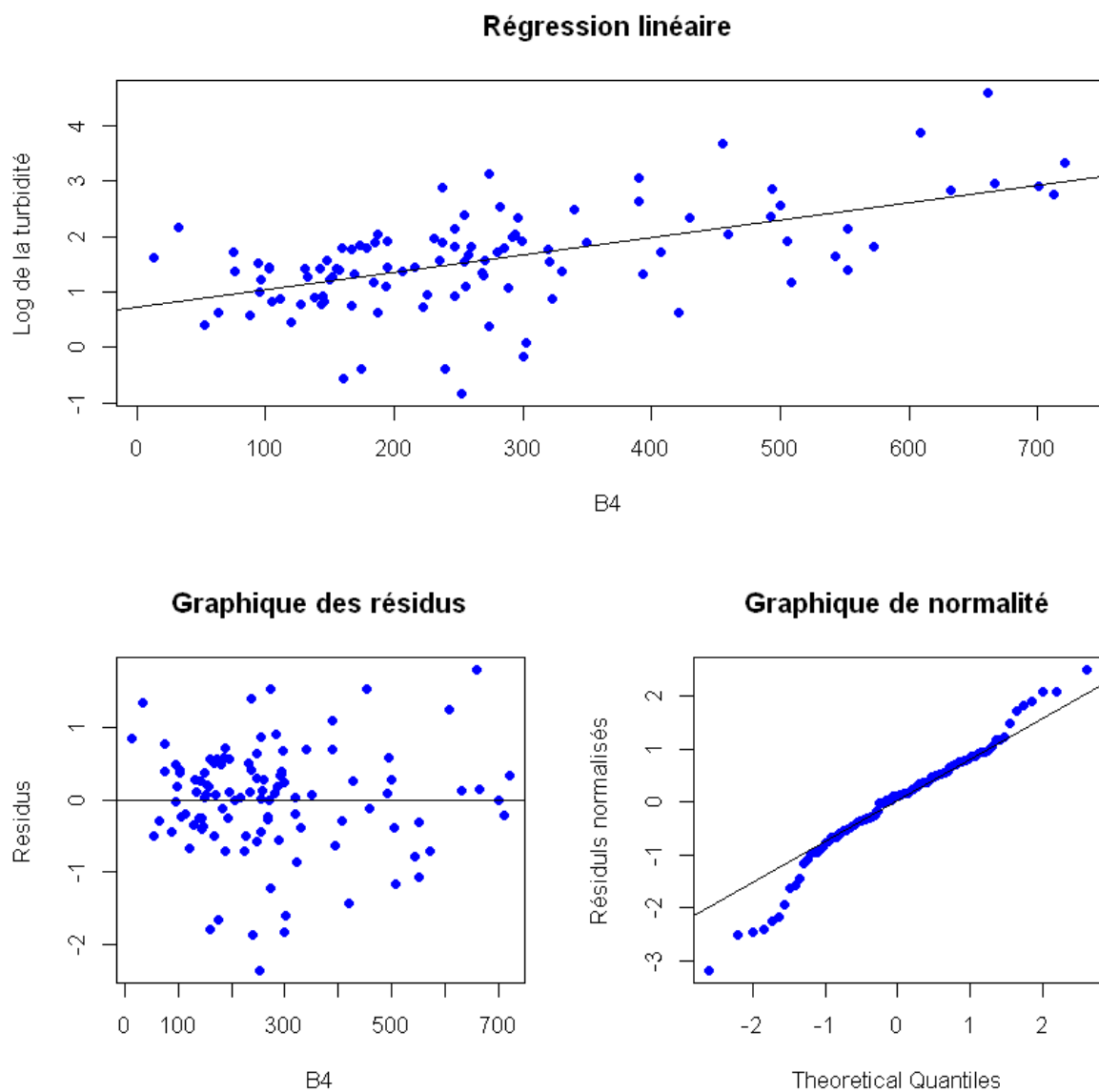


FIGURE 23 – *Graphiques du logarithme de la turbidité en fonction de la bande B4.*

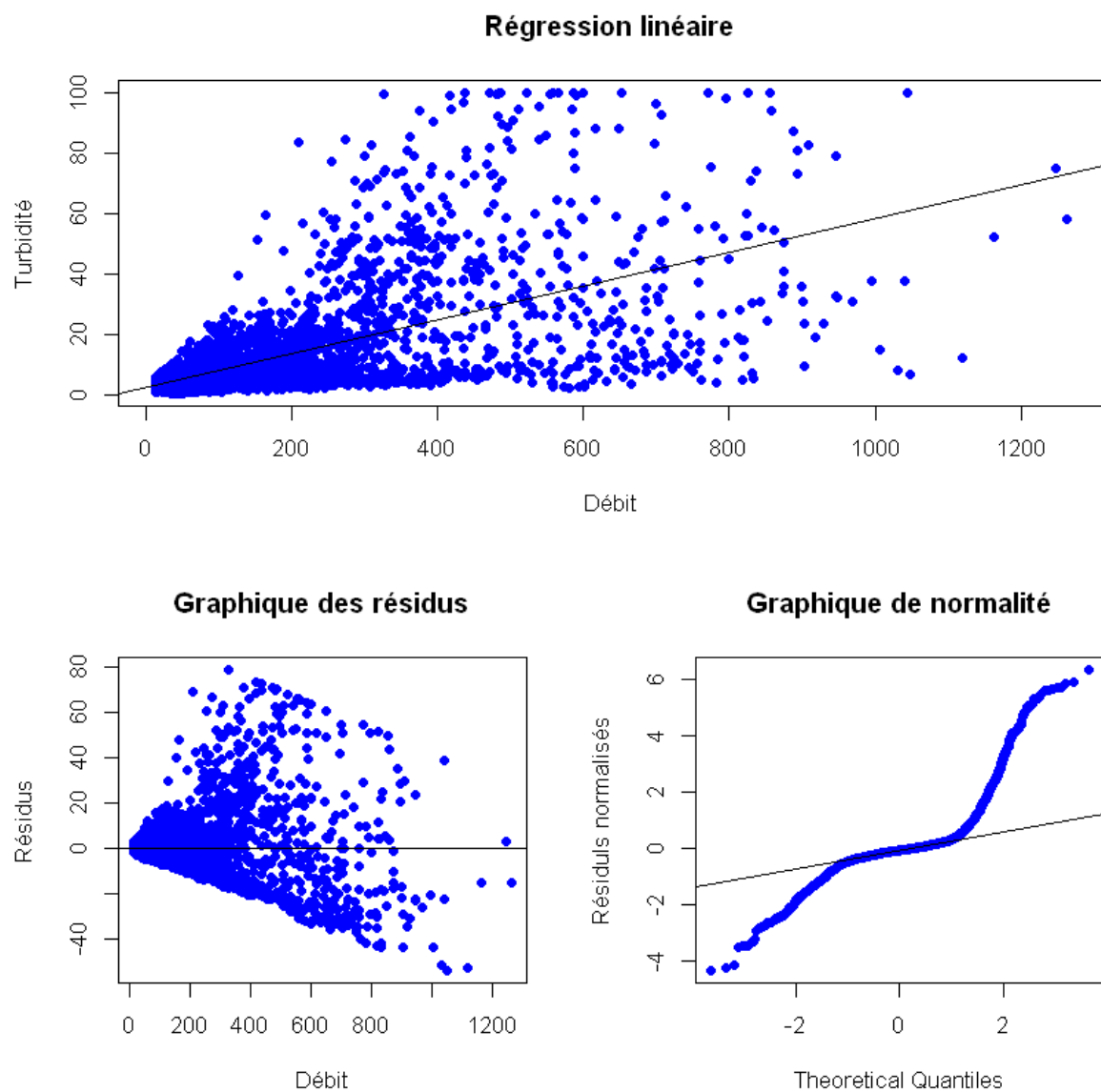


FIGURE 24 – *Graphiques de la turbidité en fonction du débit.*

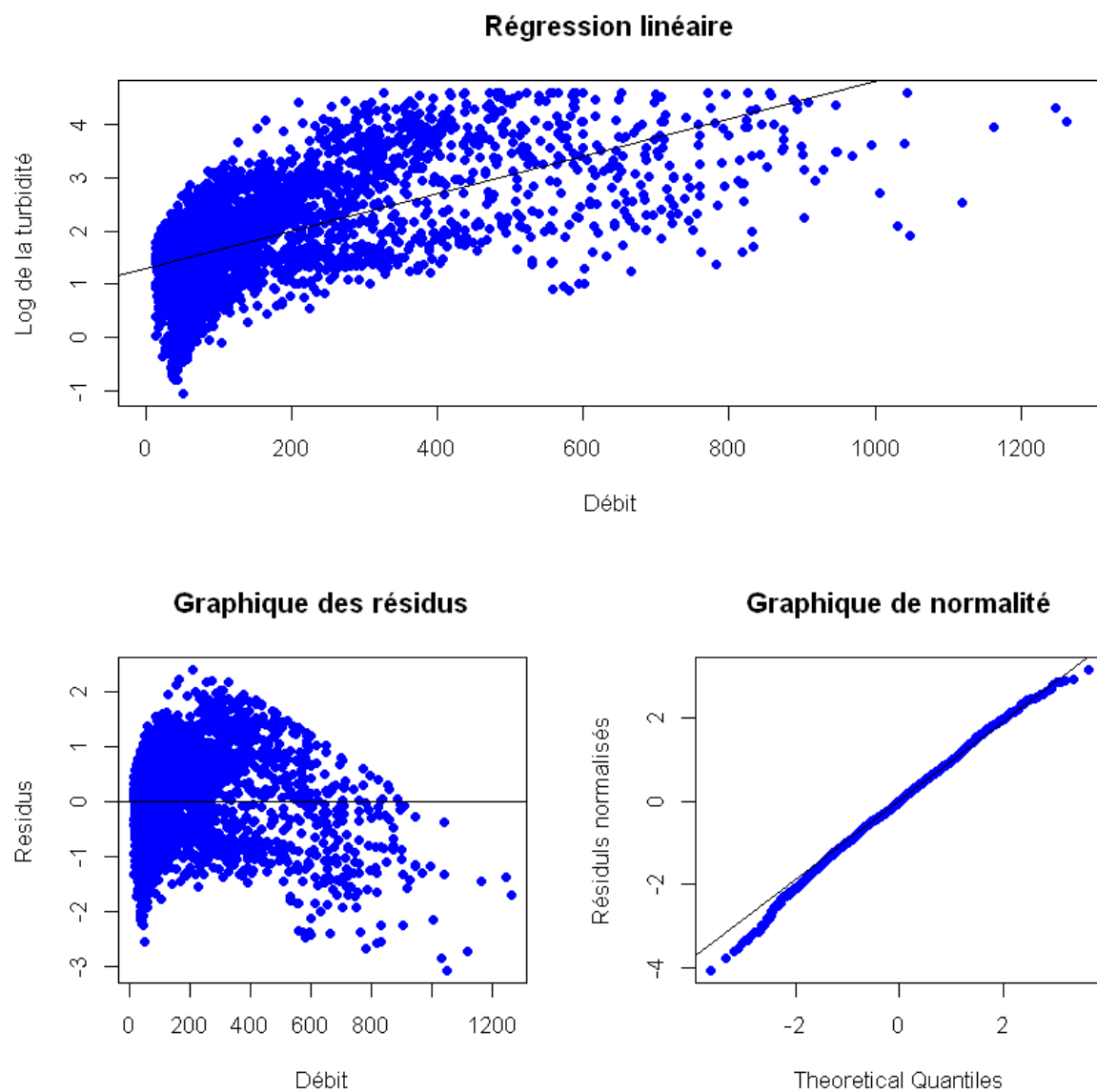


FIGURE 25 – *Graphiques du logarithme de la turbidité en fonction du débit.*

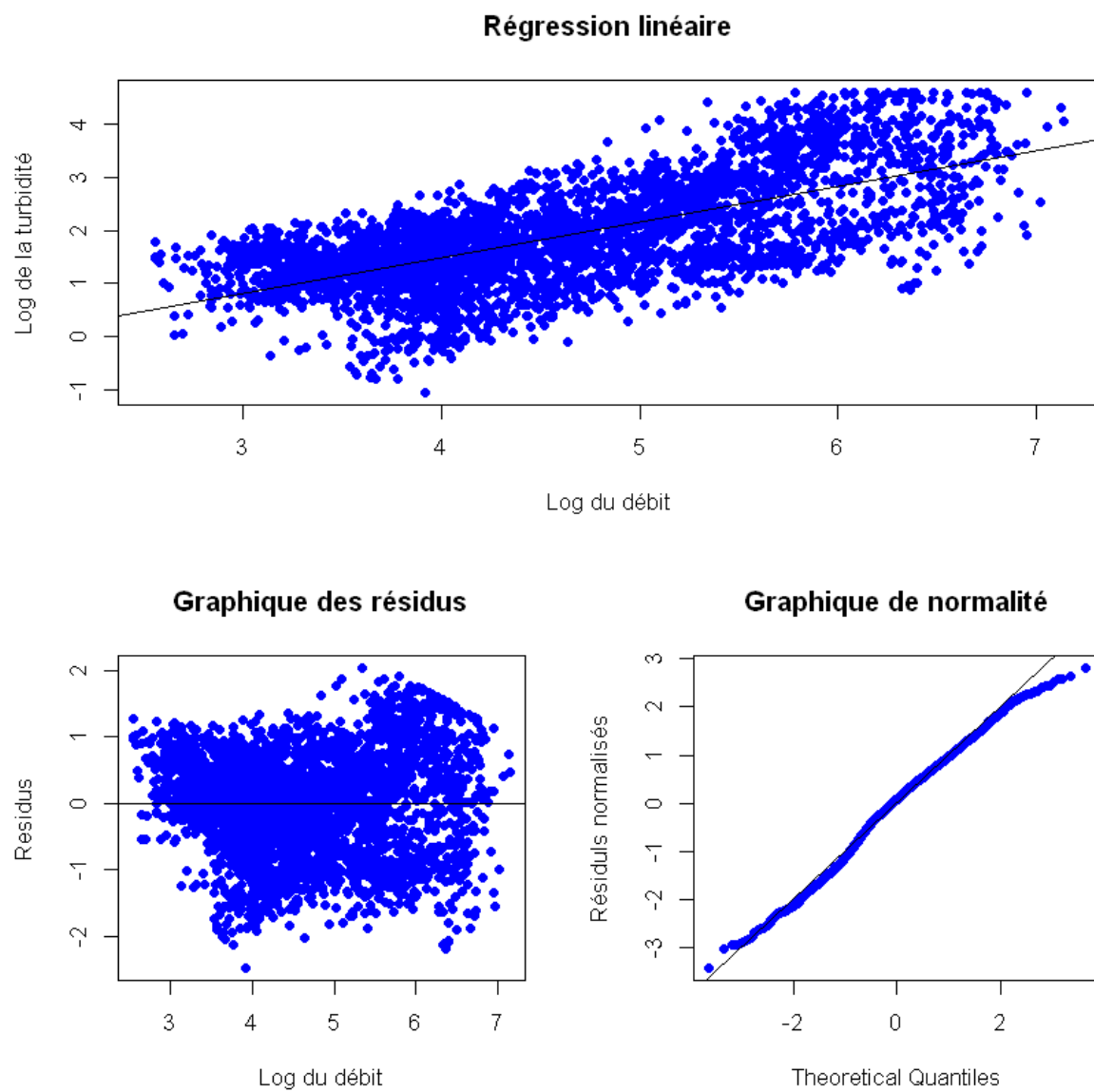


FIGURE 26 – *Graphiques du logarithme de la turbidité en fonction du logarithme du débit.*

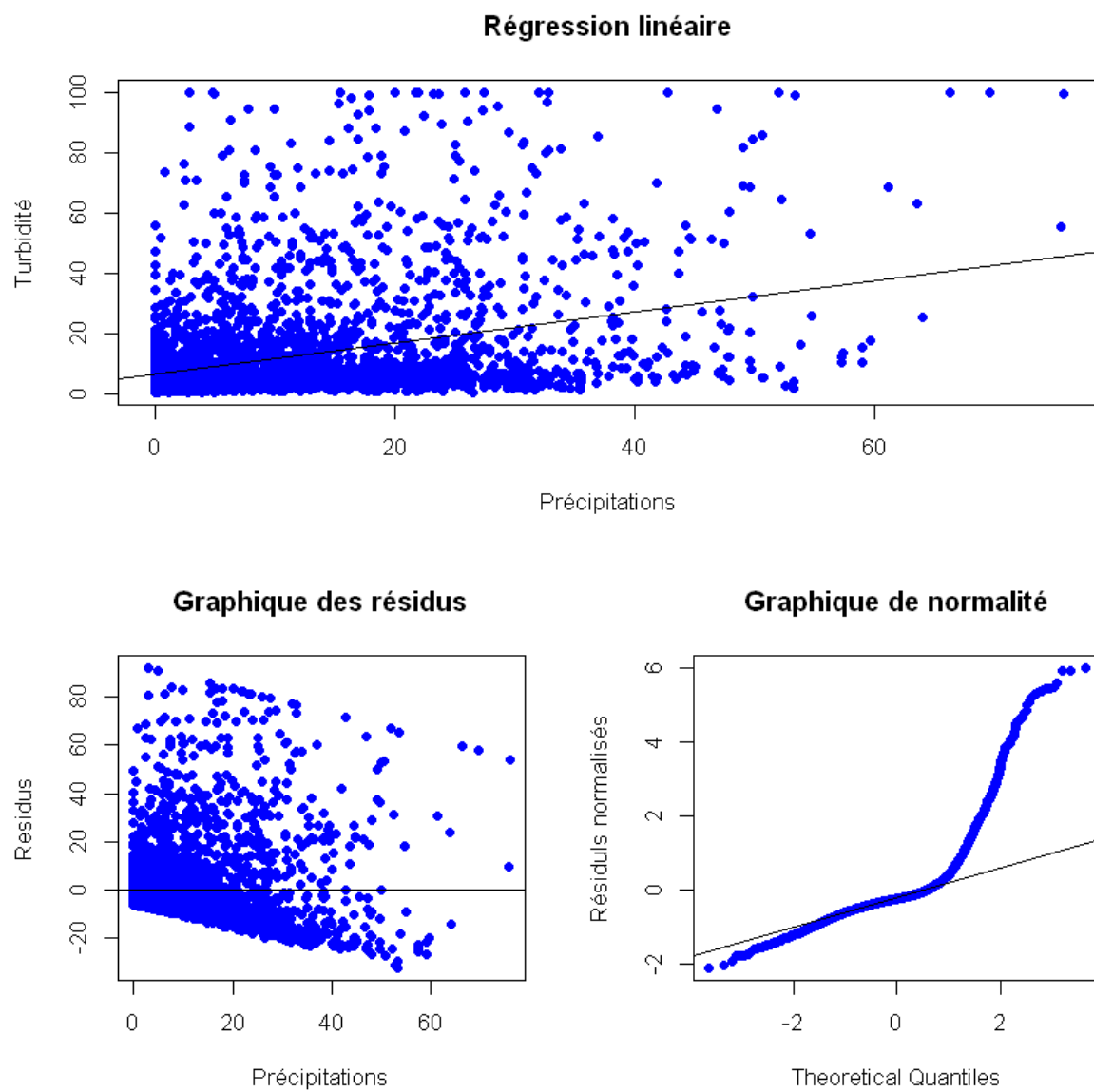


FIGURE 27 – *Graphiques de la turbidité en fonction des précipitations.*

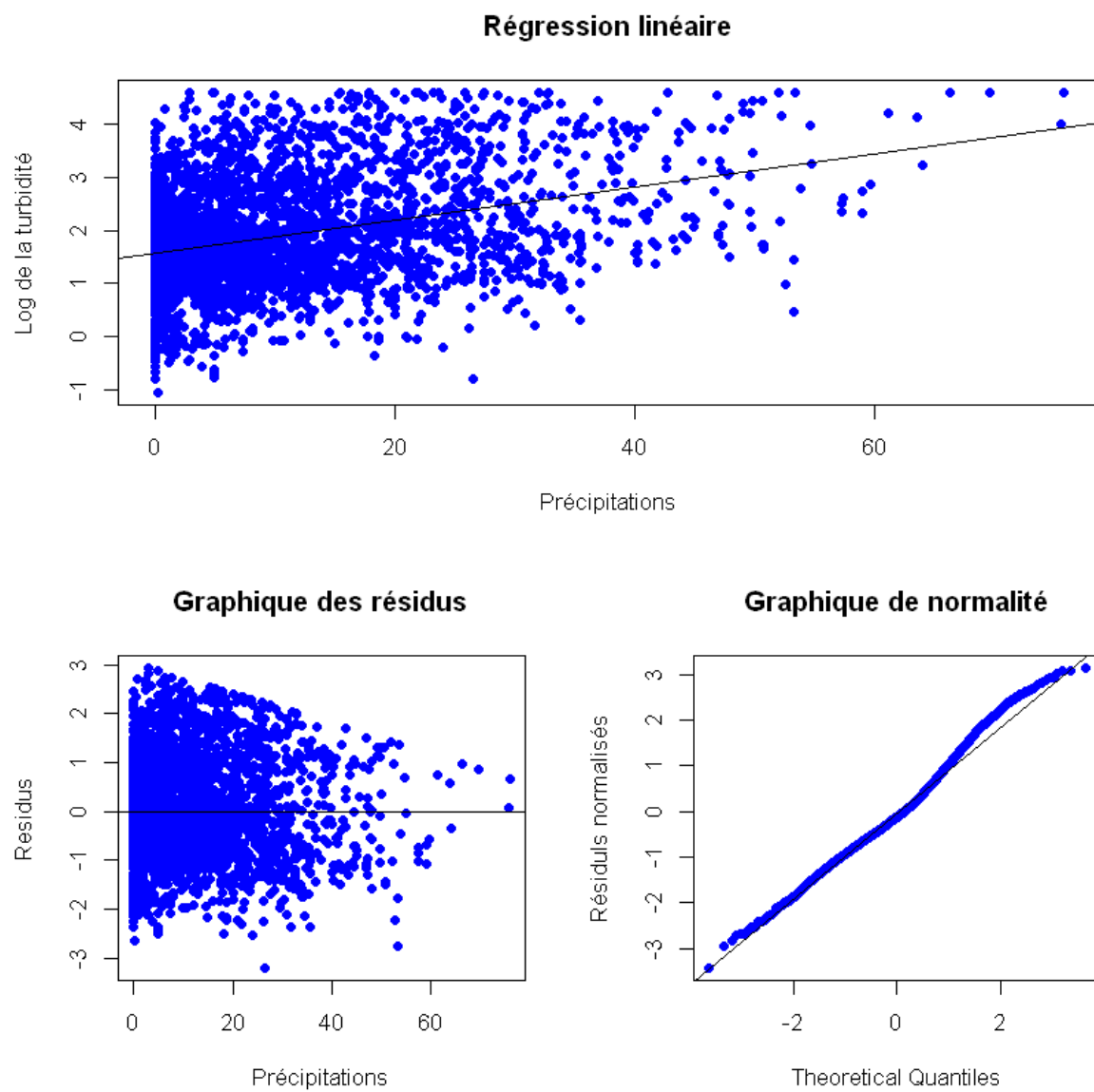


FIGURE 28 – *Graphiques du logarithme de la turbidité en fonction des précipitations.*

D Régressions linéaires associées aux modèles prédictifs

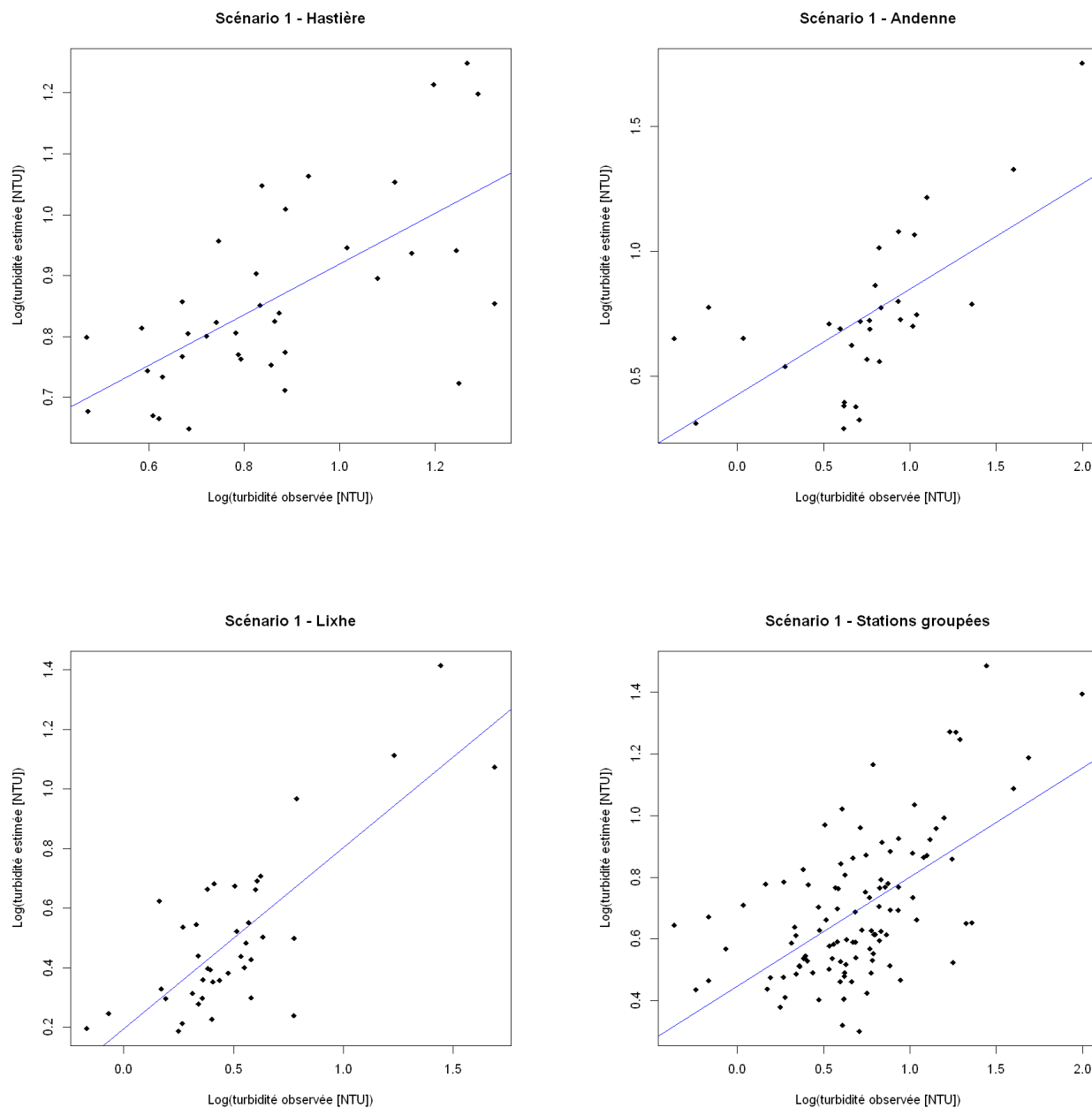


FIGURE 29 – Régressions linéaires du scénario 1

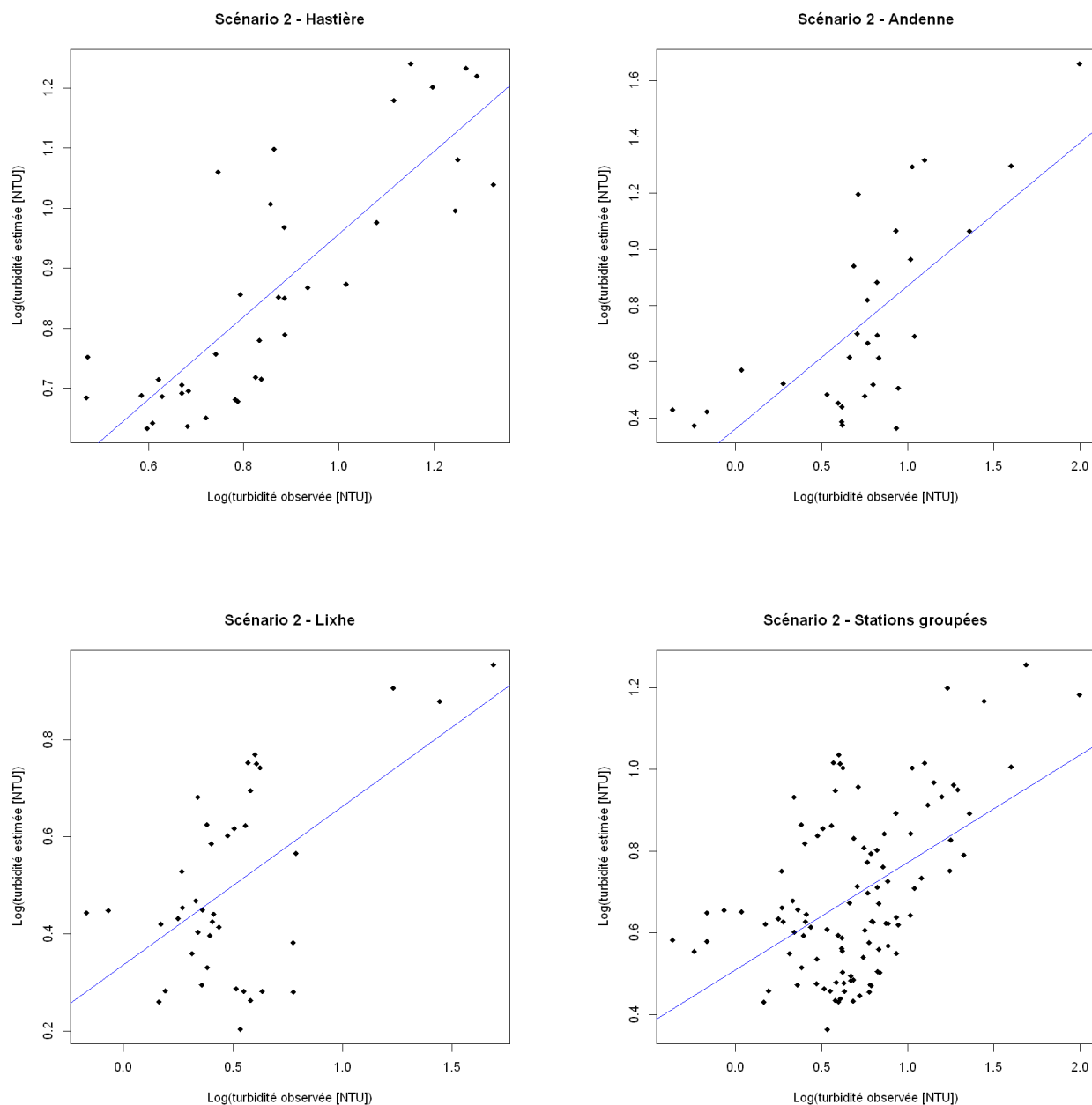


FIGURE 30 – Régressions linéaires du scénario 2

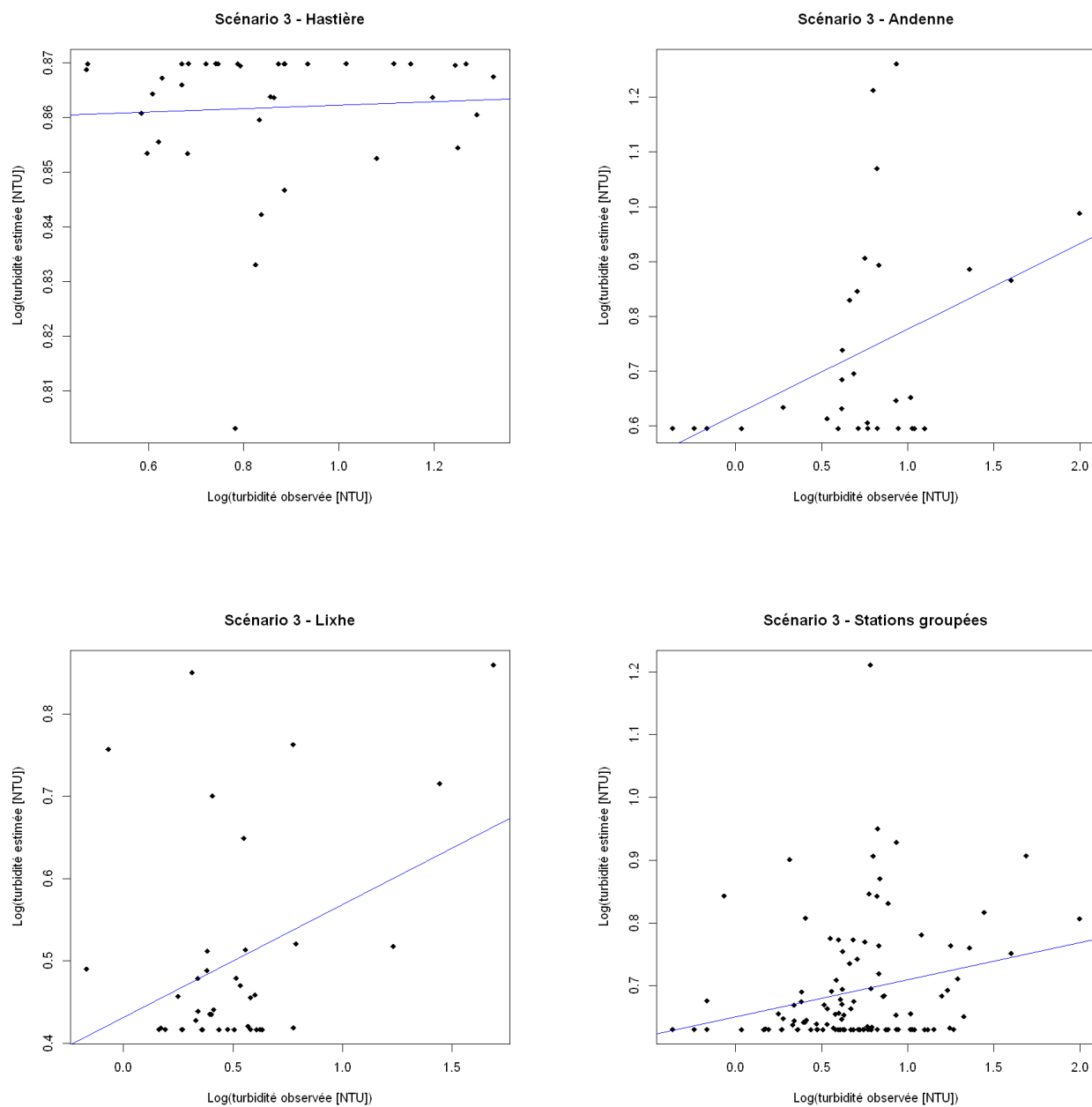


FIGURE 31 – Régressions linéaires du scénario 3

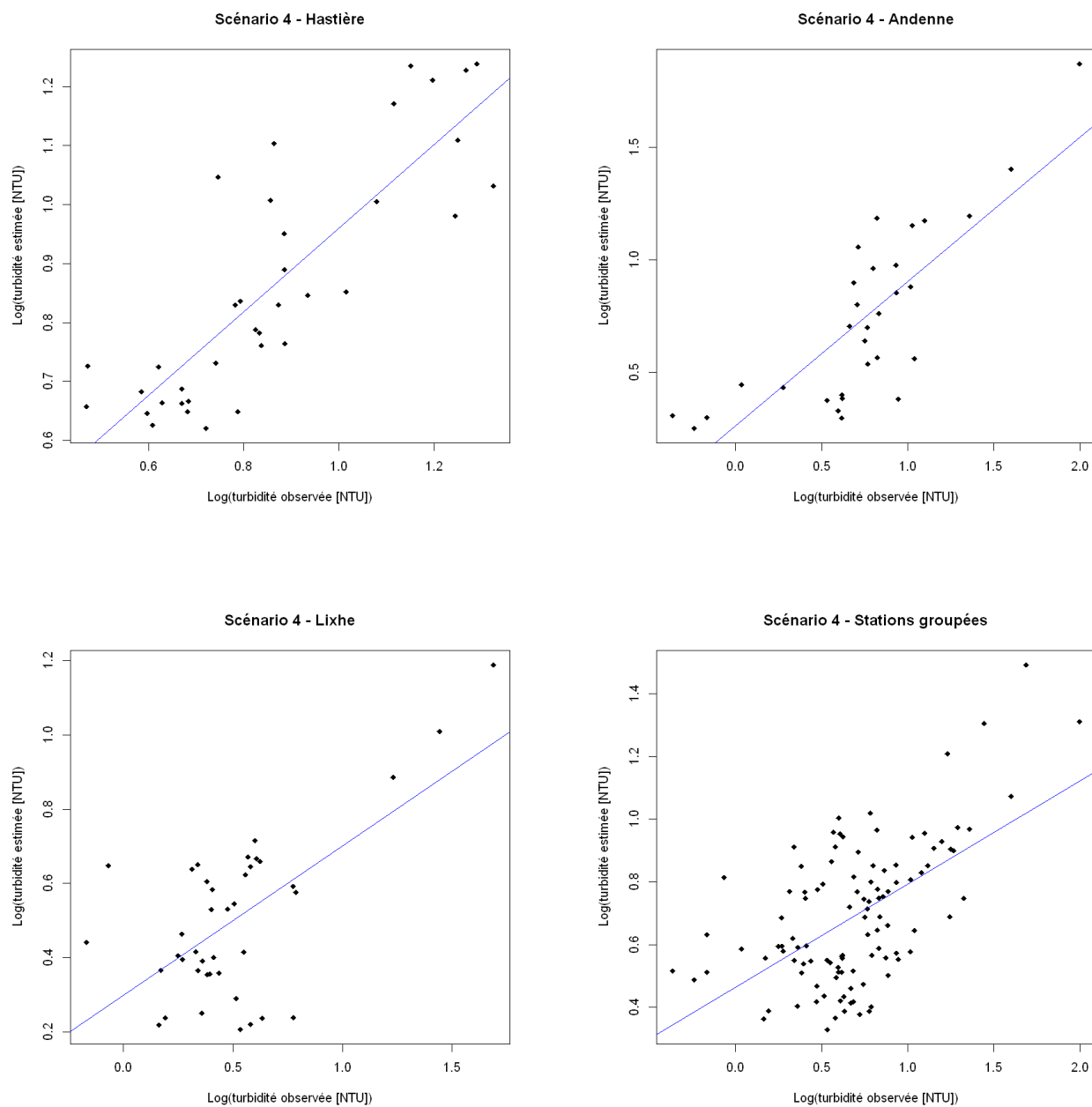


FIGURE 32 – Régressions linéaires du scénario 4

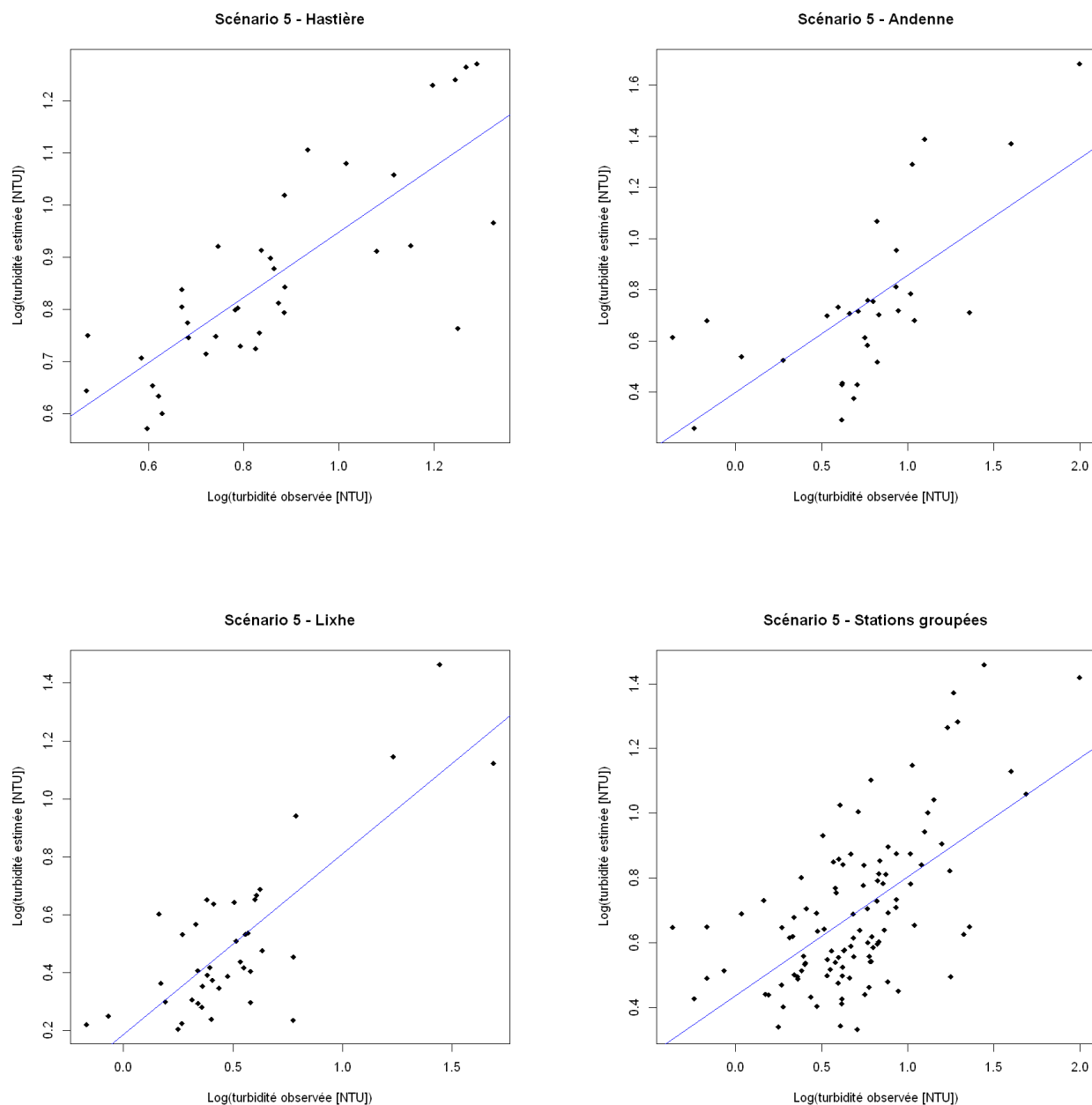


FIGURE 33 – Régressions linéaires du scénario 5

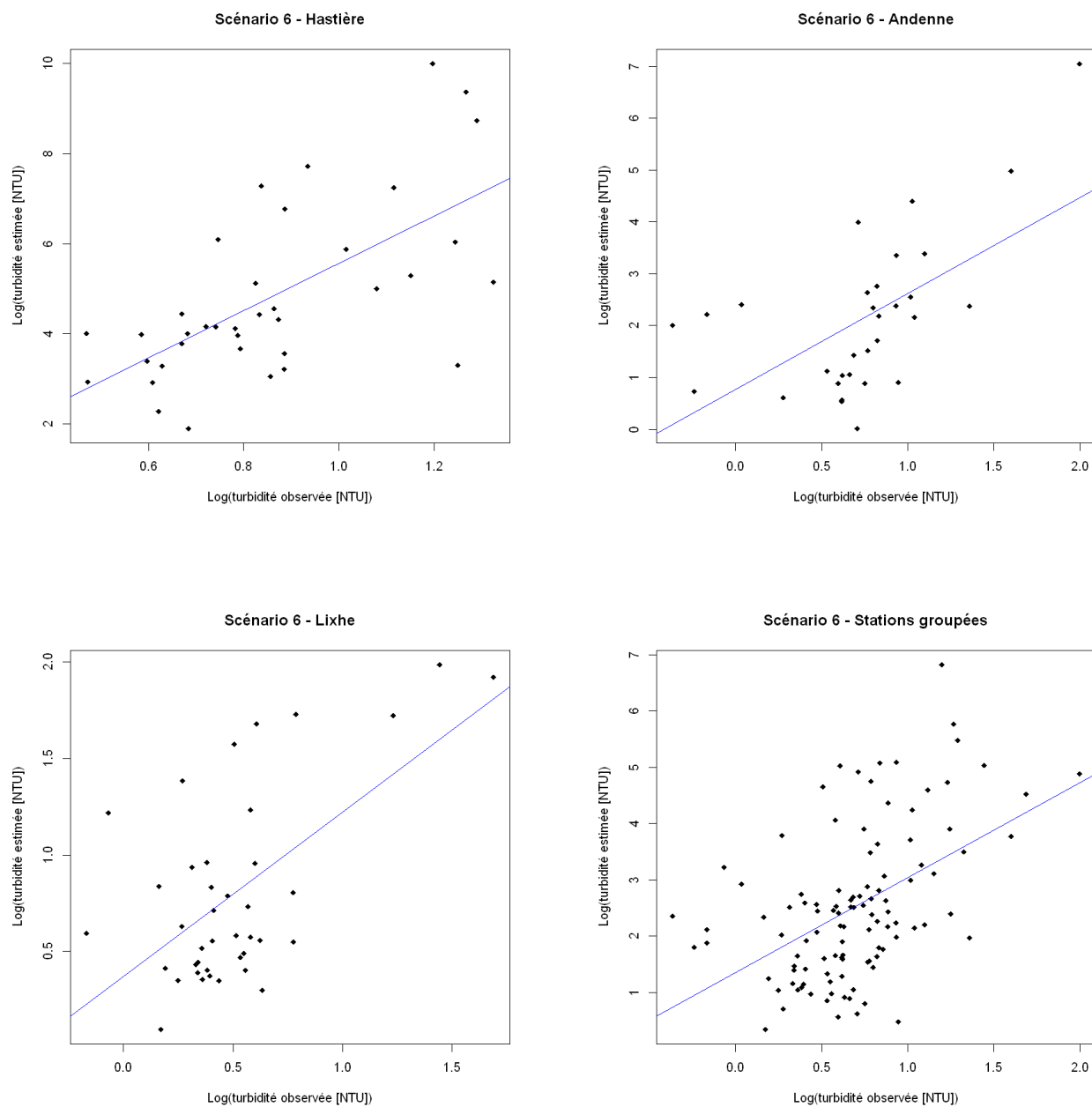


FIGURE 34 – Régressions linéaires du scénario 6

APPROCHE HYBRIDE POUR LA SURVEILLANCE DE LA QUALITÉ DE L'EAU A GRANDE ÉCHELLE

Surveillance de la turbidité de la Meuse grâce aux images Sentinel-2

Axelle Dagnelie

Les rivières sont des écosystèmes essentiels qu'il est important de surveiller et dont la gestion est devenue un enjeu majeur. Grâce aux récents progrès en matière de télédétection, il est aujourd'hui possible d'observer la qualité de l'eau des rivières par l'utilisation des produits satellitaires. C'est dans cette optique que ce mémoire s'inscrit pour le cas d'étude de la surveillance de la turbidité dans le segment wallon de la Meuse à l'aide des images Sentinel-2 de niveau 2A.

L'objectif était d'évaluer les limitations des images Sentinel-2 en parallèle avec d'autres variables de prédiction de la turbidité et d'éventuellement créer un modèle général à la rivière d'étude.

A cette fin, des données de débit et de précipitations ainsi que les données de réflectance dans les différentes bandes spectrales mesurées par le capteur de Sentinel-2 furent récoltées pour trois endroits différents le long de la Meuse et intégrées individuellement ou combinées en tant que variables prédictives dans différents scénarios. Dans le cas où les variables étaient plus de deux dans un même scénario, elles furent implémentées dans des analyses en composantes principales pour en tirer l'information principale. A la suite de cela, plusieurs modèles furent testés selon les scénarios et pour les données de chaque localisation ainsi que pour l'ensemble des données.

Les résultats de ce mémoire montrent des résultats moyennement satisfaisants ouvrant la porte à des perspectives d'amélioration d'observation de la qualité de l'eau par télédétection.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

Faculté des bioingénieurs

Croix du Sud, 2bte L7.05.01, 1348 Louvain-La-Neuve, Belgique | www.uclouvain.be/agro