

Qualification des données Quadriges²

Bilan 2018



SOMMAIRE

1	Qualification : définition et mise en œuvre	3
2	Avancement général de la qualification en 2018.....	3
3	Détails des travaux de qualification 2018	5
3.1	Qualification du jeu de données VELYGER	5
3.2	Qualification des données « Chimie »	6
3.2.1	L' « Imposex » : le programme RNOPHY	6
3.2.2	Le RNO Matière Vivante.....	7
3.2.3	Qualifications ponctuelles.....	7
3.3	Qualification hydrologie Antilles 2007-2017	8
3.3.1	Critères de sélection des données	8
3.3.2	Qualification automatique réalisée.....	8
3.3.3	Corrections manuelles en marge de la qualification.....	10
3.3.4	Qualification experte Martinique.....	11
3.3.5	Zoom sur les données qualifiées Douteux / Faux.....	12
3.4	Qualification hydrologie métropole 2016	13
3.4.1	Critères de sélection des données	13
3.4.2	Avancement des travaux 2018.....	14
3.4.3	Bilan du processus 2018.....	15
3.5	Préparation de la qualification des données REPOM Sédiments	16
3.5.1	Contexte	16
3.5.2	Programme de travail établi.....	16
3.5.3	Bilan des travaux 2018	17
3.6	Communication	20
3.6.1	Fiches qualification.....	20
3.6.2	Journées REPHY.....	20
4	Bilan – discussion	21
	Annexe 1 : Qualification automatique hydrologie Antilles 2007-2017 : Nombre de données détectées et corrigées par anomalie.....	23
	Annexe 2 : Commentaire de qualification des données bonnes déclassées à Douteux ou Faux.....	26

1 Qualification : définition et mise en œuvre

La qualification a pour objectif d'attribuer un niveau de qualité aux données : Bon, Douteux ou Faux. Le processus de qualification consiste à rechercher des anomalies plus ou moins complexes dans les données stockées, les anomalies étant définies par les responsables de programme Quadrigé ou responsables thématiques. Ces recherches d'anomalies sont faites en lien avec la Cellule Quadrigé.

On distingue plusieurs types de qualification :

- **Qualification ponctuelle de quelques résultats** : Exemples : on sait qu'un échantillon a été mal conservé et que les résultats de son analyse seront douteux, ou on a détecté un problème analytique et on sait que les résultats sont faux.
- **Qualification d'un jeu de données important** :
 - soit lors d'une reprise de données dont la qualité a été évaluée préalablement à l'intégration dans Quadrigé (dans ce cas les données sont intégrées avec un niveau de qualification déjà défini),
 - soit un expert thématique travaille sur un jeu de données précis, que ce soit pour un rapport d'étude, une publication scientifique, ou toute autre analyse statistique : il peut transmettre les résultats de son expertise à la cellule d'administration Quadrigé, et des niveaux de qualité peuvent être attribués aux données.
- **Qualification dite « automatique »** : mise en place depuis 2009 sur certaines thématiques, elle a pour objectifs de qualifier les données en routine via la recherche d'anomalies prédéfinies par des experts thématiques. Des programmes de recherche de ces anomalies sont lancés périodiquement par la Cellule Quadrigé, générant des fichiers d'anomalies renvoyés aux producteurs de données. Après vérification, les réponses des producteurs sont validées par les experts thématiques et sont intégrées dans la base : les données sont qualifiées après une éventuelle correction.
- **Qualification dite « experte »** : elle consiste à analyser les résultats selon des modèles statistiques définis par des experts thématiques avec l'équipe de biostatisticiens du service VIGIES de l'Ifremer (service auquel appartient la cellule Quadrigé).

2 Avancement général de la qualification en 2018

Toutes les données de Quadrigé ne font pas l'objet de qualification (même si elles ont toutes vocation à le faire). A ce jour cinq thématiques présentent des données qualifiées (Tableau 1). Pour la qualification « automatique » les thématiques sont définies par des critères qui sont précisés dans les chapitres suivants.

Tableau 1 : Nombre de résultats présents dans Quadrigé avec un statut qualifié (Bon, Douteux ou Faux) par thématique et par programme. Le précédent chiffrage de 2017 est présenté à titre informatif.

Thématique	Mode de qualification	Programme	Nombre de résultats qualifiés	
			2017	2018
Chimie	Automatique + experte + ponctuelle	RNOINT	50	50
		RNOMV	196 449	197 468
		RNOPHY	4 717	8 664
		RNOSED	33 961	33 961
		ROCCHAUCHIMIE	469	469
	Ponctuelle	REPOMS	887	887
Hydrologie	Automatique + ponctuelle	ARCHYD	79 503	79 503
		REPHY	288 494	317 756
		MARTINIQUE-EAU-ETUDES		15 021

Thématique	Mode de qualification	Programme	Nombre de résultats qualifiés	
			2017	2018
		RHLN	12 203	12 211
		RNOHYD	698 003	697 993
		SRN	23 857	23 857
		OMEGA	62	62
	Reprise de données	REPHY (Réseau des estuaires Bretons)	116 510	116 512
		RESLOC_44	80 386	80 386
Microbiologie	Automatique + ponctuelle	CMIC	3	3
		REMIC	15 637	15 637
		REMIE1	39	39
		REMIE2	8	8
		REMIS	130	130
		REMI-SURV	83	83
Ressources aquacoles	Ponctuelle (données historiques)	VELYGER		64 470
Zooplankton	Reprise de données + ponctuelle	IGAZOH	185 111	185 111

Les thématiques qualifiées en 2018 sont la Chimie, l'Hydrologie et les ressources aquacoles (programme VELYGER). Cette qualification a été réalisée selon différents processus : qualification ponctuelle de résultats suite à des demandes d'experts thématiques, qualification de tout le jeu de données VELYGER historique suite à son analyse et correction par les coordinateurs Ifremer, et qualification dite « automatique » pour les données hydrologiques.

Tableau 2 : Qualification des résultats Quadrigé réalisée en 2018 par thématique et mode de qualification.

Thématique	Programme	Mode de qualification	Période	Nb résultats
Chimie	RNOMV	Ponctuelle	1999, 2016	823
Hydrologie	ARCHYD	Ponctuelle	2003-2004	2
	MARTINIQUE-EAU-ETUDES	Automatique + experte	2008-2017	26 512
	REPHY	Automatique + experte	2007-2017	
	REPHY	Ponctuelle	1996	37 657
	RHLN	Ponctuelle	2016	8
	RNOHYD	Ponctuelle	1978	1
Ressources Aquacoles	VELYGER	Jeu de données	1973-2015	64 470

En termes de répartition géographique des résultats qualifiés, un gros effort de qualification a été réalisé en 2018 aux Antilles sur les données hydrologiques. La qualification des données hydrologiques métropole (REPHY + réseaux régionaux) a été réalisée pour les niveaux passages et prélèvements, les échantillons et résultats seront qualifiés en 2019 et ne sont donc pas comptabilisés dans la Figure 1. Dans les résultats métropole qualifiés en 2018, le programme VELYGER représente 99% des qualifications réalisées.

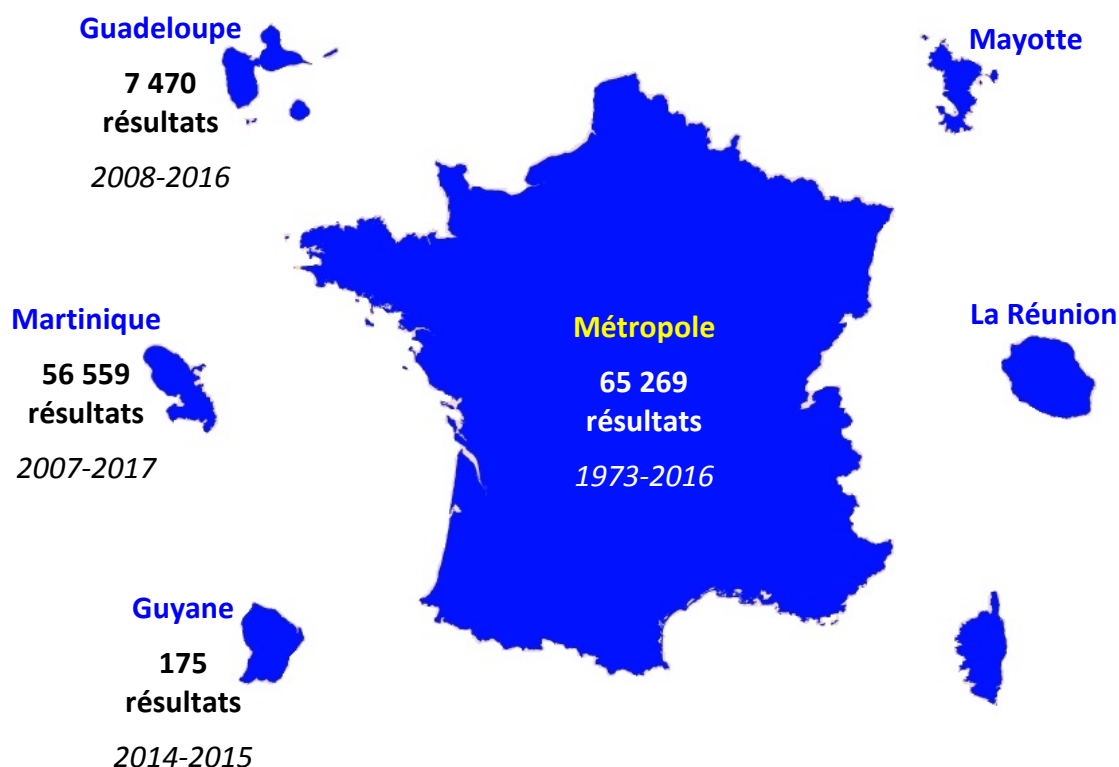


Figure 1: Nombre de résultats qualifiés en 2018 par territoire.

3 Détails des travaux de qualification 2018

3.1 Qualification du jeu de données VELYGER

Le réseau VELYGER a pour objectifs d'ObserVER, anaLYser et GERer la variabilité de la reproduction de l'huître creuse *Crassostrea gigas*. Il consiste à suivre le recrutement larvaire de cette espèce par comptages des larves dans des prélèvements d'eau.

Le LER d'Arcachon (notamment via Isabelle Auby) ainsi que le Laboratoire de Physiologie des Invertébrés d'Argenton (Stéphane Pouvreau) ont examiné l'ensemble des données acquises depuis le début de la surveillance en 1973 jusqu'à 2015 incluse, période pour laquelle Ifremer avait la charge de la totalité de la maîtrise d'œuvre, du prélèvement jusqu'à l'analyse des échantillons.

Ainsi, le LER Arcachon et le laboratoire PFOM-PI d'Argenton ont demandé à la cellule Quadrige de qualifier l'ensemble des données VELYGER de cette période à « Bon », avec comme commentaire de qualification « Qualification des données VELYGER suite à expertise du Laboratoire Ifremer Environnement Ressources d'Arcachon et du Laboratoire Ifremer Physiologie des Invertébrés d'Argenton – 2018 ».

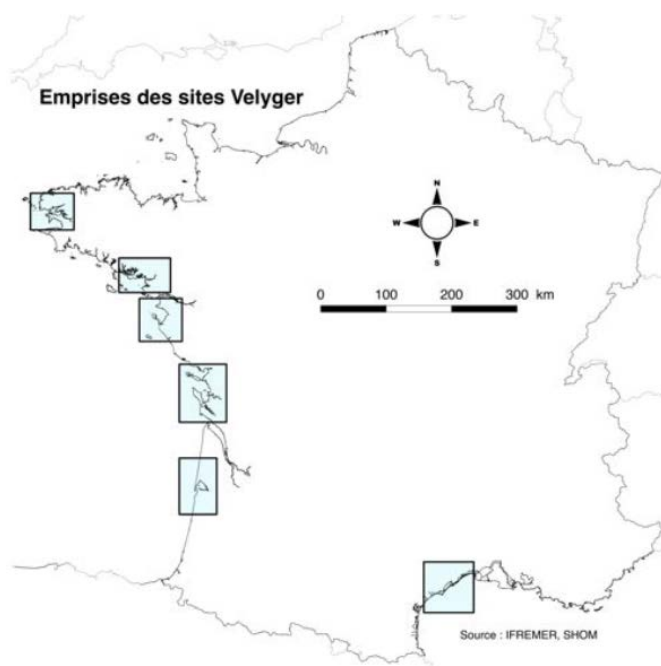


Figure 2 : Sites suivis par le réseau Velyger (source : rapport annuel 2017 : <https://wwz.ifremer.fr/velyger/Rapports-Annuels/Annee-2017>).

3.2 Qualification des données « Chimie »

3.2.1 L'« Imposex » : le programme RNOPHY

L'imposex est la réponse biologique à une contamination du milieu par le tributylétain (TBT) utilisé dans les peintures antisalissure des carènes de bateaux. Il provoque une masculinisation des femelles de certains gastéropodes marins (Imposex). Dans le cadre de la Directive Cadre sur l'Eau (le TBT est une des substances prioritaires de cette directive) et de la Convention pour la protection du milieu marin de l'Atlantique du Nord-Est (OSPAR), un suivi est réalisé chaque année depuis 2003.

Les données issues de cette surveillance sont bancarisées dans Quadriga dans le **programme RNOPHY**. Les données historiques (jusqu'à 2013 inclus) ont été qualifiées en 2014 par Didier Claisse, alors coordinateur du ROCCH. Fin 2017, les données 2014 et 2015 ont été qualifiées selon le même processus par Anne Grouhel qui occupe désormais le poste de coordinatrice du ROCCH (Figure 2).

La documentation décrivant ce processus de qualification a été rédigé en 2018 :

- Procédure de qualification des données du programme RNOPHY : elle décrit précisément les critères de sélection des données à qualifier et les règles de qualifications appliquées.
- Fiche « Qualification – Chimie – RNOPHY » : résumé en une page de la procédure. Elle ne contient pas d'aspects techniques.

Ces deux documents sont accessibles à tous publics sur le site de la cellule Quadriga (https://wwz.ifremer.fr/quadriga2_support/La-qualification-de-mes-donnees).

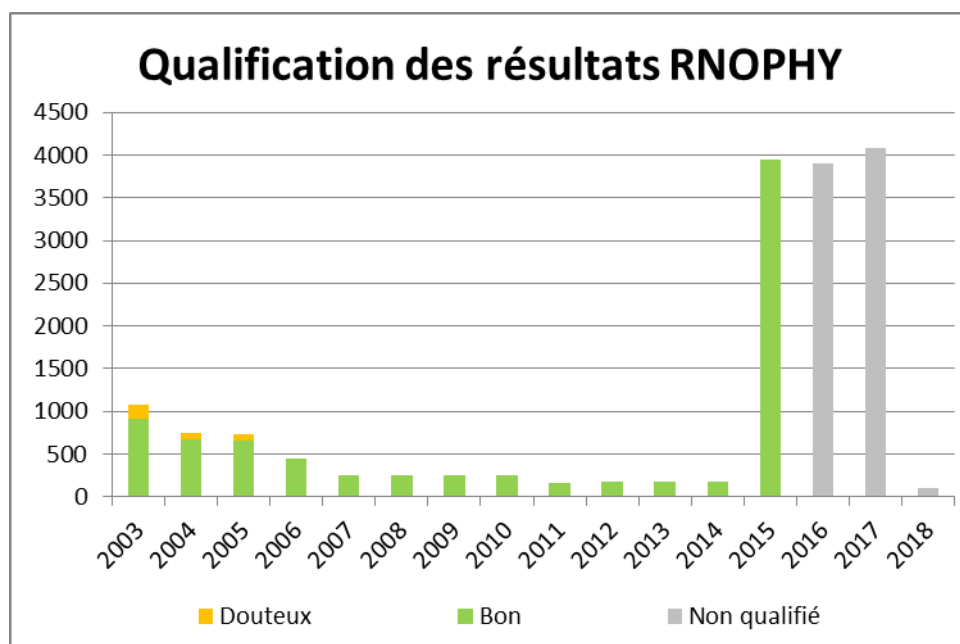


Figure 3 : Qualification des données RNOPHY (mars 2019). L'augmentation du volume de résultats à partir de 2015 est due à la saisie des données individuelles au lieu des moyennes.

3.2.2 Le RNO Matière Vivante

Le programme RNOMV (RNO Matière Vivante) correspond à la surveillance des contaminants chimiques dans les organismes marins (coquillages, poissons) pour l'évaluation des niveaux et des tendances.

Toutes les données RNOMV présentes dans la base de données Quadrige² au 1^{er} janvier 2014 ont été qualifiées par Didier Claisse avec le support de la cellule Quadrige² (qualification 2014).

En octobre 2017, la qualification des données allant de 2008 pour les résultats, de 2014 pour les Passages/Prélèvements/Echantillon, jusqu'à la fin du 1^{er} trimestre 2016 a débuté avec le support d'A. Grouhel (coordinatrice ROCCH). Il n'a pas abouti en 2018 par indisponibilité des experts pour qualifier les données.

3.2.3 Qualifications ponctuelles

En 2018 des travaux d'expertise réalisés par Ifremer/RBE/BE ont permis de qualifier des données historiques de HAPs :

- Expertise de février 2018 : 973 résultats 2016 qualifiés Douteux pour « Valeur appartenant à une série analytique douteuse (valeurs anormales par rapport aux valeurs aux autres dates / résultats) »
- Expertise d'avril 2018 : requalification des HAPs 2015-2016 avec réhabilitation de certains résultats qualifiés en Douteux en février :

Niveau initial avril 2018	Niveau final après expertise	
	bon	douteux
Douteux	240 → Résultats réhabilités	61 Déjà qualifiés
Non qualifié	129	13 « Valeurs très élevées, en rupture par rapport aux autres valeurs de la série (valeurs anormales par rapport aux valeurs aux autres dates / résultats) »

3.3 Qualification hydrologie Antilles 2007-2017

3.3.1 Critères de sélection des données

Les données hydrologiques qualifiées correspondent aux critères ci-dessous :

Programmes	Réseau REPHY MARTINIQUE-EAU-ETUDES
Paramètres	TEMP, SALI, TURB, TURB-FNU, OXYGENE, CHLOROA, NH4, NO3, NO2, NO3+NO2, PO4, SIOH, PHEO
Supports	« Masse d'eau, eau brute », « Eau filtrée »
Méthodes	Toutes sauf les analyses de pigments par HPLC (cela exclut les mesures de CHLOROA avec cette méthode).

3.3.2 Qualification automatique réalisée

La qualification automatique consiste à rechercher un certain nombre d'anomalies dans les données : ces anomalies correspondent à de potentielles incohérences dans les données qu'il faut vérifier et potentiellement corriger ou qualifier Douteux voire Faux selon les cas. Les anomalies « Hydrologie » sont décrites dans une fiche synthétique disponible sur le site de la Cellule Quadrigé¹.

A l'issue de cette qualification automatique, l'ensemble des données historiques est qualifié jusqu'en 2017 inclus à l'exception de quelques résultats 2003-2004 intégrés après les exercices de qualification de 2013 (Figure 4).

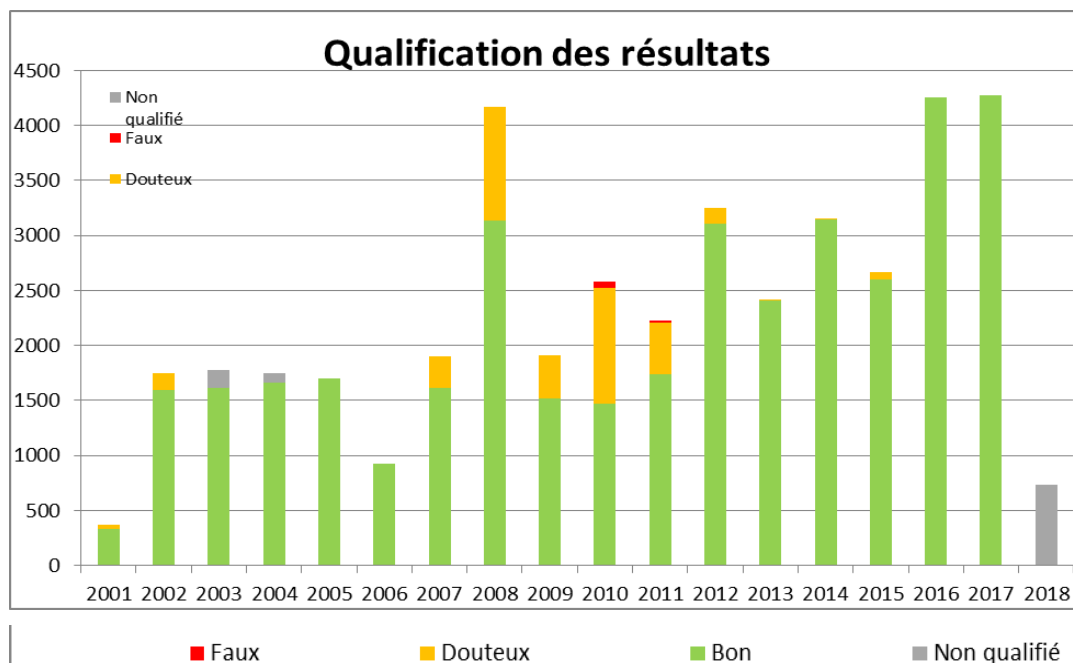


Figure 4 : Qualification des résultats Hydrologie Antilles à l'issue de la qualification automatique.

¹ https://www.ifremer.fr/quadrigé2_support/La-qualification-de-mes-donnees

Au total, les corrections apportées durant cette phase de qualification concernent les éléments suivants (détail par anomalie en Annexe 1) :

Nombre de corrections effectuées* :

- Référentiel (ajouts / mises à jour) : 1
- Ajout de données (lacunes identifiées et complétées) : 2 prélèvements recréés pour y déplacer des résultats existants
- Correction de valeurs de champs (métadonnées + résultats) :
 - Niveau prélèvement : 873
 - Engin de prélèvement : 1896
 - Préleveur : 595
 - Méthode des résultats : 2155
 - Valeur des résultats : 91
 - Précision des résultats : 13 834
 - Analyste : 1419
- Données supprimées : 31 résultats

* Ce chiffrage concerne les corrections effectuées par la Cellule Quadriga via les scripts R et des requêtes SQL. Les corrections effectuées par les qualificateurs Ifremer / Martinique ne sont pas comptabilisées ici.

Cette phase de qualification « automatique » a montré toute son utilité :

- Figure 5 : Sur 32 anomalies recherchées, 75% ont révélé des données à corriger / qualifier, sachant que l'anomalie HPAC-009 n'a révélé aucune anomalie car elles avaient été déjà corrigées dans le cadre de l'anomalie HPAC-002. Toutes les anomalies ayant révélé des données à vérifier ont suscité des corrections et/ou des qualifications : toutes ces anomalies sont utiles et permettent d'améliorer la qualité des données et l'identification de données douteuses ou fausses.
- Les anomalies détectées dans le cadre de la qualification automatique ont permis de d'identifier d'autres anomalies non détectables par les scripts, entraînant d'autres corrections dans les données.

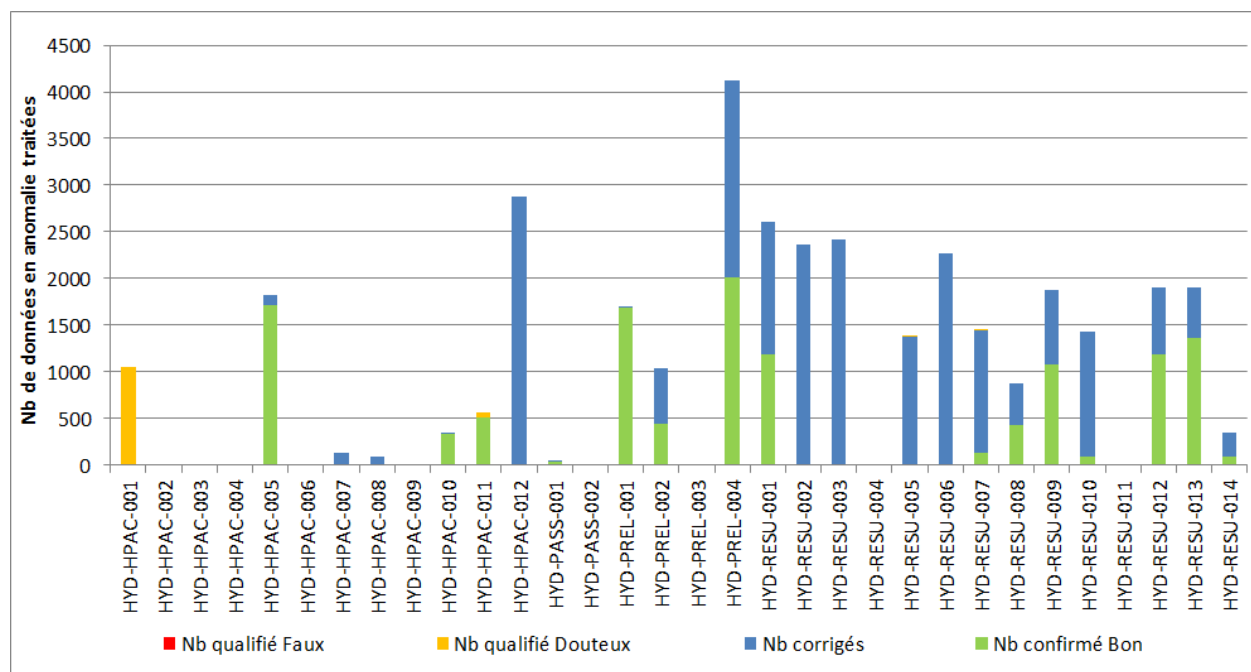


Figure 5 : Nombre de données en anomalie détectées par anomalie et résultat de leur traitement.

Le traitement des données en anomalie montre l'apport de la qualification puisqu'au total seules un tiers des données en anomalie détectées sont confirmées bonnes sans correction. Les deux tiers sont corrigées (et donc finalement qualifiées bonnes), seuls 3% des données sont qualifiées douteuses (Figure 6).

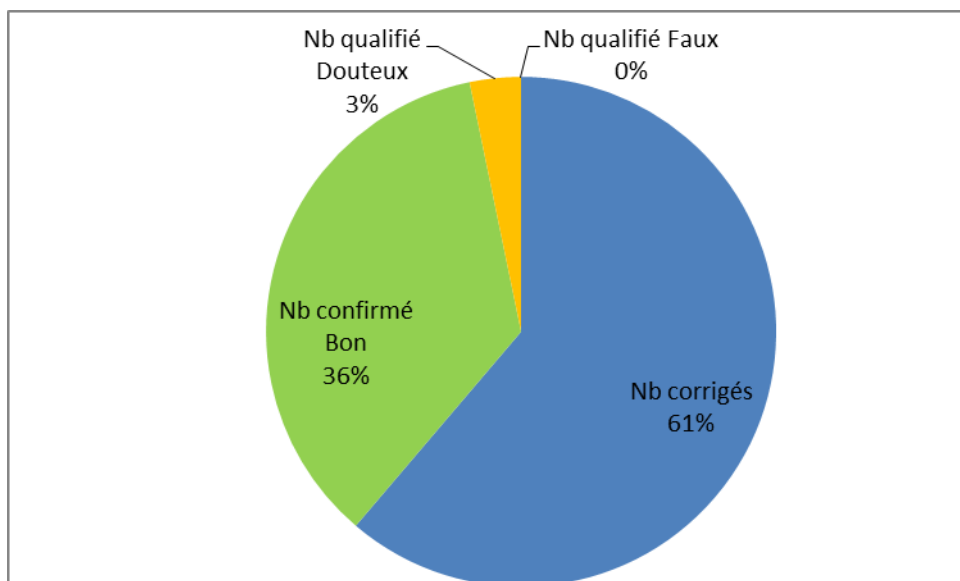


Figure 6 : Pourcentage de répartition des différentes résolutions des anomalies détectées en qualification automatique.

3.3.3 Corrections manuelles en marge de la qualification

Au cours du processus de qualification automatique il est souvent nécessaire de consulter les données dans l'application Quadrigé² ou de vérifier des cahiers de terrain, de paillasse ou rapports de laboratoires. Cela met souvent en évidence des erreurs non détectées par le processus automatique. Ces corrections sont effectuées après la qualification du « pack » (passage, prélèvement, échantillon ou résultat) en cours afin d'en tenir compte au pack suivant.

Dans le cas des données hydrologie Martinique et Guadeloupe 2007-2017, de nombreuses corrections ont été effectuées de cette façon :

- 101 Résultats de turbidité ont été corrigés : la méthode « in situ » a été remplacée par une méthode « sur échantillon » et le résultat a été déplacé sur l'échantillon correspondant
- Des heures de passages et de prélèvements erronées ont été corrigées (3 passages + 1 prélèvement)
- Des données bancarisées au mauvais endroit ont été déplacées (prélèvements, échantillons et résultats)
- Des doublons de passages / prélèvements voire de résultats dus à des erreurs de remplissage des fichiers Quadrilabo ont été supprimés ou corrigés
- 47 prélèvements ont été corrigés : niveau surface remplacé par fond (le code SANDRE du niveau « Surface » ayant été recopié sur toutes les lignes du fichier Quadrilabo alors que celui-ci contenait les données de surface ET de fond).
- 4 résultats de FLORTOT sans taxon associé ont été corrigés (taxon saisi) :

Lieu mnémo	Lieu libellé	Date	P	S	F	M	U	Valeur	Analyste	Saisisseur
124-P-016	Capesterre	29/05/2017	FLORTOT	Masse d'eau, eau brute	Sans objet	Comptage cellules au microscope eau	I-1	130	ASCONIT971	PARETO971
124-P-018	Caye à dupont	29/05/2017						4301	ASCONIT971	PARETO971
124-P-019	Main Jaune	26/07/2017						701,2623	ASCONIT971	PARETO971
124-P-022	Ilet à Christophe	18/05/2017						15681	ASCONIT971	PARETO971

3.3.4 Qualification experte Martinique

Nicolas Cimiterra et Jean-Pierre Allenou de l'Ifremer / Délégation Antilles de Martinique ont réalisé une qualification experte sur les données Martinique après leur qualification automatique. La qualification experte des données Guadeloupe est prévue pour 2019.

Cette qualification experte a été réalisée en utilisant une version des outils AURIGE développés par Ifremer/VIGIES adaptée aux cas des Antilles. Ces adaptations ont été réalisées avec le support de Mélanie Brun (Ifremer/VIGIES). Les sorties graphiques de ces scripts ont permis d'identifier des données aux valeurs exceptionnelles potentiellement douteuses ou fausses et de mettre en évidence des anomalies dans les séries de données (Figure 7).

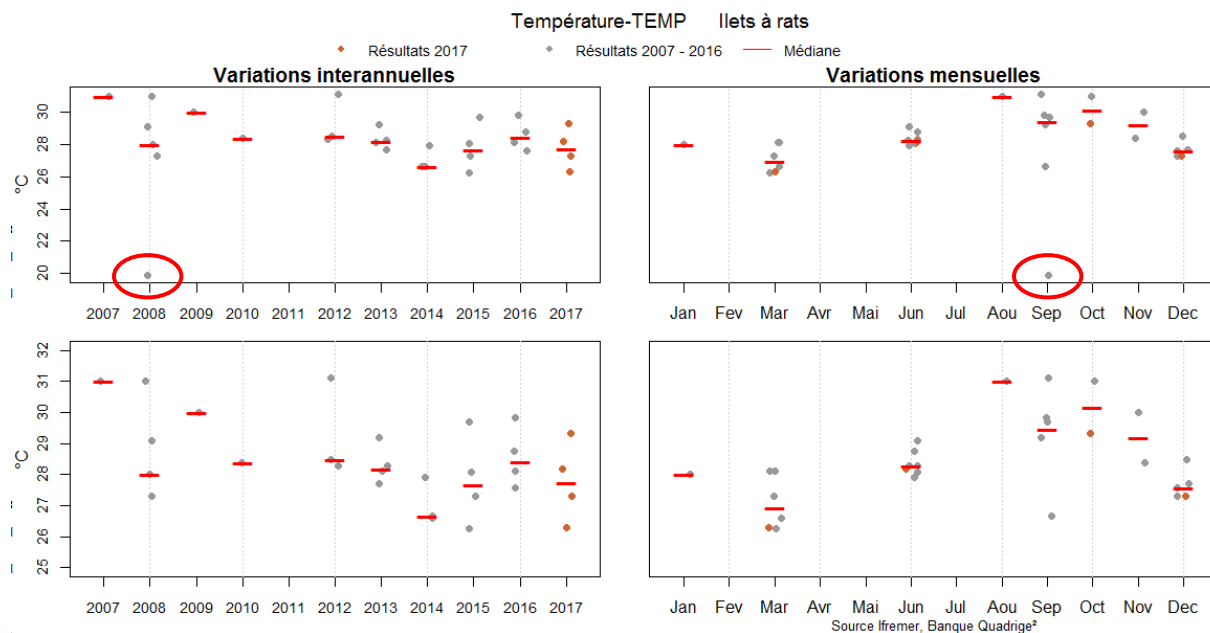


Figure 7 : Exemple d'identification de valeurs exceptionnelles qualifiées douteuses ou fausses via les scripts Aurige adaptés aux Antilles (Source : N. Cimiterra, 2018).

Ce travail a permis de qualifier 11 622 résultats qui n'avaient pas été qualifiés en qualification dite « automatique » (Tableau 3).

48% des données voient leur niveau de qualité confirmé après expertise.

Pour 51% des données le niveau de qualité est déclassé (détail des commentaires de qualification en Annexe 2).

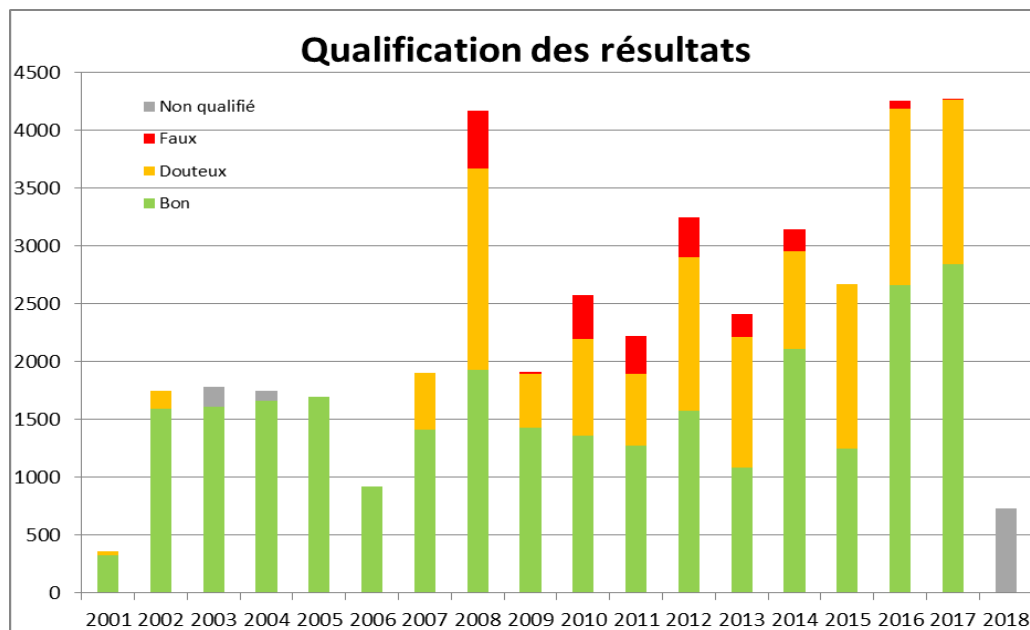
Pour 1% des données (48 résultats), le niveau de qualité est réévalué à la hausse :

- ⇒ Des résultats de température et de salinité avaient été initialement qualifiés douteux pour cause de problème de sonde, mais après vérification seuls les capteurs des autres paramètres étaient défectueux : ces résultats ont été qualifiés à Bon.
- ⇒ Des résultats de NO2 qualifiés Faux pour cause de méthode d'analyse inadaptée à l'eau de mer ont été reclassés à Douteux en référence au rapport « étude de la contamination des eaux côtières de la Martinique par les nutriments ».

Tableau 3 : Bilan des niveaux de qualité attribués en qualification automatique et experte.

Niveau de qualité		Qualification experte			Total
		Bon	Douteux	Faux	
Qualification automatique	Non qualifié	8521	3101		11622
	Bon	7972	8895	1432	18299
	Douteux	12	2123	559	2694
	Faux		36	36	72
Total		16505	14155	2027	32687

Au final, le niveau de qualité des données hydrologiques (paramètres qualifiés en qualification automatique hydrologique) Antilles est le suivant :



Les paramètres qui restent à qualifier sont :

- Les pigments phytoplanctoniques obtenus par HPLC (une partie a déjà été qualifiée car la méthode analytique et la valeur numérique ont été jugées fiables),
- Les analyses de nano- et picoplancton par cytométrie de flux,
- Les flores phytoplanctoniques par lecture sous microscope.

3.3.5 Zoom sur les données qualifiées Douteux / Faux

Depuis 2012 le taux global de résultats douteux et faux est de 1% des données (Figure 8). Pour les années 2007 à 2011 le taux dépasse 15% par an en raison des nombreuses données historiques dont certaines métadonnées manquent (qualifiées douteuses pour cause d'absence de métadonnées, notamment les méthodes et les engins de prélèvement employés).

Comme le montre la Figure 9, presque 80% des données qualifiées douteuses et fausses sur la période 2007-2017 le sont pour cause de manque de métadonnées : ces résultats restent utilisables pour des traitements avec précaution car une partie des informations sur la donnée est inconnue (méthode d'analyse notamment).

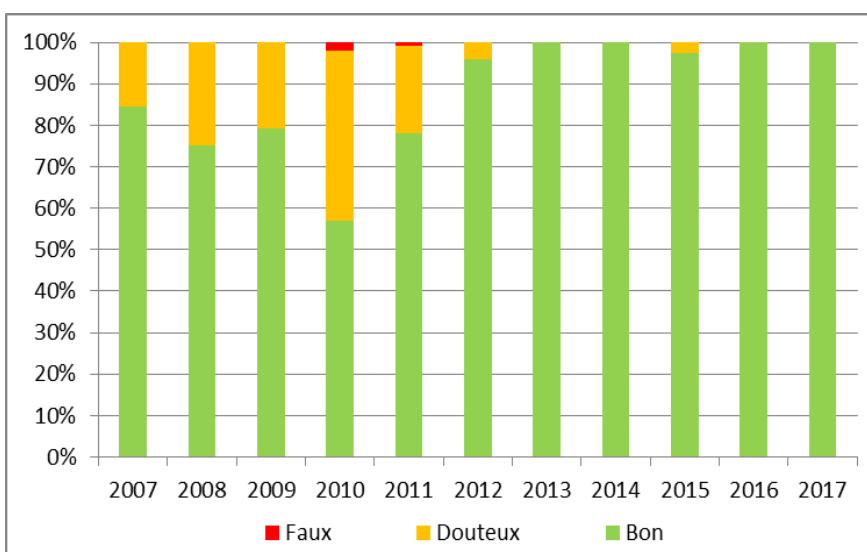


Figure 8 : Répartition des niveaux de qualité des données hydrologiques Antilles 2007-2017.

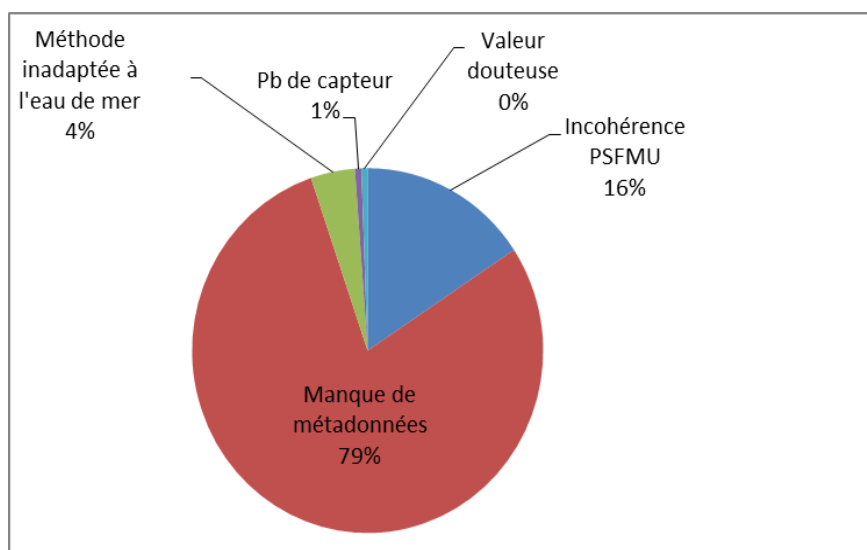


Figure 9 : Répartition des causes de qualification à douteux ou faux des résultats 2007-2017.

Les causes liées aux référentiels (incohérences de PSFMU) sont résolues dans les futures acquisitions de données (correction du référentiel Quadriga et diffusion de consignes de saisie actualisées auprès des producteurs de données – JP. Allenou 2018).

3.4 Qualification hydrologie métropole 2016

3.4.1 Critères de sélection des données

Les critères employés pour extraire les données hydrologiques métropole 2016 à qualifier sont les suivants :

Programmes	Réseau REPHY + réseaux régionaux + réseau historique RNO maintenu en Corse + Réseau de Suivi des Lagunes Méditerranéennes REPHY, REPHY-ETUDES, ARCHYD, RHLN, SRN, RNOHYD, RSLHYD
Lieux	Appartiennent aux zones marines métropolitaines*.
Paramètres	TEMP, SALI, TURB, TURB-FNU, OXYGENE, CHLOROA, NH4, NO3, NO2, NO3+NO2, PO4, SIOH, PHEO
Supports	« Masse d'eau, eau brute », « Eau filtrée »
Méthodes	Toutes sauf les analyses de pigments par HPLC (cela exclut les mesures de CHLOROA avec cette méthode) car il faudrait s'assurer que les règles de qualification des méthodes classiques sont aussi applicables avec les mêmes critères pour ces nouvelles méthodes.

*Remarque : une qualification des données hydrologiques Antilles ayant été réalisée en cours d'année 2018 sur les données de la période 2007-2017, et la même demande étant formulée en Océan Indien, les anomalies « Hors Pack » ont été recherchées sur toutes les données 2016, mais la qualification des passages, prélèvements, échantillons et résultats a été limitée par la suite à la métropole.

3.4.2 Avancement des travaux 2018

	HPAC-001	HPAC-002	HPAC-003	HPAC-004	HPAC-009	HPAC-010	HPAC-012	HPAC-005	HPAC-006	HPAC-007	HPAC-008	HPAC-011
	0	0	0	1	1	1	1	1	2	2	2	3
	Cohérence des PSFM	Conflits thématiques	Champs vides	Niveaux de saisie	Echantillons Mesures in situ	Cohérence CHLORO / PHEO	Saisies hors stratégie	Doublons	Cohérence préleveurs stratégies	Cohérence analystes stratégies	Validation données	Champs commentaires remplis
Métropole 2016	10 PSFMUs supprimés 5 PSFMUs gelés 2 PSFMUs ajoutés 29 146 résultats corrigés (PSFMU) 602 résultats qualifiés Faux	1584 prélèvements + 1519 échantillons détachés du REPHY	Pas d'anomalie	Pas d'anomalie	Pas d'anomalie	77 résultats corrigés (méthode)	278 résultats corrigés (méthode) 1 résultat déplacé	1791 résultats en doublon supprimés	57 préleveurs vides complétés dans les stratégies	52 analystes vides complétés dans les stratégies	53 passages validés (et leurs données filles avec)	27 résultats qualifiés Douteux / Faux

	PASS-001	PASS-002	PASS-003	PREL-001	PREL-002	PREL-003	PREL-004	PREL-005	ECH-001
	1	1	2	1	1	1	2	3	1
	Heure	Unité sonde	Contrôle + validation passage	Heure	Préleveur	Unité immersion	Combinaisons sonde-engin-niveau-imm-taille	Contrôle + validation prélèvement	Contrôle + validation échantillon
Métropole 2016	1 heure de passage corrigée	Pas d'anomalie	Pas d'anomalie	5 heures de prélèvement corrigées	896 préleveurs différents de celui de la stratégie : <ul style="list-style-type: none"> • 378 corrigés • 518 confirmés « Bon » 	Pas d'anomalie	En cours – à finaliser en 2019	A faire en 2019	A faire en 2019

	RES-001	RES-002	RES-003	RES-004	RES-005	RES-006	RES-007	RES-008	RES-009	RES-010	RES-011	RES-012	RES-013	RES-014
	1	2	2	2	2	2	2	2	2	2	2	2	2	2
	Analyste	TEMP	SALI	TURB	TURB-FNU	OXYGENE	CHLORO	PHEO	NH4	NO3+NO2	NO3	NO2	PO4	SIOH
Métropole 2016	A faire en 2019													

3.4.3 Bilan du processus 2018

Toutes les anomalies dites « hors pack » ont été traitées avec les coordinateurs des réseaux de surveillance et le soutien d'Anne Daniel (experte hydrologie Ifremer/DYNECO/PELAGOS Brest).

Les passages ont été qualifiés. Les prélèvements sont en cours de qualification fin 2018 comme le montre le paragraphe précédent.

Délais de la qualification

Le Tableau 6 montre les délais de traitement de chaque anomalie avec son nombre d'interlocuteurs sollicités pour le traitement de ces cas ou anomalies et le nombre de lignes d'anomalies détectées (une ligne pouvant correspondre à un passage, un prélèvement, un résultat, un PSFMU, une stratégie, un doublon, etc.).

Tableau 4 : Délais de traitement des anomalies Hydrologie Métropole 2016 traitées en 2018. Les dates de début et de fin sont colorées selon les périodes de travail (novembre-décembre 2017, janvier-février 2018, avril-mai 2018 et dernier trimestre 2018).

Anomalie	Ordre théorique	Date début	Date fin	Nb anomalies/cas détectées	Nb labos sollicités	Durée traitement
HYD-HPAC-001	1	23/11/2017	28/11/2017	15	1	5
HYD-HPAC-002	2	13/11/2017	13/11/2017	0	1	0
HYD-HPAC-003	3	20/12/2017	05/02/2018	4	1	47
HYD-HPAC-004	4	20/12/2017	20/12/2017	4	1	0
HYD-HPAC-009	5	20/12/2017	20/12/2017	0	0	0
HYD-HPAC-010	6	22/12/2017	17/01/2018	8	2	26
HYD-HPAC-012	7	22/12/2017	04/05/2018	1565	15	133
HYD-HPAC-005	8	18/04/2018	24/04/2018	3565	4	6
HYD-HPAC-006	9	20/12/2017	15/05/2018	90	3	146
HYD-HPAC-007	10	20/12/2017	04/05/2018	372	3	135
HYD-HPAC-008	11	19/04/2018	04/05/2018	221	2	15
HYD-HPAC-011	12	19/04/2018	02/10/2018	1785	3	166
HYD-PASS-001	13	18/10/2018	18/10/2018	3	1	0
HYD-PASS-002	14	18/10/2018	18/10/2018	0	0	0
HYD-PREL-001	16	25/10/2018	06/11/2018	3	1	12
HYD-PREL-002	17	25/10/2018	06/11/2018	602	1	12
HYD-PREL-003	18	25/10/2018	25/10/2018	0	0	0
HYD-PREL-004	19	14/12/2018	En cours	3245	12	

Il faut noter des **sauts importants dans les périodes de qualification** : après un mois de travaux en novembre et décembre 2017 il n'y a plus d'avancée jusqu'en avril – mai 2018. Puis s'en suit à nouveau un délai sans avancée visible jusqu'en octobre 2018.

En effet, la qualification dite « automatique » fonctionne par des échanges de fichiers : la cellule Quadriga génère des fichiers d'anomalie qu'elle transmet aux qualificateurs ou aux producteurs de données selon les anomalies. Ces personnes traitent les fichiers et renvoient la version complétée et/ou corrigée à la Cellule Quadriga qui les analyse : des allers-retours sont souvent nécessaires pour expliciter ou compléter les corrections des personnes sollicitées. Pour les anomalies des packs passage, prélèvement et résultats, l'ensemble des retours des producteurs de données sont fusionnés et envoyés au qualificateur pour validation / correction. Enfin le fichier validé par le qualificateur est renvoyé à la Cellule Quadriga qui intègre les corrections et qualifications en base de données.

Le nombre d'étapes du processus associé au nombre d'interlocuteurs et d'anomalies à traiter joue sur les délais de qualification. Selon les disponibilités de chacun, un délai de traitement de plusieurs mois est souvent constaté.

3.5 Préparation de la qualification des données REPOM Sédiments

3.5.1 Contexte

Quadriges accueille les données de surveillance du Réseau de suivi des Ports Maritimes depuis sa mise en place. Ce réseau de surveillance de la qualité des eaux et des sédiments portuaires a débuté dans les années 1990 et se poursuit aujourd'hui uniquement sur la matrice sédiment (suivis eau arrêtés en 2007). Il est coordonné par le Cerema (Marion Messenger et Mohamed El Fadili).

Dans Quadriges1 puis Quadriges2, la saisie des données était opérée par les SPELs. Cette saisie directe s'est révélée non efficace dans le temps. Peu de données ont réellement intégré la base par saisie directe entre 1997 et 2014. Une partie des données historiques non saisies a été reprise « en masse » dans Quadriges en 2014 (travaux C. Leguyader du Cerema et A. Huguet d'Ifremer).

Début 2015, un progiciel développé en VBA nommé QUADRISPEL est mis en production par le Cerema pour servir d'interface entre les saisisseurs et Quadriges2. Possédant un référentiel interne codifié SANDRE, QUADRISPEL propose des interfaces de saisie thématiques puis permet l'export des données au format Quadrilabo, intégré ensuite dans Quadriges. Le Cerema a réalisé un gros travail de reprise d'historique sous Quadrispel. Aujourd'hui, la très grande majorité des données REPOMS 2010 – 2017 sont intégrés à Q² et la saisie des nouvelles données se fait en routine.

En 2018, les données du REPOMS ont été identifiées comme utiles à **l'évaluation du Bon Etat Ecologique (BEE) réalisé au titre de la DCSMM**. Or, des problèmes de qualité de ces données ont empêché leur utilisation.

Afin de permettre une correction de ces données et l'historisation « propre » des données du REPOM, ainsi que de les rendre utilisables pour la prochaine évaluation DCSMM d'ici 2 ans, il est nécessaire d'entamer le processus de Qualification attendu dans le cycle de vie de la donnée.

Le Cerema a donc sollicité la Cellule Quadriges pour l'assister dans ce processus de qualification.

3.5.2 Programme de travail établi

Une première réunion de travail a eu lieu le 19 octobre 2018 pour évaluer les besoins du Cerema et présenter les outils actuellement opérationnels au sein de Ifremer/VIGIES. Il a été convenu de qualifier les données par tranches de dates suivantes :

- 2015-2017 : volume de données raisonnable, saisisseurs toujours en poste (facilité de récupération des informations), données utiles à la DCSMM pour les prochaines évaluations, les règles de contrôle identifiées comme nécessaires seront utiles pour les futures saisies en routine
- 2010-2017 : un seul programme « REPOMS », totalité des données saisies sous Quadrispel, séries temporelles utiles à la DCSMM
- < 2010 : données historiques plus difficiles à corriger car informations plus difficiles à retrouver, données moins utiles pour les évaluations DCSMM.

En respectant ces périodes, le programme de travail proposé est le suivant :

- 1) **Utilisation des scripts de qualification dite « automatique » volet « hydrologie »** : la thématique « hdro » a été préférée à celle du « RNO Sédiment » car cette dernière limite la recherche d'anomalie aux problèmes ciblés sur les campagnes RNO entièrement sous maîtrise Ifremer et dont les données ont été saisies par un unique saisisseur.

Les résultats de ces recherches d'anomalies permettront au Cerema d'évaluer l'ampleur des corrections à apporter et de les prioriser.

Ces résultats permettront au Cerema de proposer des actions lors du Comité de Pilotage REPOM de novembre 2018.

- 2) **Adaptation du logiciel Quadrispel par le Cerema** pour éviter de nouvelles anomalies :
 - a. Nettoyage du référentiel pour ne laisser que les éléments actifs (suppression des référentiels liés au REPOM Eau qui étaient utiles uniquement à la reprise de l'historique)

- b. Limitation des listes de valeurs possibles pour les saisisseurs (ex : niveau limité à « horizon 1 » et « couche superficielle (5-10cm) », plus de « Surface (0-1m) » ou « Fond »).
- 3) Un travail spécifique sur les unités de mesure est à réaliser : lors des saisies sous Quadrigé notamment, les valeurs numériques bancarisées n'ont pas toujours été cohérentes avec l'unité de mesure associée. Une **production de séries de graphiques** montrant des nuages de points de valeurs par couple paramètre / unité permettrait de déceler des anomalies. L'Ifremer/VIGIES pourrait aider le Cerema en développant des scripts R produisant ces graphiques et les tableaux associés.

3.5.3 Bilan des travaux 2018

Les scripts de qualification automatique « Hydrologie » ont été appliqués aux données REPOMS et ont donné les résultats suivants :

N°	Description	Commentaires
REPOMS-HPAC-001	Vérification de la liste des PSFMs (état actif / gelé, cohérence paramètre – support – fraction – méthode – unité) Cohérence PSFMs - analyste	La liste des PSFMUs permettra d'identifier des erreurs de saisie de type référentiel (erreur dans le choix de l'unité par exemple). Pour certains paramètres, la fraction analysée ne peut être que « Fract < 2mm ». Les paramètres DBT (dibenzothiophène dans Q ²) et DIBUT ont été confondus à la saisie → à uniformiser. Pour les organoétains (TBT, MBT, etc.), plusieurs unités sont possibles : or une seule unité était disponibles en saisie dans Q ² jusqu'en 2014. Les données saisies avant peuvent être soit corrigées, soit qualifiées Douteux « Incertitude sur l'unité de mesure ». De nombreuses méthodes étant des méthodes « internes », la cohérence du PSFMU et de l'analyste sera également vérifiée. Les unités de mesure utilisées sont en partie erronées (ex : mg/kg au lieu de µg/kg). Des sorties graphiques de nuages de points par paramètre / unité permettraient d'identifier les résultats avec une mauvaise unité. L'application de bornes à chaque paramètre / unité permettrait également d'identifier les valeurs ne correspondant pas à l'unité : ceci pourrait être vu en anomalies de résultats. Le nombre de PSFMUs différents par an est très élevé en 2015 (1344 PSFMUs différents pour moins de 600 les années précédentes). Il est possible que des erreurs de saisie soient à l'origine de cette démultiplication des PSFMs (cf. Annexe)
REPOMS - HPAC-002	Conflits thématiques : recherche des autres programmes liés aux données REPOMS, et des résultats d'autres thématiques rattachés aux données in situ REPOMS	Après vérification, seuls 10 passages compris entre 2007 et 2014 sont partagés avec les programmes REPOMO et DRAG. Y sont associés 15 prélèvements, 15 échantillons également partagés avec ces programmes et 149 résultats communs avec le programme DRAG. Il y a également 48 résultats appartenant au programme DRAG seul sur des passages communs DRAG – REPOMS.
REPOMS - HPAC-003	Champs vides : vérification que les champs qui n'ont pas lieu d'être remplis dans le cadre du RNOSED sont bien vides. Ont été contrôlés : Passages : observations terrain et habitat observé Prélèvements : Lot, taille de prélèvement et nombre d'individus Echantillons : taxon et groupe de taxon support, taille de l'échantillon et nombre d'individus	La liste des champs vides est à établir par le Cerema. Cette recherche devrait ramener peu d'anomalie, d'autant que la mise en place de QUADRISPEL limite les saisies aux seuls champs présents dans le progiciel et utiles à la thématique. Cette anomalie concerne donc plutôt les données historiques saisies manuellement dans Q ² .

N°	Description	Commentaires
REPOMS - HPAC-004	Niveau de saisie : vérification que toutes les PSMFs des stratégies REPOMS ont le même niveau de saisie (échantillon en l'occurrence), puis vérification que tous les résultats sont saisis au niveau prévu dans la stratégie	Normalement, toutes les données REPOMS sont saisies sur échantillon. Mais Quadrispel permet de choisir les niveaux de saisie PASS, PREL et ECHANT. Après vérifications : - 42 PSMFs ont des niveaux de saisie PREL dans des stratégies REPOMS - 3630 résultats saisis à un niveau différent de celui prévu dans la stratégie (cf. Annexe).
REPOMS - HPAC-005	Doublons de passages : vérification qu'il existe un seul passage par lieu – date.	3 « doublons » de passage ont été trouvés sur tout l'historique REPOMS (cf. Annexe)
REPOMS - HPAC-006	Doublons de résultats : vérification qu'il existe un seul résultat de mesure par lieu – date – niveau (horizon) – paramètre – fraction – méthode.	2015-2017 : 388 cas 2010-2014 : 1170 cas 2000-2007 : 137 cas
REPOMS - HPAC-007	Données non validées	2015-2017 : aucun cas 2010-2014 : 31 passages et 1272 résultats 2000-2007 : 400 passages et 13857 résultats Cf. Annexe
REPOMS - PASS-001	Heure de passage = 00:00:00 (à supprimer)	
REPOMS - PASS-002	Sonde nulle ou =0 ou >= 300m	Cf. Annexe : liste des valeurs de sonde possibles dans le REPOMS
REPOMS - PASS-003	Unité de sonde du passage <> « m »	Aucun cas (cf. Tableau 5)
REPOMS - PREL-001	Heure de prélèvement = 00:00:00 (à supprimer)	35 cas trouvés
REPOMS - PREL-002	Unité d'immersion « cm »	295 cas trouvés : à étudier
REPOMS - PREL-003	Cohérence Lieu – préleveur	
REPOMS - PREL-004	Combinaison engin – niveau – immersion – support : vérification de la cohérence des 4 valeurs	
REPOMS - ECH-001	Aucune anomalie recherchée	
REPOMS - RESU-001	Bornes par paramètre	Etablir des bornes par paramètre pour détecter les erreurs d'unité

Emeric Gautier a quant à lui développé des scripts R générant des graphiques des couples paramètres / unités et les tableaux associés (Figure 6). Ces éléments ont été fournis au Cerema le 14 novembre 2018.

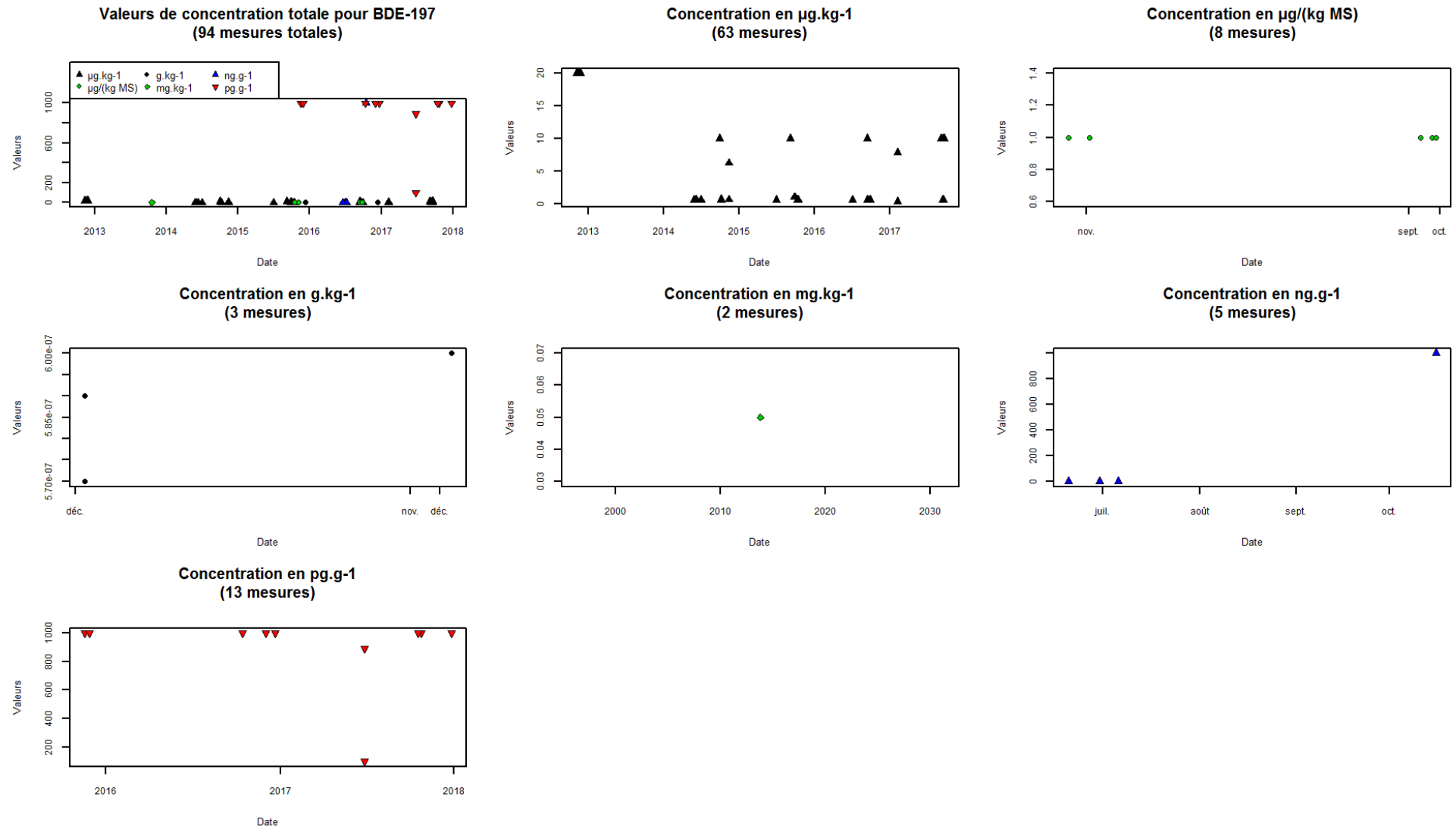


Figure 10 : Exemple de sortie graphique d'analyse des couples paramètre / unité REPOMS (E. Gautier - Ifremer/VIGIES).

3.6 Communication

3.6.1 Fiches qualification

Des fiches synthétiques décrivant le processus de qualification ont été rédigées et mises en ligne sur le site de l'assistance Quadrigé :

- Fiche de définition générale de la qualification des données Quadrigé
- Qualification des données « Hydrologiques »
- Qualification des données « Chimie – Imposex ».

Ces fiches ont été validées par l'équipe qualification VIGIES et par les qualificateurs (Anne Daniel et Nadine Masson pour l'hydrologie, Anne Grouhel pour la chimie).

3.6.2 Journées REPHY

La qualification des données a fait l'objet d'une présentation spécifique lors des journées REPHY nationales (journée interne, 29 novembre 2018). L'avancement du processus automatique pour l'hydrologie a été présenté et les consignes à respecter par les producteurs de données ont été rappelées.

Le besoin exprimé par les producteurs de données

Les agents des LERs se trouvant parmi les participants ont relevé un **intérêt pour recevoir les anomalies plus au fil de l'eau** : le processus de qualification portant sur les données de l'année n-2, il y a toujours un retard de correction. Les anomalies détectées ne sont corrigées que dans les données n-2 et non dans les données récentes. La quasi-totalité des saisies se faisant par duplication de la données précédemment acquise (et donc récente), si cette donnée contient des erreurs de saisie elles sont dupliquées elles aussi. La correction au fil de l'eau éviterait cette propagation des erreurs.

Une solution serait de mettre en place un **envoi trimestriel aux LERs des résultats de la recherche de toutes les anomalies simultanées dans les données du trimestre T-2** (Figure 11).

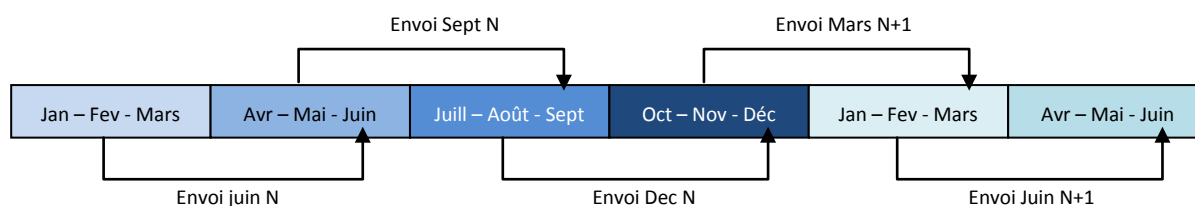


Figure 11 : Schéma du déroulement chronologique des envois d'anomalies aux LERs proposé.

Ainsi, en mars d'une année, juste avant la production des bulletins de la surveillance, les LERs recevraient les dernières anomalies de l'année précédente. Cela permettrait de corriger les données avant l'édition des bulletins et d'améliorer la qualité des données au cours du temps.

Comment le mettre en œuvre ?

Actuellement les scripts de qualification « automatique » tournent par pack et niveau de priorité (ex : pack « Passage » - niveau 1, puis pack « Passage » - niveau 2, etc.). Il faudrait les adapter pour qu'ils puissent tourner tous les uns à la suite des autres sans interruption sur une seule extraction Quadrigé.

Certaines anomalies se basent sur un fichier « support » listant les valeurs déjà connues pour certains champs et attribuant un statut « N – normal » ou « A – Aberrant » à chaque valeur ou combinaison de valeurs. Les fichiers d'anomalies générés flaguent les données comportant les valeurs « A ». Or, en cas de nouvelle valeur dans le fichier de données à traiter, les scripts s'arrêtent pour permettre l'envoi de ces nouvelles valeurs au qualificateur qui leur attribue un statut « A » ou « N », le fichier support est complété avec ces nouvelles valeurs et statuts et le script est relancé pour détecter les anomalies. Dans l'hypothèse d'un envoi trimestriel

de toutes les anomalies, pour éviter le blocage du processus en cas de nouvelle valeur dans le jeu de données, il faut que les scripts soient adaptés pour flaguer à « Inconnu » le statut des nouvelles valeurs.

Enfin, il faut automatiser l'envoi des fichiers générés aux producteurs de données. Un script permet actuellement l'envoi de mails à chaque LER avec des pièces jointes, mais il est limité à un pack – niveau de priorité et doit être lancé manuellement.

4 Bilan – discussion

Rapidité et efficience du processus de qualification

En 2018, les données hydrologiques Antilles d'une période de 10 ans ont été qualifiées en 2 mois (incluant des périodes de congés estivaux). La qualification des données hydrologiques métropole de la seule année 2016 a débuté fin 2017 et n'a pas été finalisée en 2018.

Dans le cas des Antilles, la qualification a été particulièrement rapide malgré un gros volume de corrections à apporter grâce :

- Au nombre restreint d'interlocuteurs : échanges bilatéraux entre la Cellule Quadriga et N. Cimiterra, avec relai vers les producteurs de données par Ifremer/DA et validation interne des corrections par JP Allenou (Ifremer/DA).
- A la disponibilité des ressources sur un temps retreint : afin d'assurer l'aboutissement du processus pendant la durée du contrat de N. Cimiterra, la Cellule Quadriga a priorisé ces travaux sur la période estivale.

En métropole, le nombre d'interlocuteurs par anomalie peut être important (jusqu'à 15 pour l'anomalie HPAC-012 sur les saisies hors stratégies - cf. Tableau 6). Il faut attendre le retour des 15 interlocuteurs pour pouvoir lancer les scripts des étapes suivantes. Puis le qualificateur doit examiner tous ces retours dans un fichier unique et éventuellement revenir vers les producteurs de données pour compléter et/ou corriger les corrections et qualifications apportées. Puis la Cellule Quadriga traite ce retour en l'intégrant dans la suite du processus. Les délais de traitement de chaque pack-niveau sont hétérogènes et peuvent être longs comme le montre le Tableau 5.

Tableau 5 : Comparaison des délais de traitement des anomalies entre le processus avec de nombreux interlocuteurs (métropole 2016) et le processus avec un seul interlocuteur (Antilles 2007-2017).

Pack – niveau priorité	Date d'envoi - Délai de traitement			
	Métropole 2016		Antilles 2007-2017	
PASS – 1	18/10/2018	7j	26/07/2018	1j
PREL – 1	25/10/2018	60j	26/07/2018	1j
PREL – 2	24/12/2018 au qualificateur 30/01/2019 aux producteurs	> 90j <i>En cours au 22/03/2019</i>	27/07/2018	51j <i>Incluant 30j de congés</i>
RESU – 1	/	/	11/09/2018	7j
RESU – 2	/	/	18/09/2018	25j <i>Finalisé le 04/10/2019</i>

L'espacement temporel entre les réceptions / envois de fichiers pose des problèmes à tous les acteurs de la qualification :

- Aux producteurs de données qui reçoivent des fichiers à des intervalles de temps éloignés : les consignes de remplissage des fichiers ne s'assimilent pas d'autant que la structure des fichiers est complexe et variable selon les anomalies,
- Aux qualificateurs : comme pour les producteurs les consignes de remplissage des fichiers sont difficiles à assimiler. Les réponses des producteurs étant étalées dans le temps, les échanges non synchronisés entre tous les correspondants sont difficiles à suivre,

- A la Cellule Quadriga : le déroulement technique du processus est difficile à intégrer car lancé trop épisodiquement. Les erreurs de remplissage des fichiers par les différents correspondants sont corrigées par E. Gauthier pour éviter des allers-retours chronophages avec les producteurs (ex : mélange des lignes d'un fichier par le producteur : restauration du fichier dans le bon ordre, introduction dans le fichier de qualification par le producteur de données d'autres corrections hors qualification : traduction de ces corrections supplémentaires dans un fichier séparé, puis correction du fichier de qualification pour qu'il soit lisible par les scripts).

La durée du processus de qualification automatique pose un autre problème : si des corrections sont apportées dans les données directement dans Quadriga pendant le processus de qualification, ces corrections risquent d'être écrasées lors de l'intégration de la qualification (les valeurs corrigées présentes dans les fichiers de données qualification sont inscrites en base avec le niveau de qualité). Plus le délai de traitement de la qualification est long, plus ce genre de conflit est probable. Pour résoudre ce problème il faudrait : (i) corriger le processus de qualification pour qu'il tienne compte de la date de mise à jour des données en base avant d'y appliquer ses propres modifications, (ii) réduire au maximum la durée de l'ensemble du processus de qualification.

Conclusion : le lancement du processus de qualification producteur de données par producteur de données semble plus efficace, quitte à le lancer sur une période plus grande (l'expérience des 10 années de données Antilles montre que le volume de données n'est pas un frein à la qualification).

Les apports de la qualification

Ce processus de qualification fait ses preuves en matière d'amélioration de la qualité des données :

- Des erreurs nouvelles sont détectées au cours du processus, erreurs non décelables par des scripts : dans le cas des Antilles, de nombreuses erreurs de structuration des fichiers Quadriga permettant l'import des données dans Quadriga ont été recensées,
- Ces anomalies sensibilisent les producteurs de données à la qualité de leurs données : les consignes de saisie sont mieux assimilées,
- Les consignes de saisie peuvent être consolidées / rappelées : ainsi JP. Allenou a profité de l'exercice de qualification pour actualiser et rediffuser les consignes de remplissage des fichiers Quadriga et N. Cimiterra a actualisé les listings référentiels fournis aux producteurs. N. Masson, assistante à la coordination du REPHY, a présenté les anomalies et consignes à respecter pour les éviter lors des Journées REPHY de novembre 2018.

Les perspectives 2019

La finalisation de la qualification des données hydrologiques 2016 métropole est la priorité en 2019.

L'Ifremer / Délégation Océan Indien a sollicité la Cellule Quadriga pour réaliser le même exercice qu'aux Antilles.

La qualification automatique des données Chimie devrait être relancée.

Le Cerema doit revenir vers la Cellule Quadriga pour corriger / qualifier des données selon le plan de travail établi en 2018.

Des réflexions doivent être menées pour réfléchir à la poursuite du processus, notamment pour l'hydrologie métropole : N. Masson en collaboration avec les LERs a identifié de nombreuses corrections à faire dans les données REPHY 2017-2018 : il serait donc prématuré de lancer la qualification de ces données avant de les avoir corrigées. En revanche de nombreuses données antérieures restent à qualifier, notamment sur la période 2013-2015. La question est de savoir comment relancer ces qualifications : pour tous les producteurs confondus ou par producteur de données ou par zone géographique restreinte ?

Pour les données récentes, il serait intéressant de mettre en œuvre un envoi trimestriel tel qu'évoqué au §3.6.2, à condition que les ressources le permettent (Cellule Quadriga + équipe valorisation de Ifremer/VIGIES).

A noter que lors du Comité de Pilotage du projet SIVS (Système d'Information pour la Valorisation et la Surveillance) dont fait partie Quadriga, la qualification des données a été identifiée comme une action non prioritaire malgré l'intérêt reconnu qu'elle présente.

Annexe 1 : Qualification automatique hydrologie Antilles 2007-2017 : Nombre de données détectées et corrigées par anomalie

Nb anomalies Nb corrigées / qualifiées	HPAC-001	HPAC-002	HPAC-003	HPAC-004	HPAC-009	HPAC-010	HPAC-012	HPAC-005
	0	0	0	1	1	1	1	1
	Cohérence des PSFM	Conflits thématiques	Champs vides	Niveaux de saisie	Echantillons Mesures in situ	Cohérence CHLORO / PHEO	Saisies hors stratégie	Doublons
2007-2017 Martinique	21 PSFMs "Aberrants" = 1043 résultats 1043 résultats qualifiés Douteux	Pas d'anomalie	Pas d'anomalie	Pas d'anomalie	Résolus lors des corrections HPAC-002	30 résultats 2 résultats corrigés (méthode) 28 confirmés Bon	1331 résultats 722 résultats stratégie corrigée 595 résultats méthode corrigée 14 résultats stratégie + méthode corrigées	1820 résultats 1717 à conserver (faux doublons) 32 résultats corrigés (heure passage) 31 vrais doublons supprimés 40 résultats saisis sur de mauvais passages (déplacés sur les bons PPE avec nécessité de saisie de 2 prélèvements)
2007-2017 Guadeloupe	7 PSFMs "Aberrants" = 10 résultats non qualifiés 10 résultats qualifiés Douteux	3 passages + données associées 3 passages corrigés (rattachement programme)	Pas d'anomalie	Pas d'anomalie	Résolus lors des corrections HPAC-002	305 résultats 305 résultats confirmés Bon	1544 résultats 1544 résultats corrigés (méthode)	Pas d'anomalie

Nb anomalies Nb corrigées / qualifiées	HPAC-006	HPAC-007	HPAC-008	HPAC-011
	2	2	2	3
	Cohérence préleveurs stratégies	Cohérence analystes stratégies	Validation données	Champs commentaires remplis
2007-2017 Martinique	4 préleveurs 4 préleveurs corrigés dans stratégies	124 analystes vides 124 analystes complétés dans stratégies	84 passages 2012 à 2014 84 passages validés	561 résultats avec commentaire prélèvement ou échantillon ou résultat 53 résultats qualifiés Douteux
2007-2017 Guadeloupe	Pas d'anomalie	Pas d'anomalie	Pas d'anomalie	Pas d'anomalie

Nb anomalies Nb corrigées / qualifiées	PASS-001	PASS-002	PASS-003	PREL-001	PREL-002	PREL-003	PREL-004	PREL-005	ECH-001
	1	1	2	1	1	1	2	3	1
	Heure	Unité sonde	Contrôle + validation passage	Heure	Préleveur	Unité immersion	Combinaisons sonde-engin-niveau-imm-taille	Contrôle + validation prélèvement	Contrôle + validation échantillon
2007-2017 Martinique	43 passages 42 passages heure confirmée vide 1 heure passage corrigée	Pas d'anomalie	Déjà corrigé HPAC-008	246 prélèvements 21 prélèvements corrigés (heure)	Pas d'anomalie	Pas d'anomalie	3137 prélèvements 1099 engins corrigés 47 niveaux corrigés Le reste confirmé Bon	Déjà corrigé HPAC-008	Déjà corrigé HPAC-008
2007-2017 Guadeloupe	Pas d'anomalie	Pas d'anomalie	Déjà corrigé HPAC-008	1455 prélèvements 1455 Confirmés OK (regroupements sous même passage)	1039 prélèvements 595 prélèvements corrigés (préleveur)	Pas d'anomalie	991 prélèvements 25 prélèvements confirmés Bon 797 engins corrigés 826 niveaux corrigés 255 immersions corrigées	Déjà corrigé HPAC-008	Déjà corrigé HPAC-008

Tranche de dates	RES-001	RES-002	RES-003	RES-004	RES-005	RES-006	RES-007	RES-008
	1	2	2	2	2	2	2	2
	Analyste	TEMP	SALI	TURB	TURB-FNU	OXYGENE	CHLOROA	PHEO
2007-2017 Martinique	24 résultats 20 corrigés (analyste)	1419 résultats 1417 corrigés (précision)	1476 résultats 1473 corrigés (précision)	Pas d'anomalie	1000 résultats 980 précision corrigées 4 confirmé Bon 16 qualifiés Douteux	1325 résultats 1325 corrigés (précision)	1154 résultats 1128 précisions supprimées 1 résultat corrigé (valeur et précision) 24 résultats confirmés Bon 1 résultat qualifié Douteux	869 résultats 390 précisions corrigées/supprimées 432 résultats confirmés Bon 47 valeurs corrigées
2007-2017 Guadeloupe	2582 résultats 1399 corrigés (analyste)	950 résultats 950 corrigés (précision)	943 résultats 943 corrigés (précision)	Pas d'anomalie	387 résultats 387 corrigés (précision)	950 résultats 950 corrigés (précision)	296 résultats 186 précisions supprimées 104 résultats confirmés Bon 6 valeurs nulles corrigées	Pas d'anomalie

Tranche de dates	RES-009	RES-010	RES-011	RES-012	RES-013	RES-014
	2	2	2	2	2	2
	NH4	NO3+NO2	NO3	NO2	PO4	SIOH
2007-2017 Martinique	1466 résultats 664 précisions corrigées /supprimées 801 résultats confirmés Bon 1 valeur corrigée	1027 résultats 1024 précisions corrigées /supprimées 3 résultats confirmés Bon	Pas d'anomalie	1488 résultats 668 précisions corrigées /supprimées 819 résultats confirmés Bon 1 valeur corrigée	1495 résultats 473 précisions corrigées /supprimées 1021 résultats confirmés Bon 1 valeur corrigée	Pas d'anomalie
2007-2017 Guadeloupe	415 résultats 129 précisions supprimées 271 résultats confirmés Bon 15 valeurs corrigées	396 résultats 305 précisions supprimées 81 résultats confirmés Bon 10 valeurs corrigées	Pas d'anomalie	409 résultats 36 précisions corrigées /supprimées 364 résultats confirmés Bon 9 valeurs corrigées	408 résultats 56 précisions corrigées /supprimées 344 résultats confirmés Bon 8 valeurs corrigées	345 résultats 254 précisions supprimées 91 résultats confirmés Bon

Annexe 2 : Commentaire de qualification des données bonnes déclassées à Douteux ou Faux.

Commentaire de qualification	Douteux	FAUX	Total
Donnée douteuse - se référer au rapport "étude de la contamination des eaux côtières de la Martinique par les nutriments"	5461		5461
Donnée Fausse, faisant partie d'un jeu de données très dispersées		673	673
Donnée Fausse, faisant partie d'un jeu de données très dispersées - Ensemble des données 2012 - 2013 - 2014 hors gamme - se référer au rapport "étude de la contamination des eaux côtières de la Martinique par les nutriments"		1	1
Donnée Fausse, faisant partie d'un jeu de données très dispersées		139	139
Donnée Fausse, faisant partie d'un jeu de données très dispersées - Ensemble des données 2012 - 2013 - 2014 hors gamme - se référer au rapport "étude de la contamination des eaux côtières de la Martinique par les nutriments"		7	7
Données anormalement basses sur l'ensemble des deux années 2010 et 2011	29		29
Doute suite à la mise en évidence du problème pour le paramètre salinité	154		154
Ensemble des données 2012 - 2013 - 2014 hors gamme - se référer au rapport "étude de la contamination des eaux côtières de la Martinique par les nutriments"		337	337
IDEM ChloroA - Méthode douteuse présentant des résultats incohérents	351		351
Les données du mois de juin 2013 sont aberrantes	45		45
Méthode douteuse présentant des résultats incohérents	1501		1501
Méthode douteuse présentant des résultats significativement différents de la méthode HPLC	349		349
Mise en évidence d'un problème de calibration sur la sonde - absence de rapport de métrologie	910		910
Température trop basse pour des eaux tropicales		1	1
Valeur aberrante	1	6	7
Valeur aberrante - changement de sonde multiparamètre du prestataire		47	47
Valeur aberrante - Doubte quant à la qualité du travail effectué par le producteur de données		154	154
Valeur aberrante sur les deux années 2010 et 2011 - absence de rapport de métrologie		45	45
Valeur douteuse car mesurée sur prélèvement d'eau	93		93
Valeur qualifiée à douteuse selon les remarques du rapport RNO 2016 de la Martinique	1		1
Valeur suspecte suite au changement de sonde de l'opérateur		22	22
Total général	8895	1432	10327