

データベース作成と品質管理の方法について

1 データベース作成方法

「海洋生物データの統合化技術と炭素循環解析への活用に関する研究」(以下 本プロジェクト研究)においては、角皆・才野(1999)(財)日本水路協会海洋情報研究センター(MIRC)(2001, 2002)(株)関西総合環境センター(2002)を参考にして栄養塩・クロロフィルデータベース(以下 本データベース)の作成とデータの品質管理を行った。本データベースの作成・公開は、以下の手順で進められた。

栄養塩とクロロフィル a のデータ及びメタデータの収集

中央水産研究所海洋生産部物質循環研究室(以下 本研究室)にて、各水産研究所から栄養塩とクロロフィル a のデータ及びメタデータを収集した。

メタデータとは、あるデータがどのように集められたかを示す情報で、「属性データ」とも呼ばれる。主に、観測・分析方法の異なるデータの利用および品質管理のために用いられる。本プロジェクトのように複数の研究機関によってデータが取得された場合、それらをまとめて解析するには重要な参考資料となる。本プロジェクトではメタデータとしてインベントリー情報(観測期間、海域、観測点の緯度経度、船名、機関と部局、責任者名、観測項目等)およびサンプリング・分析方法の情報をデータと共に収集した。

データの品質管理

本研究室において、データを得るための過程(サンプリングから分析、データ処理まで)において生じた人為的なエラーを発見し、それを除去するために後述の方法でデータの品質管理を行った。個々のデータに対し「問題はみられず」「問題があり検討を要する」のいずれかのフラグを立てた。「問題はみられず」とした場合はそのまま「公開」とすることを提案した。一方、「問題があり検討を要する」としたデータには、明確に異常であると判断できるものから、おそらく問題はないが多少気になる点があるものまで、様々なレベルの問題が含まれている。故に、問題ありと判断した理由と共に、各水研のデータ提供者に報告を行った。そして、個々のデータの取り扱いについて「公開」「値はあるが信頼性が低いと付記して公開」「削除」のいずれかを提案した。

データ取り扱いの決定

の結果をもとに、データ提供者がデータの取り扱いを最終的に決定した。問題ありと疑われるが明確に異常であると判断できなかったデータは、「値はあるが信頼性が低い」とコメントで付記して「公開」という扱いにした。

データベースの公開

品質管理を経たデータは中央水産研究所のホームページを経て公開すると共に、日本海洋データセンター(JODC)にも送付し、公開を依頼した。

第三者がデータを利用する際には、データを得るために用いられた方法やデータの質に関する情報が極めて重要である。そのため本データベースでは、サンプリング・分析方法、品質管理方法、品質管理の結果を、できる限りデータと共にホームページで公開することにした。

各観測航海で実施されたサンプリング・分析方法は、選択式のチェックリスト「サンプリング及び分析の方法について」を Word ファイルでデータ提供者に送付し、その回答をもとに「サンプリング及び分析の実施内容」としてまとめた。また、各観測航海のデータセットに施行されたデータ品質管理チェック（詳細は次章にて述べる）については、その結果を「データ品質管理の実施内容」としてまとめた。

「サンプリング及び分析の方法について」「サンプリング及び分析の実施内容」「データ品質管理の実施内容」はそれぞれ PDF ファイルに変換し、ホームページ内の各観測航海のページにリンクし、データと併せて閲覧できるようにした。

2 データの品質管理方法

品質管理作業には データを得る過程で同時に行うチェック（観測時の採水器の誤作動のチェックや研究室での分析時のチェックなど）と、データが一通り集められた後に行うチェックの二つの段階がある。 は、実際に観測・分析したデータ提供者にしか行えない。 はデータ管理者やデータ利用者といった現場に居合わせなかった者が行う場合もあれば、データ提供者自身が提供前に自分でデータセット全体を見回すために行う場合もある。本データベースでは、 のデータ収集後の品質管理のみを行った。

以下に、本データベースにおいて行われた品質管理項目を示す。本データベースのデータ提供機関は複数であるが、観測項目及び観測方法が統一されていないため、各観測航海のデータセットで品質チェックが実施できた項目が異なる。そのため、それぞれのデータセットに対して実施した品質管理項目は、本ホームページ内の各観測航海のページにリンクされている PDF ファイル「データ品質管理の実施項目」を参照していただきたい。

1) 観測日

実在しない日付になっていないかチェックした。

2) 観測点の位置（緯度経度）

観測点が陸域になっていないかチェックした。

3) 採水試料のクロロフィル a 値と CTD の蛍光センサーの値の関係

通常、短期間に同一海域で得られた蛍光センサーの測定値とクロロフィル a 濃度の間には良好な直線関係がある。採水キャストごとに両者を比較することにより、クロロフィル

a 値のチェックを行った。ただし蛍光/クロロフィル a 濃度の比は日射の影響を受けて低下するため (Holm-Hansen et al., 2000)、昼間の表層水のデータにおいて比が低下し直線関係からのずれが生じている場合は、問題はみられずと判断した。

4) ウィンクラー法により得られた DO と CTD の DO センサーの値の関係

CTD の DO センサーの測定値と比較することにより、ウィンクラー法により得られた DO のチェックを行った。ただし DO センサーによる測定においては、センサーの反応時間の関係から、極端な濃度変化を示す分布があった場合や、センサーの反応時間を越える速度で観測層を通過した場合は、濃度を正確に測定できないことがある (Owens and Millard, 1985)。そのため、センサーとウィンクラー法による鉛直分布パターンの傾向が一致しているかどうかについてのみ検討を行った。

5) 鉛直分布パターン

水温、塩分、密度の鉛直分布により示される水塊構造、混合層の深さや密度躍層の位置と栄養塩、クロロフィル a、DO (AOU) の鉛直分布の関係をチェックした。また栄養塩の各項目間および栄養塩と AOU の間で、鉛直分布パターンの比較を行った。

外洋のデータに関しては、主に下記の点に注目してチェックを行った。

- ・混合層中の濃度のばらつき。
- ・栄養塩躍層の上端が密度躍層の上端と同じかそれよりやや深い位置にあるか。
- ・表層近くで水塊分布が複雑である場合を除き、栄養塩躍層以深で深さと共にNO₃、PO₄、Si(OH)₄がなめらかに増加しているか。
- ・表層近くの水塊分布が複雑であったり、密度躍層の上部に光合成により生じたと考えられる DO のピーク (AOU の極小) が見られたりする場合を除き、密度躍層以深で深さと共になめらかに DO が低下 (AOU が増加) しているか。
- ・NO₃とPO₄の鉛直分布パターンがおおむね一致しているか。
- ・光合成により生じたと考えられる密度躍層の上部のAOUの極小 (DOのピーク) がある場合を除き、AOUの極大、極小がある深度がNO₃及びPO₄の極大、極小とおおむね一致しているか。
- ・NO₃、PO₄、AOU の極大よりもSi(OH)₄の極大が深い層にあるか。
- ・NO₂とNH₄の濃度は一貫して低濃度、もしくは混合層や密度躍層上部で濃度が高くそれら以深で低下するという分布になっているか。
- ・クロロフィル a は混合層や栄養塩躍層付近で濃度が高く、それら以深で低下するという分布になっているか。

沿岸においては水塊の分布や生物活動の時間的空間的変動が大きいいため、鉛直分布パター

ンのチェックという方法は外洋の場合ほど有効な手段ではないが、以下の点に注目して検討を行った。

- ・混合層中の濃度のばらつき。
- ・海面付近の栄養塩やクロロフィル a の濃度が突出して高かった場合、そこが低塩分となっていて陸水の影響で説明できるか。

6) 密度に対する分布のパターン

観測点ごとに NO_3 、 PO_4 、 Si(OH)_4 、DO (AOU) の密度に対する分布図を作成し、水塊分布が複雑である表層近くを除き、それらの分布が滑らかな曲線になっているかどうか確認を行った。鉛直的な濃度勾配が小さい深層においては、深さに対するプロットよりもこの方法の方が曲線から外れている点を見つけやすい。

7) 栄養塩の各項目間の関係

観測点ごとに NO_3/PO_4 プロットを描き、直線から外れている点がないかどうかチェックした。また $\text{NO}_3/\text{Si(OH)}_4$ もしくは $\text{PO}_4/\text{Si(OH)}_4$ のプロットがなめらかな曲線になっているかチェックを行った。

8) 同一航海での近接観測点の値との比較

NO_3 、 PO_4 、 Si(OH)_4 、DO (AOU) に関して、近接する観測点の中～深層の等密度等塩分の水の間で濃度の比較を行い、等密度等塩分の水の間での過大な変動や観測点間の系統的な誤差がないかどうかチェックした。

9) 鉛直断面図により水塊構造と栄養塩及び AOU 分布を比較

ライン観測により得られたデータセットでは、鉛直断面図を作成して水温、塩分、密度により示される水塊構造と NO_3 、 PO_4 、 Si(OH)_4 、AOUの分布の比較を行った。水塊分布の変化を伴わない不規則な濃度変化や、観測点間の系統的な誤差がないかどうかチェックした。

10) レンジチェック

水温、塩分、栄養塩、クロロフィル a、DO のデータに関して、World Ocean Database 1998 Version 2.0 (WOD98v2) の、北太平洋の外洋域及び沿岸域のデータに対する水深ごとの最大値・最小値の閾値と比較を行った (MIRC、2001)。ただし沿岸域は変動が大きいため、過去の値から逸脱していても他の項目の分布との関係から異常ではないと思われる場合は、問題はみられずと判断した。

11) 鉛直勾配チェック

水温、塩分、栄養塩、クロロフィル a、DO の観測深度間の鉛直勾配を計算し、WOD98v2 の閾値を用いて比較を行った (MIRC、2001)。

12) 密度逆転チェック

各観測層で密度を算出し、密度の上下逆転が MIRC (2001) の閾値を越えた大きさに起こっていた場合には、水温と塩分の値のチェックを行った。ただし沿岸域は変動が大きいため、基準となる値から逸脱していても他の項目の分布との関係から異常ではないと思われる場合は、問題はみられずと判断した。

13) 同じ海域の過去の観測値との比較

中～深層のNO₃、PO₄、Si(OH)₄、DO (AOU) 濃度に関し、同じ海域の過去の観測点の等密度等塩分の水との比較を行い、観測航海の間で系統的誤差があるかどうかチェックした。ただし誤差があった場合でも人為的なミスによるものと判断されない限りは修正を行わず、誤差についてのコメントをつけてデータを公開した。

14) 透明度

海洋における常識的な値 (数十 m 以内) になっているかどうかチェックした。また沿岸において採水を行った最深層より透明度が深い場合は、海底の深さをこえていないか確認を行った。

3 データ表の説明

- ・データシートの 1 カラム目はデータ識別コードである。
 - H: ヘッダー項目名の行
 - D: データの行
 - C: 品質管理フラッグを 1 としたデータに対するコメントの行
- ・品質管理フラッグ
 - 0: 特に問題が見られない場合
 - 1: 値はあるが信頼性が低い場合 (そう判断した理由をコメント欄に記載)
 - 9: データなし (この場合数値の欄には - と記入されている)
- ・有効数字は表示されている通り。

4 参考文献

(株)関西総合環境センター (2002): 平成 13 年度地球環境産業技術研究開発事業 新工

エネルギー・産業技術総合開発機構委託 二酸化炭素の海洋隔離に伴う環境影響予測技術研究開発 - 海洋調査及び CO₂ 隔離能力評価技術の開発 - 成果報告書

Holm-Hansen O, Amos AF and Hewes CD(2000): Reliability of estimating chlorophyll a concentrations in Antarctic waters by measurement of in situ chlorophyll a fluorescence. Mar. Ecol. Prog. Ser., 196,103-110

Owens, W. B. and R. C. Millard (1985): A new algorithm for CTD oxygen calibration. J. Phys. Oceanogr., 15, 621-631.

角皆静男・才野敏郎監修(1999): 全球海洋フラックス合同計画(JGOFS)における観測・測定手法、日本海洋データセンター、175pp

財団法人日本水路協会 海洋情報研究センター(2001): MIRC Ocean Dataset 2001 Documentation、MIRC テクニカルレポート No.1、169pp

財団法人日本水路協会 海洋情報研究センター(2002): 化学系データ品質管理手法、70pp