

Rapport de validation des données Pomme

3- Références

*Louis Prieur - Caroline Pocho
avec la participation de Jean Raunet*

Ce rapport fait référence à plusieurs reprises à des fichiers situés sur le CD2 de transmission à Marie Paul Torre en Novembre 2005. Quelques compléments ont été transférés vers le site proof directement en Décembre 2005. L'architecture de ce CD 2 est reproduite en Annexe 7

Ce rapport est la troisième partie du rapport sur le travail de la cellule de validation des données Pomme effectué au Laboratoire d'Océanographie de Villefranche sur Mer les deux précédentes parties sont

- 1- Etalonnage des capteurs de la CTD Rosette
- 2- Crossvalidation de donnée

3-

Préliminaires

Introduction

- 1- Nécessités d'un référencement et nature du référencement effectué**
- 2- Intérêt de la présentation choisie pour les fichiers référencés et ainsi validés**
- 3- Choix des types de référencement**
- 4- Procédures utilisées pour affecter un référencement à chaque type de données issues du site Pomme Web**
- 5- Inventaire des fichiers pomme web référencés ou non (résultats) fourni en Annexe**
- 6- Informations générales**
- 7- Comment utiliser les fichiers référencés**
- 8.- liste des annexes**

Préliminaires

Cet ensemble de rapports qui émane de la cellule de validation des données Pomme est intitulé validation des données. Dans cet ensemble un certain nombre de convention de vocabulaire a été adopté, tout en utilisant des mots d'usage courant d'une part et des néologismes d'autre part. Il convient donc d'en préciser le sens. Le sens utilisé n'a toutefois aucune prétention de généralité. Ensuite sera précisé le champ des données dont la cellule a effectué une validation d'une manière ou d'une autre. On a redécouvert dans ce travail systématique ce qui est connu depuis longtemps pour un nombre restreint de variables ou de données.

Vocabulaire

Validation, référencement, étalonnage, cross validation, harmonisation, contrôle, cellule de validation, tout terme se rapportant aux données, sont maintenant précisés.

Validation : Toute données d'observation doivent être validées. Cela relève du sens commun pour un scientifique. Lorsqu'elles sont acquises ou consignées, ce sont des données 'brutes', même lorsqu'elles proviennent d'un système d'acquisition élaboré muni d'un progiciel de traitement, et même lorsqu'elles sont éditées en unités officielles de mesure. On doit en effet réaliser ou contrôler leur étalonnage, évaluer leur précision, détecter les anomalies, les transcrire en unités en vigueur, donc les valider par rapport à des normes de référence, les rendre utilisables par tout scientifique compétent ne connaissant pas les détails de l'acquisition ou le déroulement de la campagne de mesure. Pour les données de notre discipline, il convient donc au minimum de les étalonner et de les 'géolocaliser'.

La validation comporte donc plusieurs subdivisions :

i- Validation par étalonnage, on utilisera ici le terme de calibration, bien que non strictement d'un français académique. Cette étape n'est évidemment pas réalisée pour toutes les variables par la cellule de validation. Les données transmises au site pomme web par exemple sont supposées être déjà étalonnées par les spécialistes du moins dans un certain délai après la campagne. Toutefois la cellule a réalisé les étalonnages des capteurs des sondes CTD utilisées (à l'exception de Pomme 0, réalisés par le LPO).

ii- Validation par référencement. Elle a pour but de s'assurer que les données sont correctement géolocalisées et ceci de manière cohérente avec le déroulement de la campagne d'observation. La cohérence suppose une précision, la même pour tous les jeux de données validées, sur la position géographique du navire, sur le temps (date et heure ou jour julien) et sur la profondeur de l'engin de mesure ou du prélèvement d'eau. **Cette validation est l'objet du présent rapport.**

iii- Validation par croisement de données ou d'informations, ou 'crossvalidation'. Un océanographe est capable de dire qu'une observation de température de 10°C à 4000 dbars (decakilopascals, pardon) de profondeur, n'est pas correct dans la zone Pomme. Il se réfère à une connaissance externe à la campagne, obtenue à partir d'atlas ou d'observations antérieures. Clairement une erreur de transcription, d'étalonnage ou d'autre chose est venue interférer avec la saisie initiale. Un autre type de croisement d'information est fondé sur la variabilité attendue ou observée du milieu ce sera alors une cross validation statistique. Enfin on sait que les variables ne sont pas indépendantes dans le milieu océanique, en particulier en profondeur : Relation NO₃-PO₄, salinité - alcalinité, diagramme Theta - S, etc. ; le type de cross validation n'est pas figé, et demande l'intervention d'un scientifique connaissant la zone

d'étude et les variables mesurées. Le résultat de tout croisement de données se traduit par une estimation selon un code précisé des données alors qualifiées de bonnes, anormales ou exceptionnelles, douteuses, aberrantes.

Malheureusement la situation n'est pas aussi simple et les validations ne sont pas indépendantes, dans le sens où l'ordre logique : étalonnage, référencement, cross validation ne peut pas toujours être respecté. Par exemple, pour 'étalonner' le signal issu d'un capteur d'oxygène, on a besoin de données de dosage sur des prélèvements. Avant d'effectuer la calibration on a donc besoin de valider par croisement les données acquises par la sonde et celles dosées sur des bouteilles de prélèvements. Des aller-retour entre validation par croisement de données et par étalonnage sont quasiment inévitables avant de parvenir à un lot de données validées et présentables.

Référencement : Par ce terme il est entendu que chaque donnée d'une campagne est associée à une position géographique, une date, une heure et éventuellement une profondeur. Lorsque les données sont transmises au site web de la campagne, ces informations revêtent inévitablement des formes, des précisions, des manques différents selon le scientifique en charge de ces données. L'objet de ce référencement est donc de géolocaliser d'une manière uniforme chaque donnée, sans modifier la forme de la donnée, ni porter une quelconque estimation sur la qualité ou la cohérence de la donnée.

Cross validation : lors d'une tâche de validation croisée, 'cross validation', des relations statistiques sont utilisées soit pour chaque variable, soit entre variables de manière à détecter des valeurs anormales, qui sont ensuite trouvées, aberrantes, exceptionnelles ou douteuses. Les valeurs normales ne se distinguent pas dans un nuage de points. Les procédures utilisées sont décrites dans le rapport sur la cross validation.

Cellule de validation

Louis Prieur et Caroline_Pocho (Juin –Septembre 2002 et Janvier 2003 - Décembre 2005) au Laboratoire d'Océanographie de Villefranche_sur Mer avec la participation initiale de Jean Raunet jusqu'en Décembre 2002

Principes respectés

- Pas d'élimination de données, poses de flag pour validation croisée
- Traçabilité : même structure que le site Pomme Web, nom explicite de fichier, retour possible aux expérimentateurs dans leur format, rapports journaliers de Caroline Pocho
- Le travail de validation est effectué sur des lots constitués à date de téléchargements ; des mises à jours sont éventuellement effectuées.
- Contrainte minimale : le format libre à discrétion des fournisseurs de fichiers sur le site web entraîne un éventuelle tâche de mise en forme

Champ de la validation exercé par la cellule de validation

Le répertoire 3-DONNEES_TRAITEES contient les fichiers de données validées par une des différentes méthodes pertinentes présentées dans le rapport.

Les données 'non traitées' telles que :

Temps réel M Assenbaum

Résultats de modèle
Données de flotteurs et d'ADCP
Pomme 0 (pas de fichier source de référencement)
Flux de surface

par exemple n'ont pas fait l'objet de validation, car elle avaient leur propre validation de géolocalisation complète. Un inventaire complet au 15-11-2005 des données validées ou non selon le découpage du site Pomme Web est donné en Annexe 4 et a été transmis à MP Torre :
INVENTAIRE_FICHIERS_REFERENCEMENT.doc

Nom des fichiers : dans ces rapports, il est fait référence à des noms de fichier en les mettant en italique avec leur extension. Ces fichiers ont été rassemblés sur un CD de transmission à MP Torre au 15 Novembre 2005. Une arborescence de ce CD est donnée en annexe. Certaines transmissions partielles de fichiers ont été effectuées au site Pomme Web et aux responsables sans attendre la date précédente. Toutefois les dates de transmission restent importantes dans la mesure où certaines modifications ont pu être effectuées. D'une manière générale les noms des fichiers ont alors été changés, malgré l'inconvénient de noms à rallonge. Une annexe dédiée à l'historique des transmissions a été construite.

Introduction

Pour une expérience comme Pomme, concernant plusieurs campagnes de plusieurs navires sur une même zone durant une année, ayant des aspects multidisciplinaires, faisant appel à plusieurs laboratoires et équipes et conduisant à des données météorologiques, physiques, chimiques et biologiques d'aspects très hétérogènes du point de vue de l'acquisition, de la fréquence et du type de données, il est indispensable, de s'assurer de la cohérence de la géolocalisation des données et de leur référence temporelle. Comme ce travail de contrôle de la géolocalisation des données demande en général la connaissance du trajet du navire et de son calendrier de travaux, il est souhaitable que ce contrôle soit fait par le chef de mission ou toute personne ayant l'ensemble des informations pertinentes. Cette tâche est clairement identifiée en raison de l'hétérogénéité des données et du nombre de scientifique intervenant pour alimenter la base de fichier des données d'une expérience de biogéochimie. Sur un nombre restreint de variable telle que l'analyse satellitaire de la SST par exemple, cette géolocalisation est intrinsèque au traitement de donnée et ne fait pas l'objet d'un travail spécial. Il n'en est pas de même pour une campagne de biogéochimie, intrinsèquement pluridisciplinaire.

Même si le travail exécuté ici concerne le domaine restreint des données obtenues en mer dans le cadre des navires Atalante ou Thalassa pour la partie biogéochimique essentiellement, il est relativement complexe en raison de la multiplicité des paramètres concernés et leur aspect hétérogène. Le principe retenu a d'abord été de situer ce travail en aval du traitement de données primaires ou brutes, c'est-à-dire que les différents responsables de chaque secteur de données ont fourni des fichiers de leurs résultats en format libre au site Pomme Web mise en place au LODYC (LOCEAN). A différentes dates, ces données ont été chargées par la cellule de validation afin d'effectuer les travaux de référencement, lesquels ont pour but de d'assurer une localisation unique dans le temps et dans l'espace de chaque donnée, localisation qui doit être en conformité avec la route du navire et avec la séquence temporelle des travaux réalisés. Parfois dans les fichiers reçus, cette géolocalisation est assurée, mais pas nécessairement, et pas toujours avec la précision souhaitée; quelques erreurs peuvent être présentes, malgré le professionnalisme des responsables (y compris nous-même !). Dans tous les cas il s'avère nécessaire de se repérer par rapport à des trames spatiotemporelles spécifiquement adaptées au référencement unique des données.

On aura compris que ce travail de contrôle et d'homogénéisation en quelque sorte est une étape obligée vers la mise en banque de données validées et sert son utilisation ultérieure.

Il y a quelques années, avant les ordinateurs, il était classique en océanographique d'établir de tableaux des données de manière à les croiser pour un usage scientifique de recherche. Maintenant le nombre de données s'étant considérablement accru, chacun traite éventuellement plusieurs paramètres à sa convenance selon des précisions qui varient selon l'usage. Ce contrôle de cohérence des différents tableaux fournis est donc indispensable et constitue une validation en elle-même.

Ce qui suit concerne exclusivement un rapport sur la validation par référencement, donc sur la géolocalisation précise et contrôlée des données. Cependant au cours du travail de contrôle des géolocalisations déjà présentes dans les fichiers originaux ou d'ajout à ces fichiers d'une géolocalisation complète, il est apparu utile de rendre les fichiers référencés portant sur des variables issues des prélèvements sur la rosette compatibles entre eux du point de vue de leur présentation. Le référencement effectué conduit donc à des fichiers pour les données de la rosette aisément juxtaposables pour effectuer des croisements entre variables appartenant à des fichiers différents. La conséquence est que la présentation des fichiers est homogène, mais peut être différente des fichiers originaux. Toutefois les valeurs des variables dans les fichiers résultats ne sont pas modifiées et, *dans cette étape de validation par référencement*, aucun flag n'a été imposé. Signalons à ce point que les méta informations du site Pomme web, non reproduites dans le CD de transmission restent valides pour les fichiers référencés en ce qui concerne la méthode de dosage ou de traitement voire de contrôle, mais le repérage des colonnes, si décrites, dans les méta-fichiers est modifiée par le référencement. En revanche ces colonnes sont repérées par le même nom que dans le fichier original et elles sont également indiquées dans le fichier compte-rendu de Caroline Pocho. Ainsi le principe de traçabilité est respecté et permet à un fournisseur de données de retrouver les données pour un contrôle ultérieur éventuel.

1- Nécessités d'un référencement et nature du référencement effectué :

1.1- Pourquoi est-il nécessaire de contrôler la géolocalisation des données issues des campagnes de biogéochimie ?

Tout océanographe sait qu'il faut consigner l'heure et la position d'une donnée. Le problème n'est pas là. Dans la pratique, à bord d'un bateau, on dispose de plusieurs heures (heure locale, heure TU différents affichages), de plusieurs positions (navigation intégrée, plusieurs GPS). Quelle heure notée, celui du début de la station, du prélèvement ? Chacun a des exigences particulières. De plus les affichages ne sont pas toujours à jour et les erreurs de transcription ou d'acquisition sont fréquentes. Il est donc indispensable au niveau de la validation de définir une heure officielle et une position officielle cohérente. Si ce n'est pas fait dans une base de fichier, croiser des données ayant des géolocalisations de précision différentes n'est pas toujours chose aisée. Tel prélèvement a-t-il été effectué sur la même bouteille que tel autre ? La réponse à cette question pertinente n'est pas chose aisée sans validation. Heureusement la contrainte généralement imposée par le chef de mission est de toujours noter le numéro de station et le numéro de bouteille. Il est donc nécessaire de transcrire ces références en géolocalisation par rapport à une heure et position officielle.

1.2- Géolocalisation validée et harmonisée : Les chercheurs qui ont fourni des données au site web Pomme, ont souvent localisé leur données, mais la présentation de cette localisation n'était donc pas imposée ni la précision. Leur localisation doit donc être contrôlée et, en cas d'absence totale ou partielle), précisée. La géolocalisation complète minimale adoptée est :

- date et heure TU (jour julien 2001, jour 1 à 0h00 = 1.00000) du prélèvement pour les continus de surface ou du début du profil CTD pour les données rosette
- position géographique en Latitude et Longitude en format degrés et fraction de degré (5 chiffres décimaux)
- Pour les données acquises en station :
 - le n° de station (ex : 2010 pour la station 10 de Pomme 2)
 - le n° de la bouteille de prélèvement utilisée
 - la profondeur officielle en déci bars pour les données issues de la rosette de bouteilles de prélèvement.

Le n° de station et le n° de bouteille a servi de base de référencement pour les données de prélèvement (rosette) ; ce repérage était la seule contrainte a priori imposée, en référence au principe de contrainte minimale. Inévitablement des exceptions ont été trouvées et elles ont été consignées dans les rapports journaliers joints à tout fichier validé.

Toujours pour les données en station, les choix suivants ont été pris :

1.3- Le référencement est un complément explicite d'information des fichiers individuels. Pour respecter le principe de tracabilité, le référencement officiel est *juxtaposé*, et placé dans les premières colonnes du fichier référencé, aux données de dosage ou de sondes même dans le cas où une géolocalisation partielle ou

complète était présente dans les fichiers originaux. Dans ce cas également deux localisations sont donc trouvées dans le fichier, mais une seule est officielle. Aucune comparaison systématique n'a été effectuée sur la correspondance entre ces deux localisations, l'objet d'une telle comparaison n'étant pas pertinent.

1.4- Le travail de référencement a été effectué par leg. Le résultat est donc un fichier par leg pour chaque groupe de variables du site Pomme web et quel que soit le nombre de données effectives dans le fichier ; le nombre de ligne des fichiers pour un leg donné est le même pour toutes les variables. Les fichiers des différents groupes de variables sont donc homogènes et comparables par concaténation directe des colonnes. Le nombre de ligne est défini par le nombre d'item constitué par l'assemblage nombre de station nombre de bouteilles déclenchées pour chaque station. Un tableau est joint en Annexe avec les exceptions non détaillées ici.

2- Intérêt de la présentation choisie pour les fichiers référencés et ainsi validés

La présentation adoptée rend aisée le croisement entre deux données appartenant à des fichiers référencés différents. Un fichier rosette unique aurait par exemple été réalisable, mais peu commode d'emploi en raison d'un nombre considérable de colonnes (plus d'une centaine) dont certaines avec seulement une vingtaine de données. La traçabilité en aurait été affectée. Un défaut de la méthode choisie est cependant une redondance d'information éventuelle (plusieurs valeurs de T ou de S pour un même référencement selon la nature des fichiers originaux. Mais une seule valeur de T et de S officielle est à considérer, celle issue de la sonde CTD dans les fichiers cross validés et ensuite référencés. Afin de mieux guider un utilisateur des données référencées des exemples seront présentés au paragraphe 7 ci-dessous 'comment utiliser les données référencées'. Un avantage de la présentation retenue est aussi que l'intégration des résultats en base de donnée ou dans un système d'information géographique est rendu aisée.

Le groupage des données référencées respecte le groupage du site Pomme Web pour traçabilité et repérage, malgré donc des redondances signalées ci-dessus.

Le fichier reproduit en Annexe 4:

INVENTAIRE_FICHIERS_REFERENCEMENT.doc

rappelle les contenus de ces groupage ainsi que la correspondance avec les numéros de variables selon la nomenclature PROOF/JGOFS

3- Choix des types de référencement

Tout en respectant les composantes d'une bonne géolocalisation, les différentes stratégies d'acquisition conduisent à trois types de référencement afin d'inclure toutes les données acquises sur le navire: chacun des trois types conduisent à l'établissement d'une trame de fichier appelé **fichier source de référencement** (Répertoire ad hoc) qui ne contient que la géolocalisation à l'exclusion de tout autre donnée en principe.

Afin d'homogénéiser la géolocalisation des données, nous avons défini un **système de référencement qui est commun à la totalité des données navires**. Pour cela, nous avons disposé de fichiers sources de référencement qui ont constitué nos trames de référence et à côté desquelles ont été juxtaposées (par concaténation) les différentes données selon les fichiers PW traités. **Les fichiers sources de référencement¹ sont utilisés en fonction du type de données, à savoir profils verticaux, horizontaux ou autres. Pour les profils verticaux, la trame de référence utilisée est la « trame rosette ». Pour les profils horizontaux, la trame de référence utilisée est « thermoall ». Le fichier de référencement « coord operalamer » est utilisé pour les autres types de données ainsi que pour le contrôle global des opérations à la mer.**

Une fois le système de référencement établi, les données sont homogènes, il est alors possible de finaliser les données et de les intégrer au sein de la base de données Pomme.

3.1- Les trois types de « fichiers sources de référencements »

Ces fichiers sources sont au nombre de trois pour chaque leg de chaque campagne Pomme.

Le premier fichier source de référencement, situé sur le CD2 sous le répertoire « *coord_operalamer*² », indique la succession temporelle des opérations à la mer avec les coordonnées. Le référencement est attaché à l'opération à la mer effectuée. Il tient donc compte de toutes les opérations à la mer effectuées mais la géolocalisation dans ce fichier n'est pas continue et n'est donnée que lorsqu'une nouvelle opération a été effectuée, quelle que soit sa nature.

Le second fichier source de référencement, répertoire « *ficros_coord*³ », constitue la trame de référence pour les fichiers rosette complets (=dernière version à date) en disposant des informations : coordonnées (latitude et longitude en degrés centésimaux), n° de station, date et heure en jour julien et profondeur. Il s'agit donc d'un type de fichier 'rosette coordonnées' qui constitue les fichiers trame de chaque leg. Ce référencement est donc attaché aux données 'bouteilles' acquises lors des profils verticaux ascendants de la CTD munie d'une rosette.

Le troisième fichier source de référencement, répertoire « *ficsurfok*⁴ », il s'agit des fichiers du thermosalinographe corrigés (surface) qui fournissent la position du navire à chaque minute. Il constitue notre fichier source en terme de référencement officiel. Ces fichiers contiennent une absence d'informations pour pomme 3 leg 1 pour la journée du 30 août 2001 de 7h55 à 19h00.

Pour ces fichiers, les résultats en température et salinité de surface sont conservés dans le fichier source de référencement.

¹ Une copie du fichier explicatif de transmission sis ...\\2_FICHIERS_SOURCES_DE_REFERENCEMENT\ *fichiers_sources_de_referencement.txt* décrivant en détail tous les fichiers sources de référencement est visualisable en Annexe 1

² copie en annexe 1 du fichier explicatif du fichier source de référencement situé dans « *coord_operalamer* » : *transcoordoperalamer.txt*

³ copie en annexe 1 du fichier explicatif du fichier source de référencement situé dans « *ficros_coord* » : *transficroscoord.txt*

⁴ copie en annexe 1 du fichier explicatif du fichier source de référencement situé dans « *ficsurfok* » : *transthermoall.txt*

3.2- Les noms des fichiers source de référencement proprement dits.

Le détail de la construction des 3 fichiers trames ou de référencement pour chaque leg de Pomme est donnée en Annexe 1. Les noms génériques sont :

result_concat_julthermo_pXIY_cor.xls dans «*coord_operalemer*»
ficrosdficcoordpXIY_modif_cor.xls dans «*ficros_coord*»,
pXlgYTHERMOall.txt dans «*ficsurfok*»,

X et Y indiquent respectivement le nom de la campagne Pomme (1, 2 ou 3) et le nom du leg (1, 2).

4- Procédures utilisées pour affecter un référencement à chaque type de données issues du site Pomme Web

4.1- Le choix du fichier de référencement (décrit précédemment) à utiliser

Ce choix a été dicté par la nature différente des travaux de nos campagnes, l'objectif étant de pouvoir référencer chaque opération et les données qu'elles induisent. Les données acquises en permanence, navire en route ou en station, demandent une référence de temps à la minute, elles sont donc croisées avec le fichier référencement ad hoc du leg concerné situé dans «*ficsurfok*» : Les données issues des prélèvement bouteille acquises en station lors d'une opération CTD avec rosette sont croisées avec le référencement ad hoc situé dans «*ficros_coord*». Et les données obtenues en station ou en route lente à partir d'autres engins (filets, profileur vidéo, grappe optique etc...) trouvent leur référencement officiel dans le fichier de référencement ad hoc du leg situé dans le répertoire «*coord_operalamer*». En fait, pour être complet il existe un quatrième groupe de données qui, comme indiqué plus haut, n'a pas subi de référencement proprement dit, bien qu'elle soient sur le site Pomme Web.

4.2- Récupération des fichiers à référencer

À dates discrètes sur le Web de Pomme, tous les fichiers ont été rapatriés 'tels que'. Une liste a été établie avec la nature des variables présentées, ainsi qu'avec le contrôle du référencement interne de chaque fichier (pour le continu horizontal, l'heure TU et pour les données bouteilles le n° de station et le n°de bouteille). Ensuite les fichiers à référencer ont été identifiés. Eventuellement une mise en forme a dû être faite pour les rendre aisément lisible en procédure automatique 'Matlab'. C'est-à-dire que le contenu des fichiers doit être en tableau avec pour chaque ligne le n° de bouteille et le n° de station, ou bien l'heure TU pour le continu horizontal.

4.3- Procédure de concaténation, valable pour les données issues de la rosette et le continu de surface. Voir Annexe 1 pour plus de détail.

Principe du programme générique utilisé: le fichier source de référencement et le fichier de données à référencer, qui comporte en fait moins de lignes, sont chargés en matrice (tableau 2 D) Matlab et nommés *rostm* pour le référencement et *ydos* pour le tableau de données. Un tableau *rosdos* résultat, de même nombre de lignes que le fichier référencement (*rostm*) est construit avec le même nombre de colonne que le fichier de données (*ydos*). Ce tableau est initialisé avec des NaN. Une boucle est scrutée sur le numéro de ligne du fichier source de référencement. Il est cherché dans cette boucle si des données correspondent à cette ligne dans le tableau *ydos*. Si oui le tableau résultat est rempli, sinon les NaN sont laissés pour la ligne correspondante. La recherche de correspondance s'effectue selon la réunion des deux repères n° de station, n° de bouteille pour les données issues de la rosette et selon la date et l'heure à la minute près dans le cas de mesures continues de surface. Dans le cas où une ligne et une seule est trouvée correspondre dans le fichier de données, le tableau résultat *rosdos* est documenté par la ligne correspondante de données. Si plus d'une ligne de données *ydos* sont trouvées (cas rare, mais répertorié), la ligne du tableau résultat est laissée en NaN, et une inscription est portée à l'écran de l'opérateur avec le n° de ligne. Ceci se produit lorsque deux dosages ont été effectués sur la même bouteille.

A la fin nous avons le tableau de données *rosdos* compatible en lignes avec le référencement *rostm*. Les deux fichiers *rostm* et *rosdos* sont d'abord sauvegardés avec les en têtes dans un fichier Matlab, *.mat*, puis simplement concaténés en colonne [*rostm*, *rosdos*] et écrits en ASCII, *.txt*, et conservés. Enfin les premières lignes du fichier concaténé *.txt* sont documentées sous Excel grâce aux en-tête des fichiers originaux sauvegardés sous Matlab. Les noms des colonnes sont inscrits dans les programmes Matlab et conformes aux noms des colonnes des fichiers initiaux de données pour la partie données et ceux des fichiers référencement pour le fichier référencement.

Si plusieurs données sont trouvées munies du même couple station - bouteille elles sont ensuite insérées manuellement, et inventoriées. C'est la seule manipulation manuelle sur les fichiers résultats qui est donc effectuée, ceci afin de minimiser les erreurs de manipulation. Dans les fichiers résultats sous Excel, le nombre de décimales pour chaque variable est contrôlé et sauvegardé.

Au total, Pour chaque fichiers données Pomme Web référencé, trois fichiers résultats de même nom explicite sont sauvegardés (et transmis), les fichiers Matlab (extension: *.mat*), Excel (extension: *.xls*), et Excel (extension: *.csv*). Cette dernière extension ayant un séparateur autre que 'espace' est plus condensée et conserve mieux le nombre de décimale de chaque donnée même lorsque l'on ouvre Excel sous un autre système. Dans ces deux derniers types, les variables sont tronquées au nombre choisi de décimales, sous Matlab et ASCII, elles restent en pleine précision. Un compte rendu est le jour même rédigé par Caroline Pocho et est placé dans le même répertoire que les fichiers résultats. Ce compte-rendu rappelle les noms et répertoires des fichiers initiaux, l'inventaire des lignes ajoutées, et les manipulations éventuellement effectuées sur le fichier Pomme Web original pour le rendre compatible avec le programme de référencement. Si les données initiales étaient présentées en tableau cumulant en ligne toute les stations et les bouteilles

sur lesquels étaient effectués les dosages, aucune manipulation préalable n'était effectuée.

Les noms génériques des fichiers référencés sont par exemple :

ficros_carbinorgdiss_p211_cor.csv pour carbone inorganique dissous (DIC) transmis au SPW par M. Gonzales Davila ; ce fichier est situé dans le répertoire *concat_carbinorgdiss*, qui contient tous les résultats de chaque pomme et chaque Leg pour ce groupe de variable, lui même trouvé dans le répertoire *3_DONNEES_TRAITEES\AUTRES_DONNEES_POMWEB\ficrosdcoord_donpomweb*

Nota Bene : les fichiers de référencement dans le répertoire « *coord_operalamer* » contiennent, de par leur construction, le référencement de chaque opération réalisée à bord. La description de cette construction est fournie dans l'Annexe 1. Les données fournies donc par toutes les opérations autres que la CTD rosette ou le continu de surface n'ont comme référencement que l'heure et la position précise et officielle du début de l'opération. Toutefois d'autres indications sont obtenues avec le fichier de référencement tels que la température de surface, la salinité.

5- Inventaire des fichiers pomme web référencés ou non (résultats) fourni en Annexe 3

Toutes les données Pomme Web n'ont pas fait l'objet de validation d'où un premier tri. Parmi celles qui en ont fait l'objet, un tri supplémentaire a du être effectué pour déterminer quel type de validation (référencement, cross validation, opération à la mer) devait être réalisée, selon la nature des données et les informations disponibles .

Le premier tri a été établi sur la base des informations ou la nature des fichiers. Les données de Pomme 0 n'ont fait l'objet d'aucune validation par la cellule. Toutefois certaines comparaisons ont été effectuées, par exemple dans les eaux profondes afin de s'assurer que les données des autres Pomme étaient cohérentes avec Pomme 0, traité par le LPO. Les données de Flotteurs, le travail temps réel (M. Assenbaum), le suivi des température (F. Gaillard) n'ont fait l'objet d'aucune comparaison avec les données traitées par la cellule. Nous avons considéré que ces données étaient référencées comme toutes les autres campagnes et surtout que nous n'avions pas l'information complète pour construire les fichiers de référencement des données bouteille.

Un cas à part ont été les données météo de Pomme 1 et 2 (G. Caniaux) que nous n'avons pas directement validées par référencement, car une travail de comparaison des dates, heures et position du navire a montré la cohérence de ces fichiers avec notre fichier source de référencement pour les continus de surface. Ceci n'est pas surprenant car le récepteur GPS qui a servi à la construction de fichiers de référencement du continu horizontal était le même que celui utilisé pour les données mâât météo et le même que pour les ADCP. Il ne nous a pas semblé utile de changer la forme des fichiers. En effet une conséquence de la validation par référencement aurait entraîné une incorporation en définitive fâcheuse de lignes sans information puisque le fichier trame de référence est séquencé à la minute alors que le fichier météo est séquencé aux dix minutes.

Le *second tri* effectué pour choisir le type de validation a été réalisé selon l'opération de saisie à la mer à l'origine :

Pour les opérations autres que continues de surface et CTD rosette, la succession officielle des opérations dans *coord_operalmer* est la seule validation. Il est à noter qu'une convention est adoptée pour le rattachement de chaque opération (filets à nappe, profileur Vidéo, grappe optique etc) à un numéro de station CTD. Le critère choisi est la proximité géographique, tout en consultant les horaires journaliers affichés à bord par le chef de mission.

Pour choisir si les données issues de la CTD rosette pouvaient être validées par cross validation (visuelle ou statistique), le critère a été le nombre de données disponibles pour un leg (plus de 500 données toutes stations confondues par leg). Par ailleurs certaines données Pomme Web retenues pour la validation par référencement (qui ne fait que géolocaliser) ont aussi été validées par validation croisée (sels nutritifs, par exemple). Signalons également que des inventaires par opération sont disponibles dans le fascicule constitué avec signets explicites sous Adobe Acrobat au format pdf : *info_synoptique_campagne_pomme.pdf* sous le répertoire « *dossier_info_synop_camp_pom* ».

6- informations générales

Conventions initiales adoptées : numérotation des stations, numéro de bouteilles

Un système d'observation central dans nos travaux est la sonde CTD (SBE 911 avec Rosette, en l'occurrence), qui délivre donc des profils de variables (T, S, O₂, Fluorescence et atténuation optique) en fonction de la profondeur d'une part et d'autre part de l'eau prélevée à l'aide des bouteilles (24 en principe) de la rosette, déclenchées à des profondeurs choisies par les opérateurs selon les objectifs scientifiques. L'eau est ensuite analysée pour différents paramètres dosés à bord ou au laboratoire. Pour l'ensemble des campagnes Pomme des navires Ifremer le nombre de stations ainsi réalisées est tout à fait conséquent (près de 600 stations, hors Pomme 0). Afin de pouvoir croiser les différents résultats en provenance de différentes équipes il a été convenu de référer chaque échantillon au numéro de station (profil CTD et au numéro de bouteille de prélèvement sur la rosette. Pour avoir un repère unique dans les campagnes Pomme il a été prévu dès le départ de noter les numéros de station CTD selon un chiffre type CSSS, où C est le n° de Pomme et SSS le numéro de station (3 positions). Cependant la numérotation est incrémentée à chaque nouveau profil CTD, que cela soit lors d'une opération CTD avec Rosette ou lors d'un Towyo. Ainsi la numérotation des seules CTD Rosette n'est pas continue, mais des listes inventaires existent dans la transmission des données (liste de Jean Raunet, fichier inventpXIY ou inventtowyo° ces listes sont aussi trouvées dans les informations synoptiques

(4_Info_cartes_campagne_pomme/operations_a_la_mer_pomme)

et dans le fascicule pdf Atlas , *info_synoptique_campagne_pomme.pdf*, ainsi que dans les fascicules graphiques CTD sous forme pdf.. Enfin le fichier de validation *coord_operalamer* contient le détail de la succession des opérations. Bien évidemment toutes ces listes sont cohérentes entre elles car elles ont toutes été fabriquées à partir des fichiers inventaires de Jean Raunet qui a contrôlé dans le détail les informations de géolocalisation des stations et profils towyo.

Il est à noter que dans le fascicule Atlas existe une liste synoptique de toutes les profondeurs échantillonnées par bouteille avec le numéro des bouteilles, ainsi que des listes complémentaires des prélèvements effectués sur ces bouteilles. Ces listes ont été établies directement à partir des feuilles de station en tenant compte des croix apposées par les

responsables de prélèvement. Nous n'avons pas effectué de croisement entre ces listes que l'on peut qualifier listes de principe, avec les données effectivement transmises, bien qu'en théorie cela aurait été réalisable. En réalité le travail accompli par les responsables de prélèvement est tout à fait prenant compte tenu de nombre de stations et le marquage des croix n'est pas en général leur priorité malgré de fréquents rappels à l'ordre. Les listes inventaires de principe ne donnent donc que des indications sur les types de prélèvement simultanés. Dans une base de fichier ou de données, il existe des outils qui donnent automatiquement ce type de liste et qui alors donne une information exacte sur les données en base.

Quelques rappels ou indications générales à propos du travail de référencement :

- le travail de référencement n'a pas agit directement sur le contenu de chaque fichier téléchargé pour ce qui concerne les valeurs des données
- le classement des données du site pomme web a été conservé pour des raisons de traçabilité
- les profils CTD sont auto référencés dans les en tête. Toutefois il existe quatre types de résultats définitifs de profils CTD qui diffèrent par leur présentation, mais sont cohérents pour la pression la température et la salinité, et les références géographiques des n° de stations.

Ces quatre types de fichiers résultats définitifs concernant les profils CTD Rosette sont les suivants :

- i-* Fichiers *intXXXX* (5 décibars), transmis au Sismar et au site Pomme web en 2002. Ils sont situés sur le CD2 (Nov 2005) sous les répertoires ; *1_fichiers_profils_ctd_O2_rosette/données_PTS_JeanRaunet/pXlgYPTSCint5*. Ces données, avec en-tête auto documentée, concernent les profils descendants uniquement avec valeurs chaque 5 mètres, y compris les contenu thermiques et halins; les informations de traitement se trouvent dans le fichier commentaireLP.DOC sur le site ftp anonymous de Louis Prieur (oceane.obs-vlfr.fr cd /pub/prieur/pomme/result_def/ficsurfok/)
- ii-* Fichiers *pomXXXX* (1décibar), transmis au Sismar et site Pomme web en 2002 et validés par le Sismar. Ils sont situés sur les CD2 (Nov 2005) sous les répertoires : *1_fichiers_profils_ctd_O2_rosette/données_PTS_JeanRaunet/ pxlyPTSCvalid*. Ces données, avec en-tête auto documentée, concernent les profils descendants uniquement avec valeurs chaque décibar, de pression température in situ salinité et conductivité. les informations de traitement se trouvent dans le fichier commentaireLP.DOC sur le site ftp anonymous de Louis Prieur : (oceane.obs-vlfr.fr cd /pub/prieur/pomme/result_def/ficsurfok/)
- iii-* Fichiers *oxyglistd_XXXX.txt* et *oxyglistupXXXX.txt* (1décibar), transmis au site pomme web sept 2003. Ils n'apparaissent plus maintenant. Cependant ils sont toujours sur les site ftp anonymous de Louis Prieur (oceane.obs-vlfr.fr cd /pub/prieur/pomme/pommeresultat_avecO2/resdefO2/....) où l'on trouve le descriptif des fichiers : *DESCRIPTIF_oxyglistd_ou_up.doc*. Sur les CD2 Nov 2005, ces mêmes fichiers sont situés :

\1_fichiers_profils_ctd_O2_rosette_PTS\fichiers_resultats_points_O2, avec un descriptif similaire *transoxygdef.txt*.

Ces fichiers avec entête auto documentée concernent les profils CTD descendants et ascendants chaque décibars avec les valeurs d'oxygène en plus de P, T in situ et S.

- iv- Fichiers *oxygdefFatt.txt* à transmettre. Ces fichiers se trouvent sur le site FTP anonyme de Louis Prieur, ainsi que les fichiers graphiques associés au sein desquels se situent les commentaires sur ces fichiers :

oceane.obs-vlfr.fr

cd/pub/prieur/pomme/resultpommeresultat_avecO2/resdefO2Fatt

Ces profils contiennent en plus des précédents les valeurs étalonnées de la fluorescence en Total Chlorophylle a et en coefficient d'atténuation de la lumière. Des fichiers explicatifs sont situés au même niveau que les répertoire de fichiers données ou graphiques. Par ailleurs les procédures d'étalonnages sont décrites dans le rapport Pomme validation- Etalonnage

Tous ces fichiers concernant les profils de la CTD rosette contiennent dans le nom du fichier le numéro officiel de la station XXXX, le premier X correspondant au n° de Pomme des navires de l'IFREMER. Une information synoptique sous forme d'inventaire est disponible sur le site pomme WEB et sur le CD2 sous la rubrique :

\4_Info_cartes_campagne_pomme\result_def\inventCTD_Pomme.zip ;

dans ces listes inventaire le numéro de station est référencé de la même manière que précédemment avec le préfixe *ascXXXX*

Enfin on trouvera sous le répertoire *result_def* au niveau immédiatement supérieur les répertoire *ficsurfok* qui contient les fichiers primaires étalonnées du Thermosalinographe avec le fichier descriptif *CommentaireLP.doc* associé. Noter que ces fichiers Thermoall.txt sont en fait les fichiers sources de référencement des continus de surface décrits par ailleurs.

7- Comment utiliser les fichiers référencés

L'avantage de disposer d'un référencement cohérent et uniforme n'est pas ici repris. Il apparaît cependant que la géolocalisation complète des données par les référencements utilisés conduit à des redondances de variables, néfastes pour la clarté du choix de l'utilisateur. Comme la validation croisée de chaque variable avec toutes les autres n'a pas été effectuée systématiquement, un fichier unique rosette n'a pas été construit. Il convient de guider l'utilisateur qui désire croiser deux variables appartenant à des fichiers validés par référencement différents. Enfin, il est rappelé que la validation par référencement n'a pas effectué de contrôle de la qualité intrinsèque des données autre que leur repérage dans la géolocalisation.

7.1- Exemple de redondance

On peut trouver dans les fichiers validés par référencement (FVR) plusieurs colonnes se référant au même paramètre. Ceci arrive lorsque le fichier original transmis sur le site Pomme web était déjà un fichier concaténé par un des responsables de dosage pour son utilisation personnelle, pour un contrôle ou bien parce qu'il en avait besoin pour un calcul. Un exemple est par exemple trouvé dans le fichier carbone inorganique dissous *ficros_carbinorgdiss_p111_cor.xls* dans lequel une colonne salinité est trouvée en col. V et une colonne NO3 en col. AV. Dans le fichier référencé issu de la cross validation pour le même leg, *ficroscontr_cfinauxp111_Atal.xls*, on trouve aussi une colonne salinité en col. L et une colonne NO3 en col. P. Ce qui a été validé par croisements statistiques de données par la cellule de validation est le second fichier. Le premier n'étant validé que par référencement. La validité des S et NO3 dans le premier fichier n'a pas été contrôlée et reste sous la responsabilité du fournisseur de fichier (dans ce cas Melchior Gonzales Davila). Très probablement les deux colonnes NO3 donneront les mêmes résultats puisque dans ce cas les données originales viennent de l'équipe de Patrick Raimbault, lesquelles ont été utilisées pour faire la crossvalidation statistique. Dans ce cas précis donc, les colonnes NO3 à utiliser est la colonne du fichier *ficroscontr_cfinauxp111_Atal.xls*, au sein duquel on dispose d'ailleurs de l'information flag. Toutefois il y a eu plusieurs versions de NO3 fournis à différentes dates. Une note à ce sujet précis est en cours de rédaction. En ce qui concerne S les valeurs officielles et contrôlées sont bien dans le fichier cross validé.

Par ces exemples la redondance d'information dans la base de FVR est donc source d'ambiguïté. Il aurait été impossible de contrôler la cohérence de chaque redondance ou à tout le moins d'en rendre compte de manière claire, puisque la cellule de validation ne dispose pas nécessairement de la traçabilité des fichiers concaténés avant la transmission au site Pomme Web. Toutefois ces exemples montrent que la cellule de validation, tout en respectant les principes de traçabilité et de non modification des données au cours de la validation par référencement, doit fournir une liste des choix conseillés de colonne et de fichier pour chaque paramètre, en se référant au plus près aux dernières fichiers fournis par le responsable de chaque variable.

Cette liste conseillée est donc fournie en annexe.

7.2- Problème de Concaténation de deux FVR.

Ce cas se présente lorsqu'un utilisateur veut par exemple croiser un paramètre situé dans un fichier et un autre paramètre, voire le même situé dans un autre. Avec les deux fichiers cités précédemment il n'y aura aucun problème, puisque le nombre de lignes de ces deux tableaux est le même (1893), c'est-à-dire le nombre de ligne du fichier trame ayant servi au

référencement des données rosette. Il suffira d'extraire les colonnes pertinentes dans chaque fichier et de les comparer. Toutefois même pour un leg donné, tous les FVR n'ont pas exactement le même nombre de lignes. En effet il arrive que le fichier de données primaires arrivés au site web pomme, comporte des 'doublets', c'est-à-dire des données de dosage différents ayant la même référence en termes de n° de station et de bouteille de prélèvement. Dans ce cas le programme de concaténation ne documente pas automatiquement cette ligne du fichier trame, mais signale à l'opérateur l'existence d'un doublet. Caroline Pocho a ensuite manuellement ajouté et documenté avec le référencement ad hoc le nombre de ligne correct pour respecter les principes de non modification des données et de traçabilité, au détriment de la conservation des 1893 lignes pour le fichier validé. Cette opération manuelle est consignée dans le rapport journalier de construction des FVR. Un tableau synoptique synthétise toutes ces opérations manuelles et consigne quelles lignes sont ajoutées. Il est donc ainsi aisé de choisir d'ignorer les lignes ajoutées, d'effectuer un moyenne ou tout autre opération avant de concaténer les deux fichiers choisis pour un croisement. Le tableau synthétique mis à jour au 15 décembre 2005 est reproduit en Annexe 5 et est présent sur le CD2 de transmission à MP Torre (*ficrosdcoord_donpomweb_controle_miseajour.doc*). On y lit par exemple que pour le leg 1 de Pomme 1, le FVR concernant l'absorption particulière contient non pas 1893 lignes mais 1894, car deux lignes sont redondantes pour la station 1007, bouteille 24.

7.3- Comment utiliser les résultats de la cross validation statistique pour une variable qui n'a pas été cross validée ?

Il suffit de concaténer, après consultation du tableau synthétique précédent, le FVR où se situe la variable avec le fichier rosette cross validé *ficroscontr_cfinauxpXIY_Atal.xls* pertinent, X pour n° de Pomme et Y pour n° du Leg. Ainsi il est aisé de choisir les lignes où le flag mécanique est nul, col. U du fichier rosette, qui indique la présence effective d'une bouteille correctement fermée, ou bien les lignes où le flag variable synthétique 2, col.AD, est nul qui indique que toutes les variables soumises à la cross validation ont été validées par les tests statistiques. La construction et la signification précise de ces flags sont indiquées dans les fichiers 'rapport journalier' de CP et sont présentés dans le rapport 'Pomme validation de données – 2. cross validation'.

7.4- Problèmes issus des mises à jour du site Pomme Web et cas particuliers de TCO2 et de sels nutritifs.

On comprend aisément que la validation des données ne peut être effectuée en continu, c'est-à-dire à chaque fois que quelques données sont corrigées dans des fichiers préalablement transmis. Les mises à jour des fichiers de données par les responsables sont inévitables, car certaines erreurs peuvent apparaître au cours de leur utilisation scientifique ou au cours d'un croisement avec d'autres données. Malheureusement la tâche de validation par référencement, ou croisement de données ne peut être rendue automatique, surtout pour éviter d'imposer des formats contraignants de transmission des données. Même avec un tel format imposé, tout ne serait pas résolu dans la mesure où toute intervention extérieure demande contrôle avant mise en base de fichiers validés. Les validations par référencement ont été effectuées aux dates signalées dans les annexes (novembre 2003, octobre 2004 et novembre 2005, voir Annexe 5 par exemple) et les FVR tiennent compte en principe de ces mises à jour. La date de téléchargement des fichiers soumis à validation est dans le rapport journalier associé à chaque fichier. Il se trouve qu'à date du présent rapport, certaines variables ont été actualisées en octobre 2005, mais non téléchargées ; c'est les cas des sels nutritifs et de TCO2, dernière version. Le travail de validation par référencement reste à faire, plus exactement à actualiser,

car des FVR pour ces variables sont disponibles. Il se trouve que ces variables récemment actualisées sont intervenues aussi dans la validation croisée qui demande plus de temps. Aussi seront-elles mises à jour après le rendu des rapports et feront l'objet d'un rapport d'actualisation.

Cependant, pour les sels nutritifs, une première étude a montré que les fichiers de cross validation reste valables dans la mesure où les tests statistiques éliminaient (peut-être trop) les valeurs douteuses. En revanche les fichiers actuellement sur le site Pomme web après actualisation et non encore validés par référencement sont suspects dans la mesure où des références aux n° de bouteille semblent erronées. Une étude donc détaillée s'impose pour valider de tels fichiers corrigés par le responsable des sels nutritifs.

Par ces quelques exemples de problèmes rencontrés on s'aperçoit qu'une validation exhaustive ne pourrait-être réalisable que de manière définitive, c'est-à-dire qu'à partir d'une certaine date après une campagne, rien ne pourrait être modifié. On introduirait alors une rigidité dans la validation qui ne servirait pas forcément l'objectif recherché : mettre à disposition de la communauté internationale des données validées. D'ailleurs plutôt que validées on devrait dire contrôlées. Nous laissons à penser quel serait l'état des données d'une campagne de biogéochimie si la validation ici réalisée n'avait pas été effectuée. En général ce travail de validation est effectué par des scientifiques qui ont un objectif précis, par exemple de réaliser une climatologie. Alors les données qui apparaissent suspectes sont éliminées. Dans notre cas ce n'était pas envisageable.

8.- liste des annexes

- Annexe 1 : Description des fichiers sources de référencement : les fichiers descriptifs joints au CD2 de Novembre 2005 sont repris sans modification, après une description en début d'annexe de la construction des trois types de fichiers source de référencement.
- Annexe 2 : Schéma du contenu des fichiers validés par référencement : Ce schéma permet de visualiser les différentes parties des fichiers ainsi validés selon la nature des données.
- Annexe 3 : Arborescence de principe de traitements de validation des données Pomme : C'est un schéma général situant le travail de référencement.
- Annexe 4 : Inventaire des fichiers validés par référencement : dans cet inventaire sont indiqués l'origine (colonne 4 description) des données site Pomme web d'après la nomenclature du site (colonne 1) et les paramètres (colonne 2) qu'ils contiennent, le nom complet du fichier validé par référencement (colonne 3) et enfin en colonne 5 le numéro de référence Proof des variables pertinentes. Toutefois, il faut prêter attention au fait que certaines variables apparaissent plusieurs fois dans différents fichiers. Pour cette raison une autre annexe est construite qui indique pour chaque variable dosée le nom du fichier et la colonne qu'il est conseillé de choisir (voir § 7 de ce rapport). Cette annexe trouve toute son utilité lorsque l'on désire mettre en base les données validées.
- Annexe 5 : Contrôle et mise à jour (15 Décembre 2005) du nombre de lignes des fichiers validés par référencement pour les données rosette et pour les continus de surface. Cependant pour plus de clarté, tous les fichiers disponibles sur le site Pomme web et classé selon la nomenclature de ce site sont ici repris. Cette annexe complète donc l'annexe précédente. Elle est destinée à fournir une information synoptique complète sur les nombres de lignes de chaque fichier. Lorsque le nombre de ligne n'est pas strictement égal à celui du fichier trame de référencement correspondant, les lignes responsables de cette différence sont identifiées par le numéro de station et le numéro de bouteille. Consulter la dernière colonne qui précise parfois que certains décalages dans les fichiers originaux de données ont été corrigés. Cette annexe trouve toute son utilité lorsque l'on désire concaténer deux fichiers validés par référencement pour croiser des données apparaissant dans deux fichiers différents.
- Annexe 6 : Choix conseillé des colonnes des fichiers validés par référencement pour chaque variable issue de la rosette. Cette annexe est destinée à guider l'utilisateur des fichiers validés par référencement dans son choix, surtout en cas de redondance dans ces fichiers.
- Annexe 7 : Architecture du CD2 de transmission au 15 Novembre 2006 (Copie chez L. Prieur). Par ailleurs une copie de ce qui est reçu par MP Torre a été effectuée par MPT en Mars 2006, car certaines actualisations jusque fin décembre 2005 ont été réalisées par ftp surtout pour réparer certains bugs de lecture sur le CD2.

Annexe 8 : Historique des transmissions de données étalonnées et/ou validées

Annexe 1

Dans cette annexe sont repris les fichiers *descriptifs* (ou de transmission) des fichiers **sources** de référencements proprement dits qui sont situés sous le répertoire du CD2 de transmission :

2_FICHIERS_SOURCES_DE_REFERENCLEMENT

NB : le texte du fichier

...\\2_FICHIERS_SOURCES_DE_REFERENCLEMENT\ *fichiers_sources_de_referencement.txt* a été complété lors de l'établissement du rapport et est donc légèrement différent ici de celui inclus dans les dossiers de transmission (in fine en décembre 2005 à Marie Paule Torre). Cependant seul le présent texte a été modifié pour compléter/préciser les informations pour plus de clarté dans ce rapport. Les trois types de fichiers explicatifs des fichiers de référencement n'ont pas eux été modifiés, même les indications des emplacements sur l'ordinateur utilisé par Caroline Pocho placées là pour des raisons de traçabilité n'ont pas été éliminées, malgré le manque d'intérêt de ces indications pour tout utilisateur extérieur à la cellule de validation.

Nous disposons de trois sources de référencement et nous décrivons maintenant comment ces fichiers source/trame ont été construits :

LE PREMIER type de FICHIER SOURCE DE REFERENCLEMENT EST décrit dans

TRANSFICROSCOORD.TXT

qui accompagne les fichiers résultats source de référencement **ficrosdficcoordpXIY_modif_cor.xls**

Il s'agit de fichiers de référencement sous Excel adapté pour les données issues des bouteilles de la rosette, il contient les coordonnées géographiques du début de la station, l'heure et le jour ainsi que les profondeurs officielles des bouteilles repérées par leur numéro de position sur la rosette. Une ligne de ce fichier, dont le rang dans le fichier est en première colonne, est repérée initialement par le numéro de station, et le numéro de bouteille. Le nombre de ligne de cette trame est donc différent pour chaque leg, mais le même pour chaque 'données traitées' du même leg après concaténation. Les exceptions sont indiquées plus loin.

Le référencement est rattaché aux données obtenues lors d'une station CTD-Rosette avec prélèvements. Il est exhaustif dans le sens que les lignes successives ont été documentées par programme à partir des fichiers bouteille (.btl) fournis par le logiciel Seabird et *retraités en décembre 2001* par L. Prieur pour obtenir les fichiers rosette avec données de la CTD aux profondeurs des bouteilles qui ont été fournis aux participants à Pomme : fichiers rospoXlegYm(b).xls, présents sur le site pomme web et sur le site ftp de L Prieur (sous pomme\pommeresultat\ficrosette, depuis le 22 11 01). Ces fichiers contiennent donc **toutes** les bouteilles déclenchées par l'opérateur CTD, qu'il y ait d'ailleurs physiquement une bouteille ou non (pour Pomme 1leg1 et Pomme 3leg1).

Les informations officielles de géolocalisation de chaque station ont suivi le cheminement suivant. La position géographique, la date et l'heure (TU) de début de la descente de la CTD Rosette ont été déterminées par Jean Raunet à partir du cahier des opérateurs CTD et des en tête de station des fichiers SBE édités à bord. Les contrôles de qualité des informations ont été effectués au laboratoire. L'heure de chaque bouteille n'a pas été retenue. Ces informations de géolocalisation sont donc les mêmes que celles indiquées dans les fichiers inventaires des stations CTD Rosette (inventpxly disponibles sur le site pomme web, mais non transmis à nouveau). Elles ont aussi été trouvées cohérentes par le Sismar. Ce sont aussi les mêmes informations qui sont dans les en-tête des fichiers résultats des profils CTD 'jean raunet' (sous \donnees_PTS_JeanRaunet et des profils CTD avec oxygène (sous \fichiers_resultats_points_O2\oxyglistdxxxx.txt) ainsi que ceux avec atténuation et Chla (non encore transmises, oxygFattlistdxxxx .txt). Toutefois, un travail de croisement de ces informations avec les autres fichiers de référencement issus du thermosalinographe fait état de légères différences en position liées aux sources différentes initiales (navigation intégrée du bord pour la CTD dispensée par affichage et acquisition du canal gps1 à 6 secondes pour le thermosalinographe). Il n'a pas été jugé utile d'harmoniser ces deux sources qui ne conduisent qu'à des différences de moins de 100 m en position.

LE SECOND type de FICHIER SOURCE DE REFERENCLEMENT EST décrit dans

TRANSCOORDOPERALAMER.TXT

qui accompagne les fichiers résultats de source de référencement : **result_concat_julthermo_pXIY_cor.xls**

Il s'agit des fichiers portant sur la succession exclusive des opérations à la mer avec les coordonnées géographiques à la date et heure et jour julien à 5 décimales du début de chaque opération. A l'intérieur de ce fichier, il n'y a donc aucune régularité ni en temps ni en espace, puisque ce fichier est séquencé par le changement d'opération à bord: début de profil CTD, de Profileur video marin, de filet Bionnes, de grappe optique, mise à l'eau de carioca, de marisonde etc....

Le référencement est rattaché à l'opération à la mer effectuée (ils tiennent donc compte de toutes les opérations à la mer effectuées mais on ne dispose de coordonnées que lorsqu'il y a une opération)

Pour construire et contrôler ce type de fichier, nous avons utilisé plusieurs sources d'information, y compris les plannings journaliers des chefs de mission. Seulement le schéma général de construction et d'amélioration est décrit maintenant. A partir des compte rendus par opération et journaliers des fichiers issus de la centrale 'Casino' des navires Ifremer, Jean raunet a établi la succession exhaustive des opérations à la mer en tenant compte du cahier de mission et des rapports individuels rédigés à bord par chaque responsable. Les fichiers .txt ainsi construits (NOMS ?) ont d'ailleurs été transmis très tôt sur le site web de Pomme. Cependant les informations recueillies n'étaient pas suffisantes pour constituer un référencement complet. Il y avait seulement l'heure TU à la minute et le jour ,sans la position., mais toutes les opérations ainsi que la dénomination des sites. Ces fichiers textes ont alors été convertis par Caroline Pocho en Excel et les dates et heures (à la minute) converties en jour julien. Ces fichiers ont été appelés en interne à la cellule de validation 'jours juliens'.

L'ensemble des informations contenues ont été vérifiées à nouveau à partir des documents originaux utilisés par Jean Raunet, afin de s'assurer de la cohérence et de la précision alors souhaitée pour constituer un fichier de référencement. Un rapide contrôle de la route du navire et de la succession des dates a également validé l'état des fichiers joursjuliens (des opérations à la mer). L'étape suivante a été réalisée en fusionnant les informations issues des fichiers thermoall.txt des thermosalinographes corrigés et validés avec celles des jours juliens. La raison est simple, la route du navire précise et les jours juliens de ces fichiers est la source primaire de nos référencements. Ces informations sont cohérentes (même GPS) avec les fichiers des flux de surface et les fichiers ADCP. Ainsi est respectée la règle du référencement unique.

Cependant dans la fusion, les jours juliens n'étaient pas nécessairement identiques à 5 décimales, et le principe de concaténation utilisé pour les fichiers rosette devaient être rendus robuste en acceptant une différence de 30 secondes pour la mise en concordance fondée sur le temps entre les fichiers thermoall et les fichiers jours juliens. Sur ces critères la position de chaque opération a été trouvée unique après quelques ajustements. Un contrôle a ensuite été réalisé sur les positions géographiques des opérations CTD, qui étaient connues par ailleurs avec précision. Ce contrôle s'est révélé satisfaisant pour un critère de 100 m.

Nous avons décidé de laisser l'information complète des fichiers thermoall.txt dans le fichier de référencement obtenu **result_concat_julthermo_pXIY_cor.xls**, pour la simple raison qu'il pouvait s'avérer utile de connaître la température et la salinité de surface au début de chaque opération. Une mise en forme finale a finalement été réalisée sous excel pour documenter explicitement les colonnes. Comme la base de notre numérotation de station est l'incrémementation de l'opération CTD (rosette ou towyo), les opérations intercalaires ont reçu une numérotation conforme à la CTD la plus proche en temps pour une position proche. Cependant il est possible que pour un nombre restreint de station ce n0 affecté ne corresponde pas au numéro de station noté par l'opérateur responsable de l'opération autre qu'une CTD. Pour cette raison une colonne dédiée a été ajoutée et les mentions ont été laissées en rouge.

Les fichiers jours juliens ont été incorporés dans le fascicule 'Atlas' des *infos-synoptiques*, mais pas les fichiers de référencement réalisés après la constitution de ce fascicule pdf.

Contrôler ce qui est écrit ci-dessus avec l'info dans le transcooordoperalamer.txt, il y a confusion semble-t-il (LP), caroline a repris une descriptin du fichier thermo/ cette redondance pourrait ajouter à la confusion...

NB casino non utilisé pourquoi ? par ce que non généralisé sur l'Atalante en 2001 sauf à mettre un opérateur devant. A Dynaproc 2, toutes les opérations et capteurs étaient inclus et le fichier était séquencé au 30 sec contre 10 minutes à Pomme 1 et 2.

LE TROISIEME type de FICHER SOURCE DE REFERENCEMENT EST décrit dans
TRANSTHERMOALL.TXT
qui accompagne les fichiers résultats de source de référencement **pXlgYTHERMOall.txt**

Il s'agit des fichiers du thermosalinographe qui fournissent la position du navire à chaque minute de la campagne sans trou.

Ce type de référencement est utilisé pour les mesures effectuées en continu durant la campagne. Après divers contrôles les fichiers les plus complets étaient ceux issus du thermosalinographe qui avait une cadence d'acquisition de une minute. Toutefois nous rappelons que deux thermosalinographes étaient utilisés en

parallèle. Les fichiers utilisés sont des fichiers établis par Jean Raunet au moment des calibrations de ces données. Ils sont le résultat pour Pomme 1 et Pomme 2 d'un croisement entre le thermosalinographe Atalante (du bord) moins stable et moins précis que le thermosalinographe SBE 21 du LPCM. La position géographique est commune à ces deux fichiers (GPS1, 6 sec). Toutefois les absences de données ne sont pas les mêmes, si bien que nous avons constitué un fichier unique sans trou après harmonisation des calibrations indiquées d'ailleurs dans le fichiers. Ces fichiers comportent donc 1440 lignes par jour complet de campagne et correspondent à chaque leg. Un flag est incorporé pour indiquer quel thermosalinographe a été utilisé. Le référencement est donc le jour julien 2001, '1^{er} janvier 0h00 = jour 1.00000', la latitude et la longitude (attention en degrés sexagésimaux format DD.MMFFFFF). Ces fichiers contiennent aussi les données de T et S à la prise d'eau de mer propre du navire (4.5 m sous la surface), et par la même constituent également les fichiers de résultats du thermosalinographe.

Un croisement a été réalisé avec les fichiers dits «paramètres météo-océaniques et flux radiatifs de surface », issus du mât météorologique (Pomme 1 et 2 seulement) et traités par Guy Caniaux, fichiers réalisés aux dix minutes avec positions en degrés centésimaux. Le même GPS étant utilisé les positions sont concordantes puisque l'heureTU était utilisée. Il n'a donc pas été jugé utile de référencer ces fichiers .

Pour l'instant seul les fichiers de alcalinité_TCO2 du continu (Melchior Gonzales) ont été référencés ainsi, ceux du HIAC (A. Sciandra) ne l'ont pas été (Pourquoi ?)

Ne pas dissocier les fichiers de données validées et référencés des comptes-rendus ou bien au cours d'une mise en site faire les liens sur les compte rendus ad hoc.

Ces trois fichiers sources possèdent des indications communes et des indications particulières toutes les indications sont fournies dans les fichiers de transmission et dans les comptes-rendus de Caroline Pocho pour chaque type de données pomme Web. Parmi les indications communes sont rappelées les nombres de stations présentes et absentes, la signification du numéro de station, la convention d'appellation des fichiers etc... Sur le plan des indications particulières on trouve les anomalies de numérotation (rares), les bouteilles en double dont un inventaire général est décrit dans le fichier : XXXXXX établi lors du contrôle général de tous les fichiers pour les nombres de lignes.....

PREMIER FICHIER SOURCE DE REFERENCEMENT:

TRANSFICROSCOORD.TXT

(pour transmission fichiers rosette coordonnees = trame de référence)

Emplacement du fichier :

D:\pomme\ctdraw\etaloxlp\refdesc\oxygdefexpl\fichiers_rosette_complets\ficos_coord
Fichier de transmission transficroscoord.txt a été réalisé en août 2004
il se trouve également dans ...ctdraw\fichiers_reunion_3sept2004\ficos_coord

Objectif : explication des fichiers résultats "ficrosdficcoordpxly_modif_cor" .xls et .txt

Le x symbolise le numero de campagne : pomme 1, 2 ou 3

Le y symbolise le numero du leg : leg 1 ou leg 2

Remarque : la campagne pomme 0 n'est pas prise en compte ici

Remarque : ces fichiers "ficrosdficcoordpxly_modif_cor" se distinguent des fichiers "ficrosdficcoordpxly_modif" par des mises en forme effectuées sur le choix du nombre de décimales (chaque colonne comporte le même nombre de décimales) et l'indication des unités de mesure

Rappel :

La campagne POMME comprend trois campagnes qui se subdivisent chacune en deux leg (exemple : Pomme 1 leg 1 et Pomme 1leg 2 =p111 et p112).

Chaque leg comprend un certain nombre de stations dont la numérotation peut être continues ou non.

Pour P1L1 : les stations vont de 1001 a 1079, sans discontinuïte

Pour P1L2 : les stations vont de 1080 a 1206, avec absence de fichiers pour certaines stations

Pour P2L1 : les stations vont de 2001 a 2081, avec absence de fichiers pour certaines stations

Pour P2L2 : les stations vont de 2082 a 2416, avec absence de fichiers pour certaines stations

Pour P3L1 : les stations vont de 3001 a 3083, avec absence de fichiers pour certaines stations

Pour P3L2 : les stations vont de 3084 a 3339, avec absence de fichiers pour certaines stations

Détail des fichiers "ficrosdficcoordpxly_modif_cor" :

Les données des variables commencent à la ligne 4, les lignes 1 à 3 décrivant le nom des variables, l'origine des fichiers et leurs unités de mesure

colonne 1 : n°ligne : numéro de ligne,

colonne 2 : n°station : numéro de station,

colonne 3 : n° position btl : numéro de position de la bouteille qui correspond au numéro de la bouteille,

colonne 4 : prof : profondeur en décibars (dbars),

colonne 5 : station : numéro de station,

colonne 6 : juldeb : jour de début en jour julien de l'année (jour 1 = 1er janvier 2001) en jour centmillième

colonne 7 : lat cent : latitude en degré centmillième (d'où une précision de 10-5),

colonne 8 : long cent : longitude en degré et en centième de degré (le signe moins indique une longitude ouest),

Détails complémentaires :

Les "NaN" signifient Not a Number selon la convention IEEE. On les emploie lors d'absence de données, le langage Matlab n'acceptant pas de vide dans les tableaux.

SECOND FICHIER SOURCE DE REFERENCEMENT:

TRANSCOORDOPERALAMER.TXT

(pour transmission fichiers coordonnées opération à la mer = fichiers des opérations à la mer avec les coordonnées)

Emplacement : D:\pomme\ctdraw\etaloxlp\refdesc\pommeres_def\operalamer\coord_operalamer
Fichier de transmission "transcoordoperalamer.txt" a ete realise en août 2004
il se trouve également dans ...ctdraw\fichiers_reunion_3sept2004\coord_operalamer

Objectif : explication des fichiers résultats "result_concat_julthermo_pxly_cor" .xls et .txt

Le x symbolise le numéro de campagne : pomme 1, 2 ou 3

Le y symbolise le numéro du leg : leg 1 ou leg 2

Remarque : la campagne pomme 0 n'est pas prise en compte ici

Remarque : ces fichiers "result_concat_julthermo_pxly_cor" se distinguent des fichiers

"result_concat_julthermo_pxly" uniquement par des mises en forme (choix du nombre de décimales et précision des unités))

Ces fichiers contiennent la totalité des opérations effectuées à bord du navire

Afin de pouvoir disposer des coordonnées (en terme de latitude et de longitude) pour chaque opération, nous avons effectué une concaténation entre les fichiers jourjulien et les fichiers thermoall (Les fichiers thermoall sont les fichiers du thermosalinographe corrigés (surface))

Ces derniers disposent de la position du navire à chaque minute et constituent notre fichier source en terme de référencement officiel

Remarque : différence au niveau des fichiers résultats entre pomme 1, 2 et pomme 3 :
pomme 3 leg 1 et 2 différent au niveau de trois colonnes sur les quatre dernières :

TQP est remplacé par TPRCOR

SAL est remplacé par SALCOR (salinité corrigée)

FLAG reste inchangé: flag

TSIS est remplacé par TCUVE (température de la cuve)

Détail des fichiers "result_concat_julthermo_pxly_cor" :

Les données des variables commencent à la ligne 4, les lignes 1 à 3 décrivant le nom des variables, leurs unités de mesure et l'origine des fichiers

colonne 1 : campagne

colonne 2 : leg

colonne 3 : année campagne

colonne 4 : jours

colonne 5 : mois

colonne 6 : heures TU

colonne 7 : minutes

colonne 8 : secondes

colonne 9 : opérations à la mer (heures TU)

colonne 10 : stations

colonne 11 : stations identiques ?

colonne 12 : sites

colonne 13 : jours julien (arrondi à 3 décimales)

colonne 14 : jour julien (à 5 décimales, donc plus précis): jour julien de l'année 2001, en jour centésimal, avec 5 décimales, incrément d'une minute. le temps de référence est le temps TU, issue de la centrale Atalante.

colonne 15 : jour (JJ)

colonne 16 : mois (MM)

Rapport Pomme validation référencement : Annexe 1 Fichiers sources de référencement

colonne 17 : annee: l'année du siècle (2001), (AA)

colonne 18 : heure: l'heure TU (hh)

colonne 19 : minute (mm)

colonne 20 : la latitude, en degrés DD. Minutes MM et décimales de minute dddddd, DD.MMdddd, + pour l'hémisphère Nord

colonne 21 : la longitude, en degrés DD , minutes (MM) et décimales de minute, -DD.MMdddd, - pour Ouest

colonne 22 : TQP (ou TPCOR) en degrés celsius : TQP pour thermomètre à quartz de proue, indique la température de surface de l'eau de mer prise et mesurée à la proue du navire. Degrés Celsius, 3 décimales. TPCOR pour température de proue corrigée.

colonne 23 : SAL (ou SALCOR /1000) : SAL pour salinité (PSU, indique /1000), SALCOR pour salinité corrigée

colonne 24 : FLAG : Flag indiquant si la salinité est issue du Thermosalinographe "SIS" (flag 1) ou du thermosalinographe LPCM, (flag2).

colonne 25 : TSIS (ou TCUVE) en degrés Celsius : TSIS pour thermomètre du thermosalinographe "SIS" (Atalante), indique la température donnée à titre indicatif. TCUVE pour température de la cuve

Remarque : redondance d'information au sein des fichiers parfois car les colonnes 1 à 13 sont issues des fichiers jour julien et les colonnes 14 à 25 sont issues des fichiers thermoall

Les fichiers thermoall pour p311 contiennent une absence d'information pour la journée du 30 août 2001 de 7h55 à 19h00, donc on a mis des NaN.

Détails complémentaires :

Les "NaN" signifient Not a Number selon la convention IEEE. On les emploie lors d'absence de données, le langage Matlab n'acceptant pas de vide dans les tableaux.

TROISIEME FICHIER SOURCE DE REFERENCEMENT:

TRANSTHERMOALL.TXT

(pour transmission fichiers THERMOall)

Emplacement : D:\pomme\ctdraw\etaloxlp\refdesc\pommeres_def\ficsurfok
Fichier de transmission "transthermoall.txt" a été réalisé en août 2004

il se trouve également dans ...ctdraw\fichiers_reunion_3sept2004\ficsurfok

Objectif : explication des fichiers "pxlgyTHERMOall"

Le x symbolise le numéro de campagne : pomme 1, 2 ou 3

Le y symbolise le numéro du leg : leg 1 ou leg 2

Rq: ces fichiers s'ouvrent avec wordpad

Ces fichiers disposent de la position du navire à chaque minute et constituent notre fichier source en terme de référencement officiel

Les fichiers thermoall sont les fichiers du thermosalinographe corrigés (surface)

Les corrections effectuées sont indiquées dans l'en tête: par exemple pour p1lg1

1 ligne indique la correction appliquée aux données brutes sur la température de proue :

$$TQP = TQP + 0.655$$

1 ligne indique la correction sur la salinité appliquée pour les données issues du Thermosalinographe SIS du navire ATALANTE :

$$SALTHESES = SALTHESES - 0.194 \quad FLAG = 1$$

1 ligne indique la correction de salinité appliquée pour les données issues du thermosalinographe SBE 26 du LPCM, Labo A :

$$SALTHLPCM = SALTHLPCM + 0.0197 \quad FLAG = 2$$

Remarque : différence au niveau des fichiers résultats entre pomme 1, 2 et pomme 3 :

pomme 3 leg 1 et 2 diffèrent au niveau de trois colonnes sur les quatre dernières :

TQP est remplacé par TPRCOR en degré Celsius (TPRCOR = TPRISEBRUT - 0.088)

SAL est remplacé par SALCOR /1000 (SALCOR = SALBRUT + 0.005)

FLAG reste inchangé: flag

TSIS est remplacé par TCUVE (température de la cuve) en degré celsius

Les deux fichiers de Pomme 3 ,1 pour chaque leg, p3lgyTHERMOall.txt sont similaires à ceux de Pomme1 et 2, mais n'ont pas été constitués de la même manière puisqu'à Pomme 3 seul le thermosalinographe SBE21 de la Thalassa a été mis en route. Le même thermosalinographe du LPCM n'a pas été utilisé. Lors de Pomme 1 et 2 sur l'Atalante le thermo du navire était un thermo SIS susceptible de plus de dérive qu'un SBE 21 et pour cette raison le thermo SBE 21 du LPCM était mis en parallèle.

Dans les fichiers thermes de Genavir concernant le Thermosalinographe la navigation n'est pas incorporée. Jean Raunet a donc joint aux fichiers Thermosalinographe issus de Thermes, les fichiers navigations de Thermes après avoir interpolé à la minute et converti le temps en jour julien. La navigation, Latitude et Longitude est également exprimée en DD.MMddddd. TPRCOR est la température de proue corrigée selon l' équation indiquée dans l'en-tête du fichier.

Les valeurs de T et S n'ont pas été filtrées (Bulles, parasites etc)et n'ont donc pas subies d'autres transformations que celles indiquées précédemment.

Détail des fichiers "pxlgyTHERMOall.txt" :

Contient :

1 ligne de titre ex : POMME2 - LEG2 - SURFACE - FICHIERS THERMOSALINO CORRIGES

1 ligne indiquant la correction appliquée aux données brutes sur la température de proue

1 ligne indiquant la correction sur la salinité appliquée pour les données issues du Thermosalinographe SIS du navire ATALANTE

1 ligne indiquant la correction de salinité appliquée pour les données issue du thermosalinographe SBE 26 du LPCM, Labo A

deux lignes de titre des colonnes avec unités

Colonne 1 : jour julien de l'année 2001, en jour centésimal, avec 5 décimales, incrément d'une minute. le temps de référence est le temps TU, issue de la centrale Atalante.

Colonne 2 le jour (JJ)

Colonne 3, le mois (MM)

Colonne 4, l'année du siècle (2001), AA

Colonne 5, l'heure TU, hh

Colonne 6, les minutes, mm

Colonne 7, la latitude, en degrés DD. Minutes MM et décimales de minute dddddd, DD.MMdddd, + pour l'hémisphère Nord

Colonne 8, la longitude, en degrés DD, minutes (MM) et décimales de minute, -DD.MMdddd, - pour Ouest

Colonne 9, TQP, température de surface de l'eau de mer prise et mesurée à la proue du navire. Les valeurs brutes sont celles issue de la centrale termes. Elles ont été corrigées du fateur indiqué dans l'entête. Degrés Celsius, 3 décimales

Colonne 10, Salinité (PSU, indique /1000).

Colonne 11, Flag indiquant si la salinité est issue du Thermosalinographe SIS (flag 1) ou du thermosalinographe LPCM, (flag2). Ce dernier a été privilégié, car plus stable, amis l'autre Thermosalinographe a été retenu pour les données manquantes.

Colonne 12, Température di Thermosalinographe donnée à titre indicatif. La comparaison avec TQP permet de regarder le décalage temporelle entre la salinité et la température de proue (de l'ordre de 5 minutes) du au temps de transit dans les quelques 60 mètres de tuyaux. Ce décalage n'a pas été appliqué.

Les données sont sans trou dans le temps. Des valeurs négatives sont appliquée pour des valeurs aberrantes ou manquantes. Pour la salinité, il arrive que le circuit d'eau soit interrompu, pour calibration du fluorimètre ou involontairement.. Les valeurs sont alors proches de zéro en S.

Les deux fichiers de Pomme 3 ,1 pour chaque leg, p3lgyTHERMOall.txt sont similaires à ceux de Pomme1 et 2, mais n'ont pas été constitués de la même manière puisqu'à Pomme 3 seul le thermosalinographe SBE21 de la Thalassa a été mis en route. Le même thermosalinographe du LPCM n'a pas été utilisé. Lors de Pomme 1 et 2 sur l'Atalante le thermo du navire était un thermo SIS susceptible de plus de dérive qu'un SBE 21 et pour cette raison le thermo SBE 21 du LPCM était mis en parallèle.

Dans les fichiers thermes de Genavir concernant le Thermosalinographe la navigation n'est pas incorporée. Jean Raunet a donc joint aux fichiers Thermosalinographe issus de Thermes, les fichiers navigations de Thermes après avoir interpolé à la minute et converti le temps en jour julien. La navigation, Latitude et Longitude est également exprimée en DD.MMdddd. TPRCOR est la température de proue corrigée selon l'équation indiquée dans l'en-tête du fichier.

Les valeurs de T et S n'ont pas été filtrées (Bulles, parasites etc)et n'ont donc pas subies d'autres transformations que celles indiquées précédemment. J'ai déjà eu avec A. Paci des échanges d'information sur les bugs dans ces fichiers. Je peux faire parvenir copie des mails

Les fichiers thermoall pour p311 contiennent une absence d'informations pour la journée du 30 août 2001 de 7h55 à 19h00, donc on a mis des NaN

Détails complémentaires :

Les "NaN" signifient Not a Number selon la convention IEEE. On les emploie lors d'absence de données, le langage Matlab n'acceptant pas de vide dans les tableaux.

ANNEXE 2

SCHEMA DETAILLANT LE CONTENU DES FICHIERS TRANSSITEPOMME SEPT2004

Les fichiers résultats validés par référencement résultent de la concaténation des trames de référencements avec les fichiers de données issues du site Pomme WEB. Ces fichiers de données ont été téléchargés du site pomme web en septembre 2004 pour la plupart. Les fichiers résultats de la validation par référencement ont fait l'objet d'une transmission en décembre 2005 à MP Torre, qui incorpore les derniers fichiers téléchargés depuis septembre 2004 (voir annexe contrôle_mise à jour).

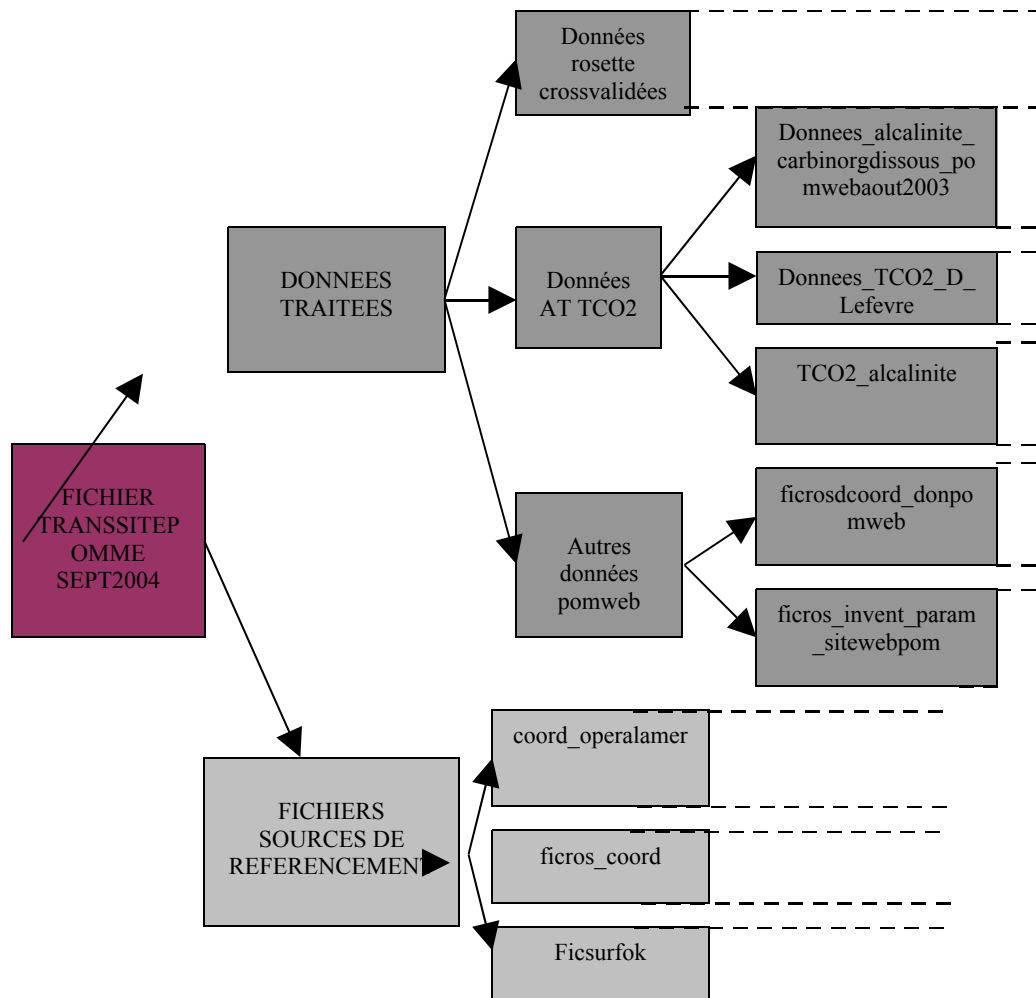
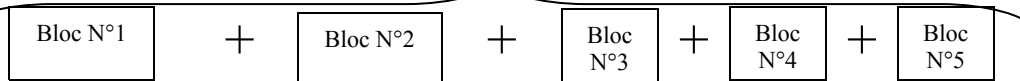
Dans cette annexe 'schéma', est expliquée la juxtaposition avec l'origine des données dans les fichiers résultats

Annexe 2 Rapport Pomme validation référencement : Schéma du contenu des fichiers validés par référencement

Dossiers contenant les fichiers de données

Blocs constitutifs des fichiers (assemblés par concaténation)

Chaque fichier est constitué de plusieurs « blocs » aboutés ou non et présentés ci-contre



Trame rosette de référencement (n° ligne, n° station, n° position btl, prof, station, jul deb, lat cent, long cent)	Fichiers rosette finaux (données sonde + O2 dosage, N03, PO4, fluo, Chla, Tchla + Pour les leg1 : flag mécanique et flagvart2 Pour les leg2 : flag mécanique avec flagvart1)	Alcalinité (AT) (données issues du site web de pomme)	TCO2 (données de Dominique Lefèvre)	Données pomme web
Trame ros réf	Fic ros finaux			
Trame ros réf	Fic ros finaux	AT		
Trame ros réf	Fic ros finaux		TCO2	
Trame ros réf	Fic ros finaux	AT	TCO2	
Trame ros réf				DonPom Web

✓ Principe de lecture du schéma à partir de l'exemple des données rosette crossvalidées :

Les fichiers comprenant les données rosette crossvalidées comportent deux groupes de colonnes, un premier groupe de colonnes ou blocs que l'on appelle « trame ros réf » (pour trame rosette de référencement), puis un second groupe appelé « fic ros finaux » (pour fichier rosette finaux)

Les fichiers de données sont classés par campagne (pomme 1, 2 ou 3) et par leg (leg 1 ou 2)

Pour pomme 1 leg 1, le fichier contient les colonnes suivantes : (le tableau est ici coupé en deux parties)

Trame ros réf								Fic ros finaux								
n° ligne	n° station	n° position btl	prof	station	jul deb	lat cent	long cent	Prof dbars	Pression dbars	Temp °C	Salinité psu	Oxygene dissous ml/l	oxygene dosage ml/l	oxygetal up ml/l	NO3 (µM) (SNPXY)	PO4 (µM) (SNPXY)
(ficrosdown)	(ficrosdown)	(ficrosdown)	(ficrosdown)	(ficcoord)	(ficcoord)	(ficcoord)	(ficcoord)									
			en dbars		en jour centmillième	en degré centmillième	en degré centmillième									

Fic ros finaux (suite)

			flag mécanique (trame ros)	flag de variables automatique (ancien)	synthese flag de variables procedure automatique (new)	P moyennes	T moy	S moy	O2 moy	NO3 moy	PO4 moy	synthese flag variable2
Chla (pigbrut)	Tchla (pigbrut)	fluoresc. reel unit (rospo1leg1mb_corrigees)										

✓ On entend par « données traitées », les données traitées par la cellule de validation

✓ Remarque : les données pomme web (téléchargées en novembre 2003) seront décrites ultérieurement. Toutefois si l'on souhaite connaître le nombre de lignes significatif, on peut se reporter à la colonne « nbre de lignes du fichier lu donpomweb » (=6^{ème} colonne) du fichier « ficrosdcoord_donpomweb_controle.doc » (emplacement : fichier_transsitepomme_sept2004\DONNEES_TRAITEES\AUTRES_DONNEES_POMWEB\ficrosdcoord_donpomweb)

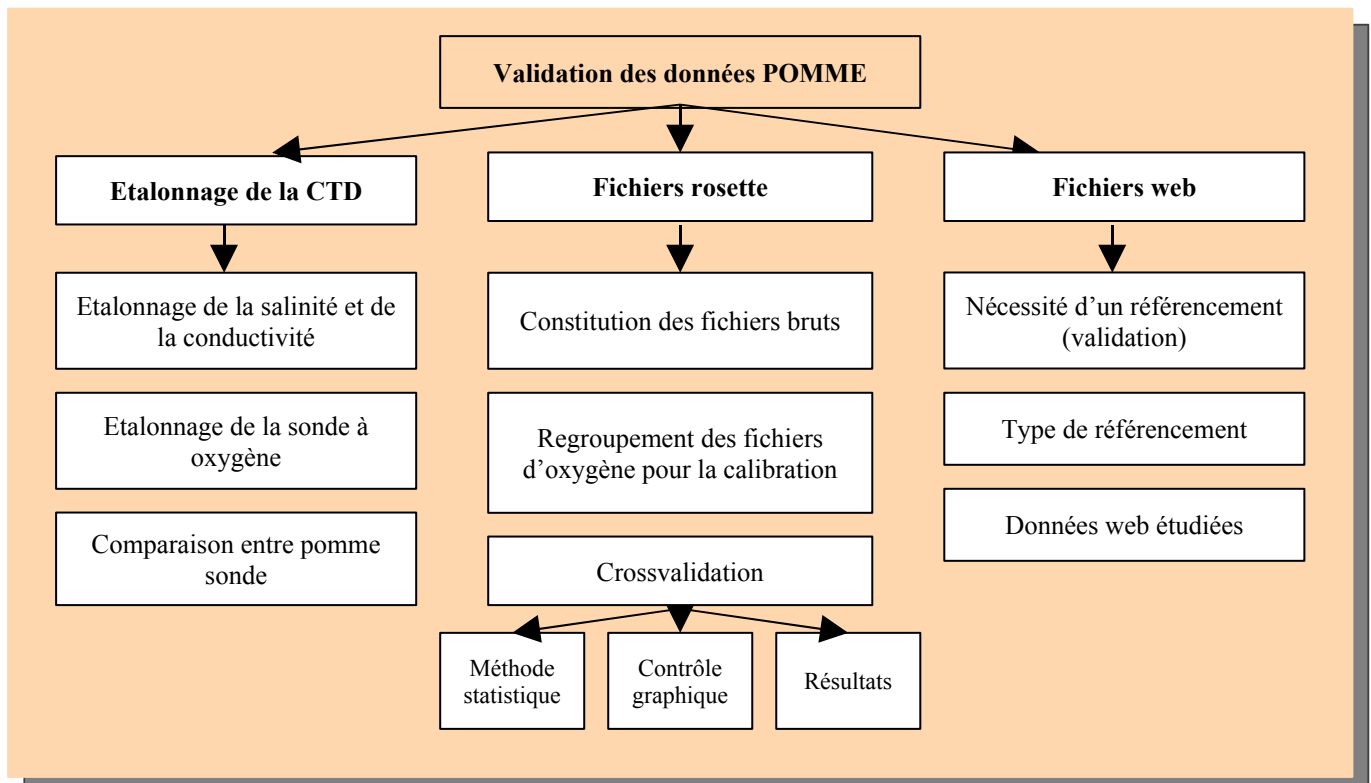
✓ A l'étape actuelle, le bloc n°3 : « AT » et le bloc n°4 : « TCO2 » ne sont pas pris en compte pour la définition du flagvar1 et flagvar2

ANNEXE 3

ARBORESCENCE POMME TRAITEMENT

29/10/03
C. Pocho

Nous avons choisi de synthétiser l'information afin de la rendre plus accessible. Il est évident qu'une seule étape présentée au sein de l'arborescence regroupe en réalité un grand nombre de traitements. La majorité de ces traitements ont été réalisés à l'aide du logiciel Matlab.



Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

ANNEXE 4

INVENTAIRE_FICHIERS_REFERENCEMENT

(emplacement : D:\pomme\ctdraw\Donnees_finales_LP_CP)

15/11/2005

C. Pocho

- ✓ Obj : réaliser un inventaire, une liste des noms des fichiers de données et de transmission par référencement
- ✓ Pour plus de détails, consultez le fichier « [reperage_fichiers_referencement.doc](#) »

TYPES DE FICHIERS / DE PARAMETRES	PARAMETRES	NOM DU FICHIER	DESCRIPTION	REFERENCE PROOF
Premier fichier source de référencement	N° ligne N° station N° position btl Prof Station Jul deb Lat cent Long cent	ficrosdficcoordp111_modif_cor.xls	fichier rosette coordonnées qui constitue notre fichier trame Le référencement est rattaché aux données CTD	
Second fichier source de référencement	Operations à la mer Jours julien Jours Mois Année Heure minute Latitude Longitude TQP SAL FLAG TSIS	result_concat_julthermo_p111_cor.xls	fichiers des opérations à la mer avec les coordonnées Le référencement est rattaché à l'opération à la mer effectuée (ils tiennent donc compte de toutes les opérations à la mer effectuées mais on ne dispose de coordonnées que lorsqu'il y a une opération)	
Troisième fichier source de référencement	Julien Jour Mois Année Heure Minute Latitude Longitude TQP SAL FLAG TSIS	p111THERMOall.txt	fichiers du thermosalinographe qui fournissent la position du navire à chaque minute	
Absorption particulaire	Absorption phytoplanctonique ou detritique de 370 à 750 nm (tous les 2 nm) en m-1	ficros_abspart_adet_pxly_cor.xls ficros_abspart_aphyto_pxly_cor.xls	Données issues du site web Pomme	98
Assimilation	absorption de l'ammonium (en nmol N/litre/jour)	ficros_assimiazote_p112_cor.xls	Données issues du site	418

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

d'azote (N15)	absorption de l'urée (en nmol N/litre/jour) absorption du nitrate (en nmol N/litre/jour)		web Pomme	
Assimilation d'azote (N15) et de carbone (C13)	<u>D'après la notice :</u> Colonne C : Concentration en nitrite (µM) Colonne D : Concentration en nitrate (µM) Colonne E : Concentration en phosphate (µM) Colonne F : Concentration en ammonium (nM) +- Explication du formatage du fichier excel pour mesures de production Colonne D : Taux de production primaire (mgC.m-3.d-1) Colonne E : Assimilation de nitrate (nmoles.l-1. d-1) Colonne F : assimilation d'ammonium (nmoles.l-1. d-1) <u>D'après nos fichiers :</u> production primaire (mg/m3/j) production nouvelle (nM/T) production régénérée (nM/T)	ficros_assimiazotecarb_p111_cor.xls	Données issues du site web Pomme	Assimilation 15N (NH4) : 226 Régénération : 172 production primaire + production nouvelle + production régénérée : 85 et 86
Bactéries hétérotrophes	OVNI cell/ml HNA cell/ml LNA cell/ml total cell/ml	ficros_bactheterot_p211_cor.xls	Données issues du site web Pomme	? non traité Si sur rosette : 78
Carbone inorganique dissous	CT pH AT fCO2 NO2 (µM) NO3 (µM) PO4 (µM) O2 µM	ficros_carbinorgdiss_p311_cor.xls Oct 2004 : ficros_carbinorgdiss_jourjul_p111.xls (continu de surface)	Données issues du site web Pomme	CT : ? pH : 36 AT : 35 fCO2 : 361 NO2 : 29 NO3 : 28 PO4 : 31 O2 : si c'est du dosage : 5
Carbone Organique total (TOC)	TOC		Données issues du site web Pomme	256
Chimie DTX	<u>Fichier annexe 1 :</u> pigments en spectrofluorescence (µg/L) chloro a chloro b chloro c phéo a phéo b <u>Fichier annexe 2:</u> Prélèvements de surface: salinité PSS 78		Données issues du site web Pomme	pigments en spectrofluorescence : 41 salinité PSS 78 : 357, mais si thermosalino : 358
Composés soufrés DMS et DMSP	DMS DMSPp DMSPp<10um DMSPp>10um DMSPd	ficros_compsoufres_p111_cor.xls	Données issues du site web Pomme	67 sur rosette

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

<p>Compteur de particules HIAC</p>	<p><u>Fichiers niveau 1:</u> Ce niveau de traitement donne les spectres bruts pour les échantillons prélevés sur la rosette ou sur le continu de surface. Col 1 à ... : données</p> <p><u>Fichiers niveau 2:</u> Ce niveau de traitement donne les concentrations, tailles moyennes et biovolume des particules pour chaque échantillon. Nbre/L diamètre (µm) surface (µm²/L) biovol (µm³/L)</p>	<p>ficros_hiac_ros_N1_p3l2_cor.xls ficros_hiac_ros_N2_p1l1_cor.xls</p>	<p>Données issues du site web Pomme</p>	<p>N1 et N2 : 19 (continu de surface) et 20 (rosette)</p>
<p>Compteur optique de plancton (OPC)</p>	<p>T1000 (biovolume en mm³/m³ des tailles de 250 à 1000 µm) T2000 (biovolume en mm³/m³ des tailles de 1000 à 2000 µm) T4000 T8000 T13000 Total</p>		<p>Données issues du site web Pomme</p>	<p>Si lab : 17 ? Si fish : 18 ?</p>
<p>CTD-rosette et surface (Atalante)</p>	<p><u>CTD toutes variables "continues", tous les mètres :</u> D/M (ex : 1) nbre séquences Fichiers pomxxxx P, T, S, C tous les mètres (1 fichier par station) Fichiers intxxxx P, T, S, C tous les 5 mètres (1 fichier par station) <u>Rosette :</u> n° station N° des btl choix de départ des profondeurs btl fichier rosette des prof btl</p> <p><u>Fichiers à télécharger :</u> TO90 sal sigma-é00 ox ml/l F1C Xmiss Upoly0 Upoly1 Nbf</p> <p><u>Thermosalino:</u> TQP SAL Flag TSIS</p> <p><u>Meteo:</u> VIT CAP TAIR HUM PA</p>		<p>Données issues du site web Pomme</p>	<p>CTD : 2 O2 sonde : 4</p>

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

	RAD <u>Navigation:</u> loch loch gyro1 gyro2 VIT CAP HDM			
CTD-Rosette (Thalassa)	Idem « CTD rosette et surface (Atalante) »		Données issues du site web Pomme	CTD : 2 O2 sonde : 4
CTD-O2 et CTD-Rosette (Atalante) (Août 2003)	<u>Fichiers profils CTD O2 rosette:</u> pression (dbar) température (°C) salinité (psu) oxygène dissous (ml/l) oxygène saturé (ml/l) oxygène dissous (µM/kg) <u>Fichiers CTD rosette (avec ou sans synthèse) :</u> Temp up sal up oxygetal up oxdos Chla Tchla NO3 PO4 fluo flag oxyg flag mecanique (manuel) flag variables (manuel) flag mecanique (autom) flag variables visuel (autom) flag variables autom (autom) Détail du flag variables autom prof flag de T flag de S flag de O2 flag de NO3 flag de PO4 Profils moyens: P moyens T moy S moy O2 moy NO3 moy PO4 moy		Données issues du site web Pomme	CTD : 2 O2 sonde : 4
CTD (D'Entrecasteaux), XCTD et XBT	- Profils de salinité (figures .ps ou .eps) - Profils de température (figures .ps ou .eps) - Profils sal temp (figures .ps ou .eps) - Profils XCTD : depth temp		Données issues du site web Pomme	CTD : 2 O2 sonde : 4

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

	sal - Profils XBT : depth temp			
Fer	DFe TdFe Remarques		Données issues du site web Pomme	97
Flux O2/TCO2 et Activité ETS	<u>Activité ETS :</u> activité ETS <u>Flux O2-TCO2 :</u> PCB O2 se PCB RCO O2 se RCO PCN O2 se PCN PCB TCO2 se pcb RCO TCO2 se rco PCN TCO2 se pcn	ficros_ETS_p3l2_cor.xls ficros_O2TCO2_p1l1_cor.xls	Données issues du site web Pomme	ETS : 152 Flux O2/TCO2 : 90
Microzooplancton	<u>Abondance (cells/l) :</u> effectifs oligotriches effectifs tintinnides <u>Biomasse (µg/l) :</u> biomasses oligotriches biomasses tintinnides % de mixotrophes	ficros_microzoopl_p1l1_cor.xls	Données issues du site web Pomme	94
Paramètres météo-océaniques et flux radiatifs	température de l'air en degré Celcius température de la mer en degré Celcius. humidité de l'air en pourcent. pression de l'air en hectoPascal. composante zonale du vent en m s ⁻¹ . composante méridienne du vent en m s ⁻¹ . intensité des précipitations en mm/h. flux de chaleur sensible en W m ⁻² . flux de chaleur latente en W m ⁻² flux solaire descendant en W m ⁻² flux infrarouge descendant en W m ⁻² flux infrarouge net en W m ⁻² bilan net de chaleur en W m ⁻² bilan radiatif net mesuré en W m ⁻² bilan radiatif net calculé en W m ⁻² stress du vent (rair u*2) en N m ⁻² composante zonale du stress du vent (taux) en N m ⁻² . composante méridienne du stress du vent (tauy) en N m ⁻² .		Données issues du site web Pomme	432
Pièges à particules	<u>Piège dérivant :</u> flux en mg/m2/j <u>Piège dérivant carbone total et azote :</u>		Données issues du site web Pomme	<u>Piège dérivant :</u> flux : 336 CT : 194

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

	<p>carbone total (mg/godet) flux carbone total (CT) (mg/m2/j) azote (N) (mg/godet) flux azote (mg/m2/j)</p> <p><u>Pièges fixes :</u> position du piège P. matière (mg.m-2) flux (mg/m2/j)</p>			<p>flux CT : 194 N : 330 flux N : 330 <u>Pièges fixes :</u> position du piège P. matière : 337 flux : 337</p>
Pigments phytoplanktoniques	<p><u>PIGMENTS:</u> Chlorophyll c3 Chlorophyll c1+c2 Peridinin 19'-Butanoyloxyfucoxanthin Fucoxanthin 19'-Hexanoyloxyfucoxanthin Prasinoxanthin Violaxanthin + Neoxanthin: (coéluent) Neoxanthin Diadinoxanthin Alloxanthin Diatoxanthin Zeaxanthin Lutein Lutein 5,8-epoxide Chlorophyll c-like non polaire 1 TChlb: (Chlorophyll b + DV Chlorophyll b) Divinyl Chlorophyll a Chl a: (allomer+ Chlorophyll a + epimer) Total Chl a: (Chlorophyll a + DV Chlorophyll a) Chlorophyll c-like non polaire 2 Carotenes : (somme des carotènes) Chlorophyll c-like1 @ 2.6 min Chlorophyllid a Flags</p>	ficros_pigmentsphyto_p111_cor.xls	Données issues du site web Pomme	42
Production bactérienne et rendements de croissance	<p><u>Production bactérienne :</u> vraie prof BP (ngC/l/h)</p> <p><u>Rendements de croissance :</u> code durée totale manipe biodegradation en jours durée observation consommation O2 (jours) concentration maxi O2 (µM O2) concentration mini O2 (µM O2) régression linéaire pente conso O2 (µM O2/l/j) moment maximum BP (jours) temps zéro BP (ngC/l/h) maximum BP (ngC/l/h) date réelle prise pour calcul BGE (jours) BP cumule sur cette période rapporte à un flux journalier (µmolC/l/j) rendement de croissance BGE %</p>	<p>ficros_prodbact_p311_cor.xls ficros_rendcroiss_p112_cor.xls</p>	Données issues du site web Pomme	<p>Production bactérienne : 151 Rendements de croissance : 93</p>
Production primaire	<p>screen or in situ incubation depth of incubation (m)</p>	ficros_prodprim_p112_cor.xls	Données issues du site web Pomme	84

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

(méthode 14C)	net carbon absorption dark bottle (nM) net carbon absorption light bottle >2µm (nM) net carbon absorption light bottle SD (>2µm) (nM) net carbon absorption light bottle GF/F (nM) net carbon absorption light bottle SD (GF/F) (nM) net carbon absorption light bottle Sum (nM) net carbon absorption light bottle SD (sum) (nM) primary production PP (nMd-1) integrated primary production IPP (nmol m-2 d-1) integrated primary production IPP>2µ (nmol m-2 d-1)			
Profils bio-optique	<p><u>Fichiers absorption-atténuation-LSS (=ac9-LSS) :</u> sens du profil (ex : P2_026_rem = station pomme 2 remontée) LSS en unités relatives absorption totale (at) en m-1 atténuation totale (ct) en m-1 atténuation dissout (cd) en m-1 flag relatif aux mesures CTD Flag relatif aux mesures at Flag relatif aux mesures ct Flag relatif aux mesures cd</p> <p><u>Fichiers rétrodiffusion : ECO-VSF analogique à 532nm (ECO_analog) :</u> sens du profil (ex : P2_026_rem = station pomme 2 remontée) bb 532 en unités relatives</p> <p><u>Fichiers rétrodiffusion : ECO-VSF digital à 650 nm (ECO_650nm) :</u> sens du profil (ex : P2_026_rem = station pomme 2 remontée) bb650 en m-1</p>		Données issues du site web Pomme	Pour AC9 LSS : 183 pour l'atténuation et 23 pour l'absorption Ne pas tenir compte de ECO VSF analogique et digital
Radio-sondages	l'altitude du lancer (m). le système de mesure du vent (GPS). Date et heure du lâcher (UTC) date du heure du réseau synoptique correspondant (UTC) <u>Données :</u> 1 enregistrement de 120 caractères par niveau avec 15 paramètres (15 x 8 caractères , cadrage des valeurs à droite) ; Les valeurs manquantes sont notées -999 1 niveau toutes les 5 secondes. minutes secondes altitude pression température humidité relative température du point de rosée temperature du thermomètre mouillé rapport de mélange tension de vapeur temperature potentielle Teat'w direction du vent		Données issues du site web Pomme	433

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

	vitesse ascensionnelle			
SeaSoar-OPCT	radiales positions début WGS84 positions fin WGS84		Données issues du site web Pomme	
Sels nutritifs (Atalante)	Concentration en nitrite (µM) : NO2 Concentration en nitrate (µM) : NO3 Concentration en phosphate (µM) : PO4 Concentration en ammonium (µM) : NH4	ficos_selsnut_Atala_p111_cor.xls	Données issues du site web Pomme	NO2 : 29 NO3 : 28 PO4 : 31 NH4 : 30
Sels nutritifs (D'Entrecasteaux)	Concentration en nitrite (µM) Concentration en nitrate (µM) Concentration en phosphate (µM) Concentration en silicates (µM) : Si(OH)4		Données issues du site web Pomme	NO2 : 29 NO3 : 28 PO4 : 31 Si(OH)4 : 33
Sels nutritifs (Thalassa)	Concentration en nitrite (µM) Concentration en nitrate (µM) Concentration en phosphate (µM) Concentration en ammonium (µM) Si (µM)	ficos_selsnut_Thalassa_p311_cor.xls	Données issues du site web Pomme	NO2 : 29 NO3 : 28 PO4 : 31 NH4 : 30 Si : 33 ? (à vérifier)
Silicium	type Si(OH)4 BSi (0.6-10µm) BSi (>10µm) BSi (total) rSi (0.6-10µm) rSi (>10µm) rSi (total) VSi LSi (0.6-10µm) LSi (>10µm) LSi (total)	ficos_silicium_p112_cor.xls	Données issues du site web Pomme	Si(OH)4 et BSi : 33 Pour le reste : ?
VM-ADCP	<u>les caractéristiques de l'ADCP et sa configuration, notamment :</u> angle des faisceaux par rapport à la verticale (Beam_Angle) angle de l'ADCP par rapport à l'axe du navire (ADCP_Angle) la profondeur de l'ADCP (Xoff) longueur des cellules (Bin_Length) la profondeur de la 1ère cellule par rapport à l'ADCP (Middle_Bin1_Depth) le nombre d'ensembles moyennés utilisé pour créer chacun des profils (Nb_ens_moy) un flag indiquant si la correction du (roulis/tangage) a été effectué ou non (Corr_PR). Ici, le (roulis/tangage) est pris en compte. les valeurs des variables utiles au calcul de la vitesse absolue l'amplitude (Amplitude) le désalignement (HeadMis) l'assiette (PitchMis) les caractéristiques du nettoyage des profils par filtre		Données issues du site web Pomme	?

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

	<p>médian (flag 2) nombre de profils à considérer de part et d'autre du profil à valider (Nb_a_Moyenner) nombre d'écart-types (Nb_Ecart_Moyenne) <u>pour chaque profil :</u> la vitesse navire (Unav,Vnav) la profondeur du milieu des cellules (Depth) la température (PTx) le cap (PHdg) le roulis (PRoll) le tangage (PPtch) le flag pour chacune des cellules du profil (flag_courant) les 3 vitesses absolues du courant (EWVelAbs, NSVelAbs, UDVelAbs) les écart-types associées (URMS, VRMS, WRMS) le pourcentage de données correctes qui a permis de déterminer ce profil lors du moyennage (Pgood) les intensités d'écho (PECI) les données de bottom-ping (BT...) Pour le D'Entrecasteaux, on ne corrige pas de l'effet du (roulis/tangage) car ces données ne sont pas disponibles (Corr_PR=0).</p>			
Zooplancton	<p><u>BIONESS :</u> n° du filet de pêche (n° 1 à 9) tranches d'eau échantillonnées biomasse <u>WPII :</u> biomasse abondance copépodes méso et macroplancton</p>		Données issues du site web Pomme	187
Paramètre fluo variable		ficrosparamfluovar_p311.xls	Données issues du site web Pomme Nouveau paramètre (oct 2004)	431 (à compléter)
Paramètre photosynthétique		ficrosparamphotosynt_p111.xls	Données issues du site web Pomme Nouveau paramètre (oct 2004)	91

FICROSDCOORD_DONPOMWEB_CONTROLE_MISEAJOUR

(emplacement : D:\pomme\ctdraw\etaloxlp\refdesc\oxygdefexpl\fichiers_roseette_complets\ficrosdcord_donpomweb)

8/12/2005 - C. Pocho

EXPLICATION DU REFERENCEMENT DES FICHIERS ISSUS DU SITE WEB DE POMME (<http://www.lodyc.jussieu.fr/POMME/>)

✓ Règles et principes :

- **Référencement des données issues du site web de pomme.** Pour cela on les concatène à un fichier de référencement qui varie en fonction du type de profil (vertical ou horizontal). On utilise le **fichier de référencement rosette pour les profils verticaux** et celui du **thermoall (2^{ème} fichier de référencement) pour les profils horizontaux (continus de surface)**
- **La concatenation s'effectue en fonction du numéro de station et du numéro de bouteille excepté pour l'assimilation d'azote** où la concaténation s'est effectuée en fonction du numéro de station et de la profondeur (on ne disposait pas du numéro de bouteille) **ainsi que pour la carbone inorganique dissous** où la concaténation en fonction du jour julien à partir du fichier thermoall comme fichier de référencement (avec une précision à la minute, une minute = 0.0007)
- **A une ligne correspond une même bouteille et une même station. Si ce n'est pas le cas, nous sommes alors en présence d'un doublet (ou occurrence), on précise alors le n° de station et de btl)**

✓ Mode de lecture de ce tableau :

- La colonne « Nbre de lignes du fichier trame ros » indique le nombre de lignes que comporte le fichier trame qui constitue notre trame de référencement (fichier trame rosette ou fichier thermoall)
- La colonne « Nbre de lignes du fichier lu donpomweb » indique le nombre de lignes que comporte le fichier de données issu du site web de Pomme
- La colonne « Nbre de lignes du fichier concaténé (trame+donpomweb) » indique le nombre de lignes que comporte le fichier concaténé. Si ce nombre est supérieur à celui de la trame c'est que nous sommes en présence de doublets (case en orange). Si vous êtes en présence d'un doublet il faut regarder les deux colonnes suivantes
- Les colonnes « Si plus d'une ligne avec meme btl et meme station : Indication des n° de stations et des n° de btl » vous indiquent le numéro de station et de bouteille du doublet. **Attention en cas de présence de doublet la règle « à une ligne correspond un seul numéro de bouteille et de station » ne peut plus s'appliquer**

✓ Actualisation effectuée en décembre 2005 : en plus des données téléchargées de novembre 2003, on insère également celles téléchargées en oct 2004 et du carbone organique total (TOC) (téléchargées en novembre 2005)

- **Données d'octobre 2004 :** Depuis le mois de novembre 2003, les mises à jour effectuées concernent les paramètres suivants : Carbone inorganique dissous, Carbone organique total, Microzooplancton, Paramètre fluo variable (nouveau paramètre), Paramètres photosynthétiques (nouveau paramètre), SeaSoar OPCT, VM-ADCP, Pièges à particules (Pièges dérivants, Pièges fixes). Les concaténations avec le fichier de référencement ont été effectuées pour les paramètres suivants : Carbone inorganique dissous, Paramètre fluo variable (nouveau paramètre), Paramètres photosynthétiques (nouveau paramètre)
- **Données de novembre 2005 :** traitement et concaténation du carbone organique total (TOC)

✓ Classement par type de profils :

○ Profils verticaux :

- L'absorption particulaire
- L'assimilation d'azote (concatenation station profondeur)
- L'assimilation d'azote et de carbone
- Les bactéries hétérotrophes
- Les composés soufrés DMS et DMSP
- Flux O₂/TCO₂ et activité ETS
- Microzooplancton
- Pigments phytoplanctoniques
- Production bactérienne et rendements de croissance
- Production primaire (méthode 14C)
- Sels nutritifs (Atalante)
- Sels nutritifs (Thalassa)
- Silicium
- Compteur de particules HIAC (rosette uniquement, continu de surface n'est pas traité)

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

- Paramètre fluo variable (en octobre 2004)
- Paramètre photosynthétique (en octobre 2004)
- Le carbone organique total (TOC) (en novembre 2005)

○ **Profils horizontaux :**

- Carbone inorganique dissous (Concaténation avec le fichier thermoall en fonction du jour julien avec une précision à la minute) (en novembre 2003 et en octobre 2004)

DONNEES REFERENCEES

Données téléchargées en novembre 2003 :

TYPE DE DONNEES	DONNEES WEB	INFORMATIONS	LEG	Nbre de lignes du fichier trame ros	Nbre de lignes du fichier lu donpomweb	Nbre de lignes du fichier concaténé (trame+donpo mweb)	Si plus d'une ligne avec meme btl et meme station : Indication des n° de stations et des n° de btl	Remarques et conséquences
--------------------	-------------	--------------	-----	--	---	--	---	---------------------------------

Annexe 5 Rapport Pomme validation référencement : Contrôle et mise à jour des nombres de lignes des fichiers validés par référencement

Données issues de la rosette	Absorption particulaire (adet+aphyto)	Adet	P1L1	1893	745	1893+1=1894	Station 1007	Btl 24 présente deux fois	
			P1L2	1281	201	1281			
			P2L1	1702	837	1702			
			P2L2	1233	201	1233+11=1244	2276	Deux fois btl 7, 8, 9, 10, 11, 12, 17, 18, 19, 21, 23	
			P3L1	1990	685	1990			
			P3L2	1334	201	1334			
		Aphyto	P1L1	1893	745	1893+1=1894	1007	Btl 24 présente deux fois	
			P1L2	1281	201	1281			
			P2L1	1702	837	1702			
			P2L2	1233	201	1233+11=1244	2276	Deux fois btl 7, 8, 9, 10, 11, 12, 17, 18, 19, 21, 23	
			P3L1	1990	685	1990			
			P3L2	1334	201	1334			
	Assimilation d'azote (N15)		P1L2	1281	38	1281			Concaténation en fonction N°station et prof
			P2L2	1233	31	1233			
			P3L2	1334	32	1334			
	Assimilation d'azote (N15) et de carbone (C13)		P1L1	1893	162	1893			
			P2L1	1702	132	1702			
			P3L1	1990	108	1990			
	Bactéries hétérotrophes		P2L1	1702	1001	1702			
	Carbone inorganique dissous		P1L1	1893	1096	1893			
			P2L1	1702	667	1702			
			P3L1	1990	1159	1990			
	Carbone Organique total (TOC)	Données éparses							Pb de décalage de lignes qui a été corrigé Pb de décalage de lignes qui a été corrigé
	Composés soufrés DMS et DMSP		P1L1	1893	234	1893			
			P2L1	1702	399	1702			
			P3L1	1990	373	1990			
	CTD-rosette et	CTD (toutes variables)							

Données téléchargées en octobre 2004 : (mise à jour)

TYPE DE DONNEES	DONNEES WEB	INFORMATIONS	LEG	Nbre de lignes du fichier trame ros ou du fichier trame thermoall	Nbre de lignes du fichier lu donpomweb	Nbre de lignes du fichier concaténé (trame+donpomweb)	Si plus d'une ligne avec meme btl et meme station :	Remarques et conséquences
							Indication des n° de stations et des n° de btl	
	Carbone inorganique dissous		P1L1	30 827	2651	30 827		Concatenation avec le fichier thermoall en fonction du jour julien avec une précision à la minute
			P2L1	29 977	1667	29 977		
			P3L1	30 929	13 459	30 929		
	Paramètre fluo variable		P1L1	1893	768	1893		
			P2L1	1702	888	1702		
			P3L1	1990	895	1990		
	Paramètre photosynthétique		P1L1	1893	118	1893+4=1897	1070 1075	Deux fois btl 23 et 24 Deux fois btl 23 et 24
			P2L1	1702	140	1702		
			P3L1	1990	98	1990		

Données téléchargées en novembre 2005 : (mise à jour)

TYPE DE DONNEES	DONNEES WEB	INFORMATIONS	LEG	Nbre de lignes du fichier trame ros	Nbre de lignes du fichier lu donpomweb	Nbre de lignes du fichier concaténé (trame+donpomweb)	Si plus d'une ligne avec meme btl et meme station :	Remarques et conséquences
							Indication des n° de stations et des n° de btl	
Données issues de la rosette	Carbone Organique total (TOC)		P1L1	1893	1224	1893		
			P2L1	1702	1248	1702		
			P3L1	1990	1080	1993		

DONNEES NON REFERENCEES

Données téléchargées en novembre 2003 :

TYPE DE DONNEES	DONNEES WEB	INFORMATIONS	LEG	Nbre de lignes du fichier trame ros	Nbre de lignes du fichier lu donpomweb	Nbre de lignes du fichier concaténé (trame+donpomweb)	Si plus d'une ligne avec meme btl et meme station : Indication des n° de stations et des n° de btl	Remarques
Autres types de données	Chimie DTX	Fichier annexe2 : Prélèvements de surface						
	Compteur de particules HIAC (continu de surface uniquement)	Fichier niveau1 : Echantillons prélevés sur la rosette ou sur le continu de surface Fichiers niveau 2 : Concentrations, tailles moy et biovolume des particules						
	CTD-rosette et surface (Atalante)	Thermosalino Meteo Navigation						
	Paramètres météo-océaniques et flux radiatifs							
	Radio-sondages	(sondages)						
	SeaSoar-OPCT	(radiales)						
	VM-ADCP							
	Chimie DTX	Fichier annexe1 : profils verticaux						
	Compteur optique de plancton (OPC) Fer	Radiales						
	Pièges à particules	Pièges fixes Pièges dérivants (Flux+Carb total et azote)						
Profils bio-optique Zooplancton	Bioness WPPII							

✓ **Un tableau de synthèse des doublets est présenté ci-dessous :**

- Rq : P0 n'a pas été traité et P3L2 ne contient aucun doublet
- Rq : Les données concernées par la présence de doublets sont :
 - Abs particulière (adet+aphyto)
 - Flux O2/TCO2 activ ETS
 - Pigments phyto
 - HIAC pour rosette N1 et N2 uniquement (pas les continus de surface)
 - Paramètre photosynthétique

Leg	N° STATIONS où doublets	N° BTL où doublets
P0	Non traité	Non traité
P1L1	1005	Deux fois btl 1
	1007	Deux fois btl 24
	1064	Deux fois btl 5, 6, 9, 10
	1069	Une btl en surface
	1070	Deux fois btl 23 et 24
	1075	Deux fois btl 23 et 24
P1L2	1087	Deux fois btl 4
	1092	Deux fois btl 12, 13
P2L1	2035	Deux fois btl 10
	2051	Deux fois btl 7
P2L2	2276	Deux fois btl 7, 8, 9, 10, 11, 12, 17, 18, 19, 21, 23
	2408	Deux fois btl 1
P3L1	3019	Deux fois btl 24
P3L2	/	/

LISEZ_MOLDOC

(emplacement : F:\SAUV_supprimees_du_disqueD\dossier_transmission_MPTorre\à transmettre\CD2_MPT_POMME_261005)
sur le CD2 emplacement : CD2_POMME_PROOF_14_11_2005\CD2_MPT_POMME_14112005

C.Pocho
10/11/2005

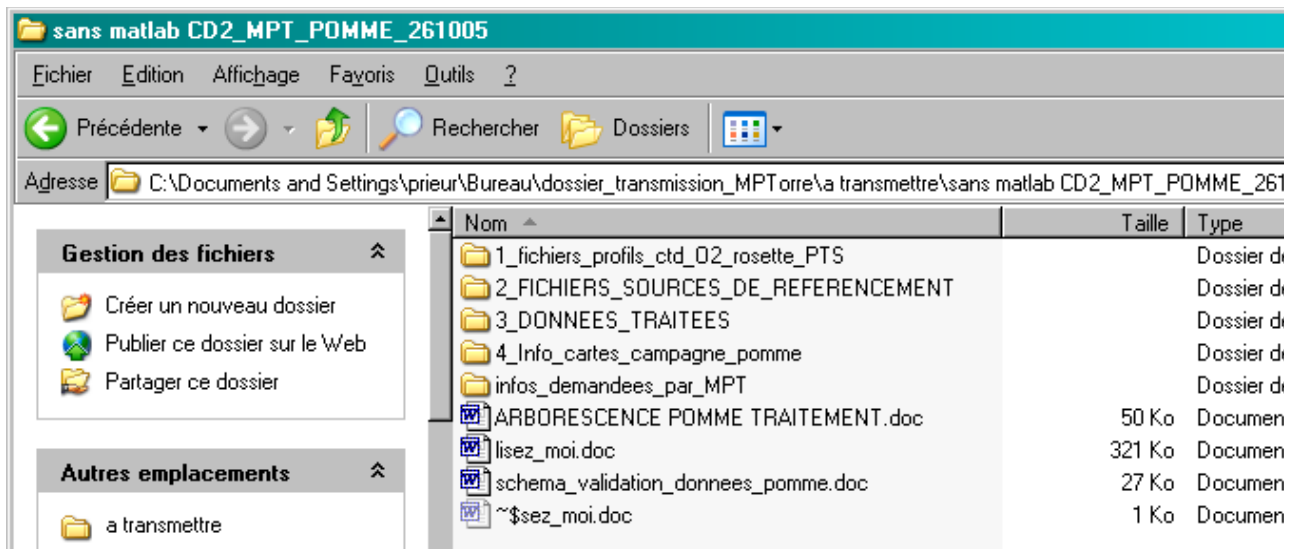
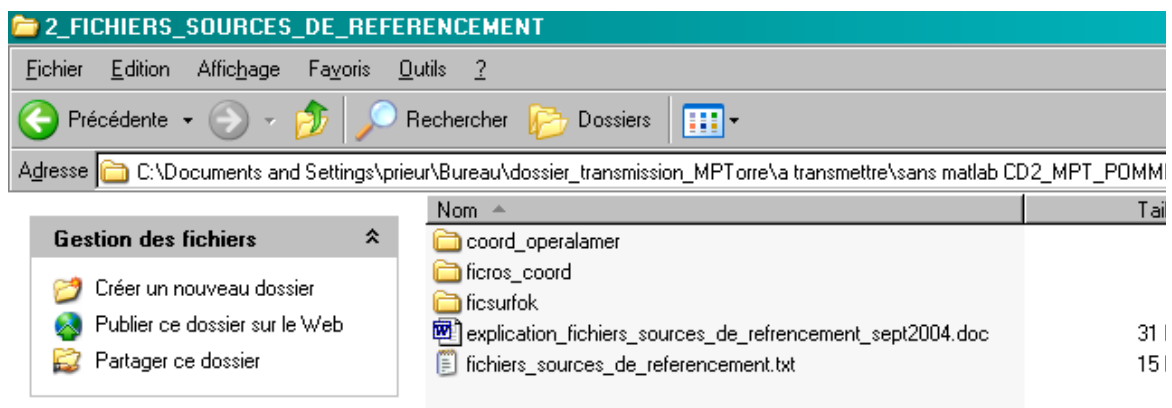
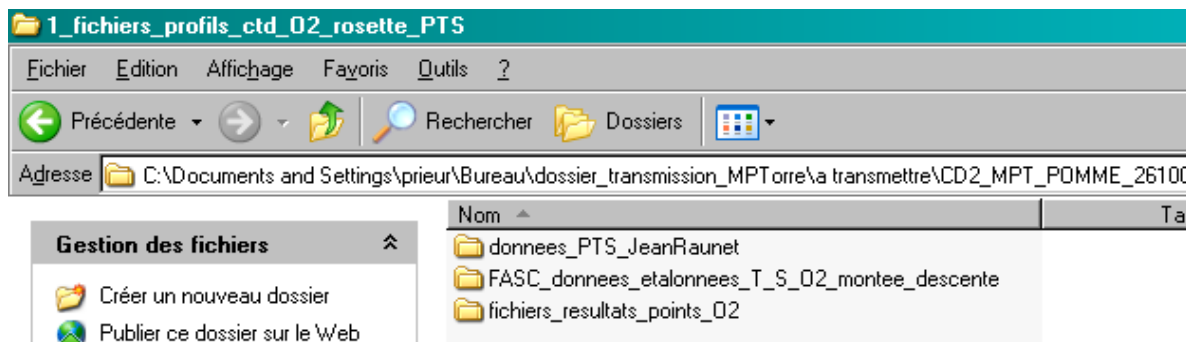
2^{ème} cd_rom des données POMME transmis à Marie-Paule Torre

✓ Rappel de quelques infos et suggestions :

- Un 1^{er} cd-rom a été transmis à MPT le 17/01/05 (avec prog matlab)
- LA REGLE PRINCIPALE DU TRAITEMENT : ON NE SUPPRIME RIEN
- Les données sont :
 - Soit des données validées par référencement uniquement (ces données n'ont pas de flag, on concatene des données originales à une trame de référencement, à une entête. C'est le cas des données issues du site web POMME : donpomweb)
 - Soit des données crossvalidées avec flag (c'est le cas du fichier rosette qui comprend les données CTD aux profondeurs des bouteilles et les données des bouteilles : O2, NO3, PO4, Chla, TChla, fluo)
- Attention il ne faut pas dissocier les fichiers
- Les profils CTD O2 transmis : il s'agit bien des profils de descente et de montée issus de l'étalonnage de la descente.
- Ces profils sont distincts de ceux de gens qui ne contenaient que P, T et S Il convient donc de mettre sur le site Proof les deux versions , ainsi qu'une 3eme (explications dans le rapport Pomme validation étalonnage). Les deux premières versions ont été déjà transmises au site pomme Web et restent valides
 - La première est celle de Jean Raunet (on peut d'ailleurs donner son nom à cette version pour simplifier les choses). Cette version comprend les données de physique étalonnées et validées : P, T, S + inter5m (fichiers interpolés tous les 5 mètres) + inter1m (fichiers interpolés tous les mètres)
 - La seconde est celle réalisée par Louis Prieur et moi. Il s'agit des dossiers « oxyglistd » et « oxyglistup » où P, T, S sont légèrement différent de ceux de Jean.
 - La troisième est la version précédente à laquelle nous avons ajouté les profils étalonnés de Chla et d'atténuation.

Ce sont les entêtes des fichiers de Jean ont servis de référencement pour les stations (cf inventaire stations + fichiers)

- Au sein du dossier « result_def » se trouvent le dossier « profbouteilles.zip » comprenant les prof des bouteilles (à rajouter je pense dans les infos campagne pomme)
- Pour permettre aux gens de se repérer par rapport aux nombreuses diffusions et versions des documents mis en ligne sur le site web de pomme et sur le ftp de Louis Prieur, ne serait-il pas conseillé de rappeler à chaque fois la date de transmission des données ?
- Les fichiers rosette peuvent être transmis en .txt pour être sûr de ne pas rencontrer de problème au niveau de la lisibilité de la légende (en fonction de la largeur de la colonne) et au niveau des décimales (conserver la plus grande précision possible).
- Nous avons choisis le format .csv pour transmettre les données qui étaient sous format excel.
- Unité de mesure de référence pour l'O2 dissous est $\mu\text{M}/\text{kg}$ mais nous avons souvent travaillé en ml/l. Il serait donc intéressant de pouvoir mettre les deux unités de mesure.
- Ne pas oublier d'insérer les cartes synoptiques qui se trouvent sur le site web de pomme
- Les rapports vont suivre

Contenu du cd-rom :**Détail du contenu de chaque dossier :**

3_DONNEES_TRAITEES

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Dossiers

Adresse C:\Documents and Settings\prieur\Bureau\dossier_transmission_MPTorre\à transmettre\sans matlab CD2_MPT_POMME

Nom	Taille
AUTRES_DONNEES_POMWEB	
DONNEES_AT_TCD2	
DONNEES_ROSETTE_CROSSVALIDEES	
explication_fichiers_donnees_traitees_sept2004.doc	41 K
schema_recapitulatif_fichiers_transsitepomme_sept2004.doc	59 K

Gestion des fichiers

- Créer un nouveau dossier
- Publier ce dossier sur le Web
- Partager ce dossier

4_Info_cartes_campagne_pomme

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Dossiers

Adresse C:\Documents and Settings\prieur\Bureau\dossier_transmission_MPTorre\à transmettre\sans matlab CD2_MPT_POMME

Nom	Taille	Type
cartes_sites_pomme		Dossier
coordonnees_distances_stations_leg1_leg2		Dossier
fascicules_graph_ctd_pomme123		Dossier
operations_a_la_mer_pomme		Dossier
POMME_Rosette_bt1_PRELEV		Dossier
result_def		Dossier
Liste_des_n°_stations.doc	27 Ko	Document
Liste_des_n°_stations_towyo.doc	26 Ko	Document
liste_n°station_des_sites_leg2_POMME.doc	26 Ko	Document
liste_stations_100m.doc	24 Ko	Document
liste_stations_n_existent_pas.doc	24 Ko	Document
liste_stations_profondes.doc	25 Ko	Document
possitesp1p2p3.rtf	14 Ko	Texte

Gestion des fichiers

- Créer un nouveau dossier
- Publier ce dossier sur le Web
- Partager ce dossier

Autres emplacements

- sans matlab
- CD2_MPT_POMME_261005
- Mes documents
- Documents partagés
- Poste de travail

infos_demandees_par_MPT

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Dossiers

Adresse C:\Documents and Settings\prieur\Bureau\dossier_transmission_MPTorre\à transmettre\sans matlab CD2_MPT_POMME

Nom	Taille	Type
DONPOMWEB_nov2003_sels_nutritifs_Entrecasteaux		Dossier
a_consulter_sommaire.rtf	200 Ko	Texte

Gestion des fichiers

- Créer un nouveau dossier

CHRONOLOGIE_TRANSMISSION_POMME

10/01/2005 - C. Pocho

CHRONOLOGIE DES TRANSMISSIONS EFFECTUEES :

1. **Sept 2000 : fichiers rosette bruts** (dossier « rosbrut » : fichiers « rospoxlegymb.xls ») comprend les données brutes (P, T, S, O2 sonde) issues de la rosette et de la sonde CTD
2. **Sept 2002 : étalonnage montée** (fascicules 1^{ère} version, dossier « FASCICULES »)
3. **Mai 2002 et mars 2004 : ftp oceane** : Données Pomme CTD et thermosalinographes transmises le 13 mai 2002 et actualisées en mars 2004 (actualisation données pomme 3 thermosalino uniquement avec navigation)
4. **Août 2003 : CD-Rom « pommeresultat_avecO2 »** qui comprend les résultats des traitements de l'oxygène pour les profils CTD rosette (étalonnage descente et montée avec fascicules graphiques non définitifs (un dossier tso2diagd et un dossier profilsd) 2^{ème} version, programmes et résultats : fichiers oxyglistd et oxyglistup) et les résultats de la procédure de détection de valeurs douteuses pour les fichiers rosette (étude de validation des fichiers rosette avec indication des bonnes valeurs et des valeurs douteuses (programmes, résultats, fichiers « travficrosDEFp111_flag_variat_2_s.xls » et « travficrosBRUTp112_flag_variat_2_s.xls »)
5. **Février 2004 : Fascicules graphiques tso2diagd et profilsd 2^{ème} version définitifs** (dossier refdesc : « fascicules_tso2diagd_profilsd ») mis sur ftp
6. **Sept 2004 : ftp + CD-Rom « FICHIERS_TRANSSITEPOMME_SEPT2004 »** comprenant les « FICHIERS_SOURCES_DE_REFERENCEMENT » (« coord_operalamer », « ficros_coord », « ficsurfok ») et les « DONNEES_TRAITEES » (« DONNEES_ROSETTE_CROSSVALIDEES » avec les données rosette crossvalidées, « graphiques_controle_2 », « controle_structplotros », les « DONNEES_AT_TCO2 » avec « Donnees_alcalinite_carbinorgdissous_pomwebaout2003 », « Donnees_TCO2_D_Lefevre », « TCO2_alcalinite » et les « AUTRES_DONNEES_POMWEB » avec « ficros_invent_param_sitewebpom » et « ficrosdcoord_donpomweb »)

FTP OCEANE : ETAT AU 11/01/2005
--

- ✓ ftp://oceane.obs-vlfr.fr/pub/prieur/pomme/fichier_transsitepomme_sept2004 (10/09/2004): comprenant les « FICHIERS_SOURCES_DE_REFERENCEMENT » (« coord_operalamer », « ficros_coord », « ficsurfok ») et les « DONNEES_TRAITEES » (« DONNEES_ROSETTE_CROSSVALIDEES » avec les données rosette crossvalidées, « graphiques_controle_2 », « controle_structplotros », les « DONNEES_AT_TCO2 » avec « Donnees_alcalinite_carbinorgdissous_pomwebaout2003 », « Donnees_TCO2_D_Lefevre », « TCO2_alcalinite » et les « AUTRES_DONNEES_POMWEB » avec « ficros_invent_param_sitewebpom » et « ficrosdcoord_donpomweb »)
- ✓ <ftp://oceane.obs-vlfr.fr/pub/prieur/pomme/graphsitesbruts> (01/07/2003) : graphiques des sites leg 2 en pdf
- ✓ <ftp://oceane.obs-vlfr.fr/pub/prieur/pomme/pobrut> (01/07/2003): données brutes pomme 1, 2, 3 (.BL, .CON, .DAT, .ROS) : fichiers descente et montée (un fichier par station) classés par leg
- ✓ <ftp://oceane.obs-vlfr.fr/pub/prieur/pomme/pommeresultat> (04/01/2005) au sein duquel nous avons un sous-répertoire :
 - « result_def » qui comprend les « fascicules_graph » (04/01/2005) des ctd valid stations et « ficsurfok » (04/01/2005) qui contient les données météo et de navigation
 - « ficrosette » : fichiers excel 2000 des données rosette de Pomme 1, 2, 3 de nom « rospoXlegYm.xls » et « rospoXlegYs.xls »
 - « ficctd1mbrut » (04/01/2005)
 - « ficctd5mbrut » (04/01/2005)
- ✓ ftp://oceane.obs-vlfr.fr/pub/prieur/pomme/pommeresultat_avecO2 (10/09/2004)
 - « graphresdef » (06/02/2004)
 - « oxygdef_caro_rosette » (31/03/2004) qui comprend les fascicules des controles graphiques par structure de plot superposé et les graphiques pour TCO2, AT et TCO2_alcalinité
 - « oxygdef-caro-profils » (06/02/2004) : fascicules profils CTD de présentation pomme 1, 2 et 3 (données étalonnées T, S, O2 descente et montée)
 - « resdef02 » (09/04/2004) :
 - « fichiersprofilsCTD02 » (30/03/2004) qui contient les programmes et les résultats (oxyglistd et oxyglistup) des fichiers ctd O2 rosette
 - « fichiersrosettefinaux » (09/04/2004) qui comprend les fichiers rosette finaux (ficroscontr_cfinauxpxly) Attention, il existe une autre version des fichiers rosette finaux (ficroscontr_cfinauxpxly_Atal ou Thal) où l'on a inséré les coordonnées (rajout de 8 colonnes) qui ne figure pas au sein de ce dossier mais au sein de « fichier_transsitepomme_sept2004 »
- ✓ ftp://oceane.obs-vlfr.fr/pub/prieur/pomme/result_def (30/07/2004) qui comprend :
 - « ficsurfok » (25/08/2004)
 - « chrono_operations_pomme.zip » (28/10/2003) : chronologie des opérations à la mer classée par leg
 - « inventCTD_Pomme.zip » (04/02/2003)
 - « profbouteilles.zip » (28/10/2003) : classé par leg

NE SONT PAS ENCORE PRESENTS SUR LE SITE WEB POMME :

- ✓ Idonnées de sept 2004 + les mises à jour d'oct 2004 concernant les données pomme web + invent param donpomweb
- ✓ les fascicules des profils CTD up and down 2^{ème} version