
Contrôle de qualité des données bouteilles (IML)

INFORMATIONS GÉNÉRALES

L'ensemble des informations présentées dans ce document est basé sur l'expertise et les travaux de la section de gestion de données à IML, qui depuis quelques années déjà compile les données bouteilles et examine à la loupe leurs qualités manuellement. Plusieurs procédures découlent directement des traitements utilisés par le National Oceanographic Data Center (NOAA) lors de la production de la base de données océanographiques mondiale (Conkright et al. 2002), et des algorithmes du manuel de l'UNESCO publié en 1990.

Le contrôle de qualité des données bouteilles permet de déterminer la validité des mesures de température, salinité, chlorophylle, oxygène dissous, nitrate, nitrite, phosphate et silicate d'un échantillon d'eau de mer. Il se compose d'un ensemble de tests effectués dans l'environnement Matlab.

Tout comme le contrôle de qualité des données CTD, le contrôle de qualité des données bouteilles se divise en 5 étapes :

Étape 1 : Contrôle des métadonnées importantes telles que le temps et la position.

Étape 2 : Contrôle des enregistrements d'un profil les uns par rapport aux autres.

Étape 3 : Comparaison des données avec une climatologie.

Étape 4 : Comparaison des profils d'une même mission.

Étape 5 : Visualisation des métadonnées et des données.

Toute métadonnée peut être modifiée, en particulier la coordonnée temps-espace sans quoi un profil est inutilisable.

Aucune donnée n'est modifiée par le contrôle de qualité lui-même. Un sémaphore de qualité est ajouté (voir section suivante) pour qualifier les données de bonnes, douteuses, erronées ou manquantes seulement pour les tests de l'étape 2. Si les données doivent être modifiées pour une raison ou une autre, ces modifications sont apportées en dehors du contrôle de qualité et le sémaphore de qualité doit être ajusté en conséquence.

Conkright, M.E., J.I. Antonov, O. Baranova, T.P. Boyer, H.E. Garcia, R. Gelfeld, D. Johnson, R.A. Locarnini, P.P. Murphy, T.D. O'Brien, I. Smolyar, and C. Stephens. 2002a. World Ocean Database 2001, Volume 1: Introduction. Edited by S. Levitus. NOAA Atlas NESDIS 42, U.S. Government Printing Office, Washington., D.C., 167 pp.

Unesco. 1990. GTSP real-time quality control manual. Intergovernmental Oceanographic Commission, Manuals and Guides no. 22.

DESCRIPTION DES SÉMAPHORES DE QUALITÉ INDIVIDUELS

Les tests de l'étape 2 ajoutent des sémaphores de qualité aux données de température, salinité, chlorophylle, oxygène dissous, nitrate, nitrite, phosphate et silicate. Le sémaphore de qualité est un entier compris entre 0 et 9. La signification des sémaphores de qualité est présentée dans le tableau suivant :

Sémaphore	Signification
0	aucun contrôle de qualité
1	la donnée semble correcte
2	la donnée semble incohérente par rapport aux autres données
3	la donnée semble douteuse

4	la donnée semble erronée
5	la donnée a été modifiée
6	réservé à un usage futur
7	la donnée est possiblement problématique (sémaphore temporaire, IML)
8	le contrôle de qualité a été fait par le producteur des données
9	la donnée est manquante

La priorité des sémaphores est croissante, le sémaphore 0 étant le moins prioritaire et le sémaphore 9 étant le plus prioritaire (sauf le sémaphore 7). Par conséquent, si un test a jugé une donnée douteuse (sémaphore 3) et que le test suivant juge la même donnée erronée (sémaphore 4), c'est le sémaphore 4 qui sera conservé. Le sémaphore 7 est une exception car il est temporaire.

SÉMAPHORE GLOBAL QCFF

Le sémaphore QCFF («*Quality Control Failed Flag*») est un sémaphore global qui permet de retracer la provenance du sémaphore de qualité. Il ne s'applique qu'aux tests de l'étape 2 et au test 3.6 du contrôle de qualité. À chaque test de cette étape est associé un nombre 2^x où x est un entier positif. Avant d'exécuter le contrôle de qualité, on attribue une valeur de QCFF de 0 à un enregistrement. Lorsqu'un test échoue, la valeur 2^x qui lui est associée est ajoutée au QCFF. On peut ainsi facilement retracer les tests échoués d'un enregistrement en décomposant le QCFF sous forme binaire. Si un sémaphore de qualité d'un enregistrement est modifié à la main, une valeur de 1 est ajoutée au QCFF.

Voici la liste des tests disponibles actuellement (cette liste anglaise est ajoutée à l'en-tête des fichiers de données lorsque le contrôle de qualité a été effectué ; la description de chacun des tests, dans la prochaine section, est en français) :

TEST	DESCRIPTION (QCFF)
Test 1.1	GTSP Platform Identification
Test 1.2	GTSP Impossible Date/Time
Test 1.3	GTSP Impossible Location
Test 1.4	GTSP Position on Land
Test 1.5	GTSP Impossible Speed
Test 2.1	Global Impossible Parameter Values (2)
Test 2.2	Regional Impossible Parameter Values (4)
Test 2.4	Profile Envelope (16)
Test 2.5	Constant Profile (32)
Test 2.7	Replicate Comparisons (128)
Test 2.8	Bottle versus CTD Measurements (TEMP, PSAL, DOXY) (256)
Test 2.9	Excessive Gradient or Inversion (TEMP, PSAL, NTRZ, PHOS) (512)
Test 2.10	Surface Dissolved Oxygen Data versus Percent Saturation (1024)
Test 3.5	Petrie Monthly Climatology (TEMP, PSAL)
Test 3.6	Brickman Monthly Climatology (NTRA, PHOS, SLCA) (2048)
Test 5.1	Cruise Track Visual Inspection
Test 5.2	Ratio and Profile Visual Inspection (station data)
Test 5.3	Replicates Visual Inspection (whole cruise data)
Test 5.4	Bottle Versus CTD Measurements Visual Inspection (whole cruise data)
Test 5.5	Ratio and Profile Visual Inspection (whole cruise data)
Test 5.6	Variable patterns with time (whole cruise data)

Si un enregistrement échoue le test 2.9, alors le QCFF sera de 512. Si en plus, le test 2.7 avait déjà été échoué, alors le QCFF sera de $512+128=640$.

DESCRIPTION DES TESTS DU CONTRÔLE DE QUALITÉ

Test 1.1 : Identification du navire

Ce test vérifie seulement si tous les profils de la mission ont été échantillonnés sur le même navire.

Test 1.2 : Date et heure

Ce test vérifie si la date et l'heure du début et de la fin du profil sont comprises à l'intérieur du temps alloué pour la mission.

Test 1.3 : Latitude et longitude

Ce test vérifie si la position du profil est possible c'est-à-dire si la position est comprise entre -90 et 90 de latitude ou -180 et 180 de longitude.

Test 1.4 : Station à terre

Ce test vérifie si la position du profil se trouve sur terre, en utilisant une carte détaillée de la côte de l'estuaire et du golfe du St-Laurent. Ce test est également fait pour les missions effectuées dans la région de la baie d'Hudson ou au large de la côte est du Canada, mais les cartes ne sont pas aussi détaillées.

Test 1.5 : Vitesse du navire

Ce test vérifie la vitesse du navire entre deux profils consécutifs. La vitesse de navire est calculée à partir de la position temps-espace de la mise à l'eau et la position temps-espace de la sortie de l'eau du profil précédent. Dans le cas où la position ou la date et heure de sortie de l'eau du profil précédent est manquante, le test utilise les coordonnées de la mise à l'eau du profil précédent pour déterminer la vitesse du navire. Le test compare la vitesse calculée à la vitesse de croisière du navire utilisé.

Test 2.1 : Valeurs impossibles globalement

Le test 2.1 vérifie si les données de température, salinité, chlorophylle, oxygène dissous, nitrate, nitrite, phosphate et silicate sont possibles globalement suivant les critères du tableau ci-dessous (tiré de WOD05, Conkright et al. 2002). Si une donnée est jugée impossible donc erronée, son sémaphore de qualité est alors remplacé par 4.

Code	Variable	Unité	Valeur minimale	Valeur maximale
TEMP	Température	°C	-2.5	35
PSAL	Salinité	(psu)	0	50
DOXY	Oxygène dissous	mL/L	0	11
CPHL	Chlorophylle	mg/m ³	0	50
NTRZ	Nitrate+Nitrite	mmol/m ³	0	515
NTRI	Nitrite	mmol/m ³	0	15
PHOS	Phosphate	mmol/m ³	0	4.5
SLCA	Silicate	mmol/m ³	0	250

Test 2.2 : Valeurs impossibles régionalement

Le test 2.2 vérifie si les données de température, salinité, chlorophylle, oxygène dissous, nitrate, nitrite, phosphate et silicate sont possibles régionalement suivant les critères du tableau ci-dessous (pour la région côtière de l'Atlantique du nord, WOD05; Conkright et al. 2002. La salinité a été modifiée pour refléter les conditions du Golfe). Si une donnée est jugée impossible donc erronée, son sémaphore de qualité est alors remplacé par 4.

Code	Variable	Unité	Valeur minimale	Valeur maximale
TEMP	Température	°C	-2.5	35
PSAL	Salinité	(psu)	0	35
DOXY	Oxygène dissous	mL/L	0	10
CPHL	Chlorophylle	mg/m ³	0	50
NTRZ	Nitrate+Nitrite	mmol/m ³	0	515
NTRI	Nitrite	mmol/m ³	0	15
PHOS	Phosphate	mmol/m ³	0	4.5
SLCA	Silicate	mmol/m ³	0	250

La région est délimitée par les coordonnées suivantes (longitude, latitude) :

(-56.0, 52.0), (-73.0, 49.5), (-73.0, 46.0), (-64.5, 46.0), (-62.3, 45.2), (-56.0, 48.2), (-56.0, 52.0)

Si la région d'étude est située à l'extérieur de ce polygone, le test n'est pas fait.

Conkright, M.E., J.I. Antonov, O. Baranova, T.P. Boyer, H.E. Garcia, R. Gelfeld, D. Johnson, R.A. Locarnini, P.P. Murphy, T.D. O'Brien, I. Smolyar, and C. Stephens. 2002a. World Ocean Database 2001, Volume 1: Introduction. Edited by S. Levitus. NOAA Atlas NESDIS 42, U.S. Government Printing Office, Washington., D.C., 167 pp.

Test 2.4 : Enveloppe du profil

Le test 2.4 vérifie si les données de température, salinité, chlorophylle, oxygène dissous et sels nutritifs se situent à l'intérieur des limites, présentées au tableau ci-dessous, par intervalle de profondeurs. La donnée est jugée douteuse si sa valeur ne se situe pas à l'intérieur de l'intervalle permis et son sémaphore de qualité est alors remplacé par 3. Les données préalablement jugées erronées ou manquantes ne sont pas considérées.

Code	Intervalle de profondeurs (m)	Unité	Valeur minimale	Valeur maximale
TEMP	0 – 50	°C	-2.5	35
TEMP	50 – 100	°C	-2.5	30
TEMP	100 – 400	°C	-2.5	28
TEMP	400 – 1100	°C	-2	28
PSAL	0 – 50	(psu)	0	35
PSAL	50 – 100	(psu)	1	35
PSAL	100 – 400	(psu)	3	35
PSAL	400 – 1100	(psu)	10	35
DOXY	0 – 30	mL/L	0	10
DOXY	30 – 200	mL/L	0	9
DOXY	200 – 1500	mL/L	0	8
CPHL	0 – 1500	mg/m ³	0	50
NTRZ	0 – 1500	mmol/m ³	0	515
NTRI	0 – 1500	mmol/m ³	0	15
PHOS	0 – 500	mmol/m ³	0	4.5
PHOS	500 – 1500	mmol/m ³	0.01	4.5

SLCA	0 – 150	mmol/m ³	0	250
SLCA	150 – 900	mmol/m ³	0.01	250

Test 2.5 : Profil constant

Le test 2.5 vérifie si les valeurs de température, salinité, chlorophylle, oxygène dissous et sels nutritifs d'un même profil sont identiques sur une partie de la profondeur du profil. Pour échouer le test, il faut qu'une variable ait exactement la même valeur à toutes les profondeurs. Son sémaphore de qualité est alors de 7 donc le profil devra être validé manuellement ultérieurement. L'exercice est répété pour tous les réplicats d'une variable. Les données bouteilles préalablement jugées douteuses, erronées ou manquantes ne sont pas considérées.

Test 2.6 : Point de congélation

La température de congélation est calculée à partir de la salinité et de la pression. Une valeur de température plus basse que la température de congélation correspondante est jugée erronée et son sémaphore est remplacé par 4. Les données de température préalablement jugées erronées ou manquantes ne sont pas considérées.

Test 2.7 : Comparaison des réplicats

Ce test compare les réplicats d'un échantillon entre eux. Il s'applique aux données de température, salinité, chlorophylle, oxygène dissous et sels nutritifs. Les écarts maximaux tolérés entre les réplicats d'une même variable sont inscrits dans le tableau ci-dessous; ces écarts ont été déterminés empiriquement en examinant plusieurs jeux de données. Tous les réplicats d'une variable qui échouent le test sont à vérifier ultérieurement, par conséquent leurs sémaphores de qualité sont remplacés par 7. Les données préalablement jugées douteuses, erronées ou manquantes ne sont pas considérées.

Code	Variable	Unité	Écart toléré
TEMP	Température	°C	0.01
PSAL	Salinité	(psu)	0.01
DOXY	Oxygène dissous	mL/L	0.5
CPHL	Chlorophylle	mg/m ³	0.5
NTRZ	Nitrate+Nitrite	mmol/m ³	3.5
NTRI	Nitrite	mmol/m ³	0.1
PHOS	Phosphate	mmol/m ³	0.5
SLCA	Silicate	mmol/m ³	4.0

Test 2.8 : Comparaison des données bouteilles avec leurs acolytes CTD

Ce test compare les données bouteilles aux données enregistrées par la sonde CTD. Il s'applique aux données de température, salinité et oxygène dissous. La chlorophylle n'est pas considérée parce que le capteur de fluorescence n'est actuellement jamais calibré. Les écarts maximaux tolérés sont inscrits dans le tableau ci-dessous. Tous les réplicats d'une même variable sont comparés aux données de la sonde CTD. Ceux qui échouent le test sont à vérifier ultérieurement, par conséquent leurs sémaphores de qualité sont remplacés par 7. Les données préalablement jugées douteuses, erronées ou manquantes ne sont pas considérées.

Code	Variable	Unité	Écart toléré
TEMP	Température	°C	0.1
PSAL	Salinité	(psu)	0.2
DOXY	Oxygène dissous	mL/L	1.0

Test 2.9 : Gradients de température ou salinité

Ce test est basé sur le *World Ocean Database 2001* (Conkright et al. 2002). Les gradients (si gradients positifs) ou les inversions (si gradients négatifs) verticaux de température, salinité, nitrate et phosphate sont calculés afin de déterminer s'ils excèdent les valeurs permises du tableau ci-dessous. Le gradient est obtenu par la différence entre une observation (V2) et l'observation précédente (V1) divisée par la différence de profondeur entre les deux observations. L'exercice est répété pour tous les répliqués d'une variable. Les valeurs manquantes sont temporairement remplacées par la moyenne des répliqués jugés corrects (sémaphore 1) et possiblement problématiques (sémaphore 7) de manière à pouvoir tout de même effectuer le test. Les observations qui échouent le test ont un sémaphore de 7. Les données préalablement jugées douteuses, erronées ou manquantes ne sont pas considérées.

Code	Variable	Unité	Inversion	Gradient
TEMP	Température	°C /m	-10	10
PSAL	Salinité	(psu)/m	-0.1	5
DOXY	Oxygène dissous	(mL/L)/m	ne s'applique pas	
CPHL	Chlorophylle	(mg/m ³)/m	ne s'applique pas	
NTRZ	Nitrate+Nitrite	(mmol/m ³)/m	-1.0	1.0
PHOS	Phosphore	(mmol/m ³)/m	-1.0	1.0
SLCA	Silicate	(mmol/m ³)/m	ne s'applique pas	

Conkright, M.E., J.I. Antonov, O. Baranova, T. P. Boyer, H.E. Garcia, R. Gelfeld, D. Johnson, R.A. Locarnini, P.P. Murphy, T.D. O'Brien, I. Smolyar, C. Stephens. 2002. *World Ocean Database 2001*, Volume 1: Introduction. Ed: Sydney Levitus, NOAA Atlas NESDIS 42, U.S. Government Printing Office, Washington, D.C., 167 pp.

Test 2.10 : Pourcentage de saturation en oxygène des eaux de surface

Ce test vérifie que le pourcentage de saturation en oxygène dissous en surface, c'est-à-dire entre 0 à 10 m, se situe entre 85 % et 150 %. Si la valeur d'oxygène dissous échoue ce test, un sémaphore 3 lui est attribué. Les données d'oxygène dissous préalablement jugées douteuses, erronées ou manquantes ne sont pas considérées.

Test 3.5 : Climatologie mensuelle de Petrie en température et salinité

Cette climatologie a été compilée par Petrie et al. (1996) pour le golfe du St-Laurent. Les températures, salinités et densités moyennes à des profondeurs fixes de 21 régions du golfe y sont calculées pour chaque mois ainsi que leurs écarts types respectifs. Ce test utilise cette climatologie pour déterminer la validité des observations d'une mission. Si l'écart entre les observations et la climatologie excède 3 écarts types, un avertissement est émis. La responsabilité revient donc à l'utilisateur de déterminer s'il doit rejeter un lot d'observations ou seulement ajouter des sémaphores à quelques observations en particulier. Le problème avec ce test est qu'une donnée ne peut pas être rejetée simplement parce qu'elle échoue le test. Il se peut que le profil reflète un événement particulier. Il se peut également qu'il y ait un problème de sonde qui rende les observations inutilisables.

Petrie, B., K. Drinkwater, A. Sandström, R. Pettipas, D. Gregory, D. Gilbert and P. Sekhon. 1996. Temperature, salinity and sigma-t atlas for the Gulf of St. Lawrence. Can. Tech. Rep. Hydrogr. Ocean Sci. 178, v+256 pp.

Test 3.6 : Climatologie mensuelle de Brickman en sels nutritifs

Cette climatologie a été compilée par Brickman et Petrie (2003) pour le golfe du St-Laurent. Les valeurs moyennes par mois et les écarts types des nitrates, phosphates et silicates y sont calculés pour 4 intervalles de profondeur dans 12 régions du golfe. Le test 3.6 utilise cette climatologie pour déterminer la validité des observations de chaque profil d'une mission. Si l'écart entre la moyenne des observations par intervalle et la

climatologie excède 3 écarts types, un avertissement est émis. La responsabilité revient donc à l'usager de déterminer s'il doit rejeter un lot d'observations ou seulement ajouter des sémaophores à quelques observations en particulier. Le problème avec ce test est qu'une donnée ne peut pas être rejetée simplement parce qu'elle échoue le test. Il se peut que les observations reflètent un événement particulier. Il se peut également qu'il y ait un problème de méthodologie qui altère la qualité des données. De plus, il y a des régions peu échantillonnées, surtout pendant certaines périodes de l'année. Les observations qui échouent le test ont un sémaophore de 7. L'utilisateur doit examiner les valeurs avec un sémaophore de 7 et déterminer quel sémaophore valide doit être attribué.

Brickman, D. and B. Petrie. 2003. Nitrate, silicate and phosphate atlas for the Gulf of St. Lawrence. Can. Tech. Rep. Hydrogr. Ocean Sci. 231, xi+152 pp.

Test 5.1 : Visualisation de la trajectoire du navire

Ce test permet de visualiser la trajectoire du navire et ainsi d'identifier des positionnements en erreur.

Test 5.2 : Visualisation des profils

Cette étape du contrôle de qualité est primordiale. Elle permet de voir, profil par profil, le résultat du contrôle de qualité. Toutes les observations de température, salinité, chlorophylle, oxygène dissous et sels nutritifs se distribuent sur 2 figures de la manière suivante :

- Diagramme TS des données de la sonde CTD aux profondeurs d'échantillonnage des bouteilles
- Température, salinité et O₂ de la sonde CTD aux profondeurs d'échantillonnage des bouteilles et données bouteilles en fonction de la profondeur
- Observations de chlorophylle, nitrate, nitrite, phosphate et silicate en fonction de la profondeur. La moyenne des réplicats jugés corrects ou possiblement problématiques (sémaophore 1 ou sémaophore 7) est tracé en trait plein.
- Rapports N:P et N:Si, en fonction de la profondeur. Les rapports sont déterminés à partir de la moyenne des réplicats jugés corrects ou possiblement problématiques (sémaophore 1 ou sémaophore 7).

Pour tous ces graphiques, les données bouteilles sont mises en évidence si elles ont été jugées possiblement problématiques (sémaophore 7) et si elles sont douteuses ou erronées (sémaophore 3 ou sémaophore 4). Aucun sémaophore n'est ajouté pour ce test. Les valeurs extrêmes sont examinées et les sémaophores sont ajustés au besoin.

Test 5.3 : Visualisation des réplicats

Le rapport entre une mesure et son réplicat devrait théoriquement donner 1. En pratique, c'est rarement le cas. Afin d'identifier de potentielles observations fautives, les réplicats de température, salinité, chlorophylle, oxygène dissous et sels nutritifs sont tracés en fonction de leurs premières mesures respectives pour l'ensemble de la mission. Les réplicats sont mis en évidence s'ils ont été jugés possiblement problématiques (sémaophore 7) et s'il sont douteux ou erronés (sémaophore 3 ou sémaophore 4). Aucun sémaophore n'est ajouté pour ce test. Les valeurs extrêmes sont examinées et les sémaophores sont ajustés au besoin.

Test 5.4 : Visualisation des données bouteilles en fonction des données obtenues avec la sonde CTD

Le rapport entre la valeur obtenue en laboratoire et celle obtenue avec la sonde CTD devrait théoriquement donner 1. En pratique, c'est rarement le cas. Afin d'identifier de potentielles observations fautives, les données bouteilles de température, salinité, chlorophylle et oxygène dissous sont tracées en fonction des données CTD correspondantes pour l'ensemble de la mission. Les données bouteilles sont mises en évidence si elles ont été jugées possiblement problématiques (sémaophore 7) et si elles sont douteuses ou erronées (sémaophore 3 ou sémaophore 4). Si la calibration d'un senseur de la sonde CTD est inadéquate, la relation avec les données bouteilles en sera nécessairement affectée mais les valeurs à l'écart de cette relation pourront tout de même être identifiées.

Test 5.5 : Visualisation de l'ensemble des profils de la mission

Les données bouteilles de température, salinité, chlorophylle, oxygène dissous, sels nutritifs, rapport N:P et rapport N:Si pour l'ensemble de la mission sont tracées en fonction de la profondeur afin d'identifier de potentielles observations fautives. Les données bouteilles sont mises en évidence si elles ont été jugées possiblement problématiques (sémaphore 7) et si elles sont douteuses ou erronées (sémaphore 3 ou sémaphore 4). Aucun sémaphore n'est ajouté pour ce test. Les valeurs extrêmes sont examinées et les sémaphores sont ajustés au besoin.

Test 5.6 : Visualisation du patron des données bouteilles en fonction de la profondeur, du temps et de l'espace

Ce test permet la visualisation simultanée des données bouteilles de chlorophylle, oxygène dissous, nitrate, phosphate, silicate, rapport N:P et rapport N:Si pour l'ensemble de la mission en fonction de la profondeur et du temps. Les données de profondeur sont d'abord tracées sous forme d'histogramme. Les données moyennes des variables à ces profondeurs y sont superposées. Seules les données jugées correctes (sémaphore 1) ou possiblement problématiques (sémaphore 7) sont utilisées pour calculer les moyennes. Ce type de graphique permet d'identifier les données qui s'écartent du patron général.

RÉSULTATS DU CONTRÔLE DE QUALITÉ

Les données bouteilles qui subissent ce contrôle de qualité doivent être réexaminées lorsque le sémaphore de qualité attribué à une donnée est 7. Ce sémaphore est temporaire. Il indique qu'il y a possiblement un problème avec la donnée. Ce peut être un problème de répliquats par exemple, et dans ce cas, il faut revoir tous les répliquats pour identifier le ou les fautifs et ajuster les sémaphores de qualité en conséquence. Une fois ce réexamen effectué, le sémaphore 7 ne devrait plus apparaître dans les jeux de données finaux.
