

重力データの基本ファイルについての考察

徳弘 敦*

A FEW COMMENTS ON THE BASIC DATA FILE OF THE MARINE GRAVITY

Atusi Tokuhiro*

Received 1977 October 12

Abstract

Since the opening the marine gravity measurement in the Hydrographic Department, the acquired data have been accumulated to tens of thousands. The tendency of such data accumulation was accelerated when the new T.S.S.G. with a minicomputer designed for on-line real time data processing was installed aboard the *Shoyo* in 1972 and the *Meiyo* in 1974. The problems with which we are concerned for the huge data stock today are stated on the three subjects, 1) the data format basically works for the data exchange, and new techniques to retrieve the needed data out of the data stock must be developed for the data users, 2) the data sequence in the data files has a significant meaning from a view-point of data user, 3) COBOL must be useful in the scientific data handling.

Finally, as an example, mesh area data file drew from the on board data files are proposed as a basic data file from which user's file can be compiled.

1. まえがき

昭和42年の水路部観測報告「天文測地編」に当部による最初の海上重力測定の成果として報告された測点数は253点であったが昭和52年の同報告ではその数が9,256点に増大した。昭和52年度までの測得総数は報告されたものだけでも約5万点に及んでいる。このデータの大部分は水路部が昭和42年に開始して51年に終了した大陸棚の海の基本図のための測量によって取得したものであって、これに続く計画として考えられている大洋の海の基本図のための測量が実施されれば毎年1万点以上の割合で更に十年以上にわたってデータの蓄積が続くことになるであろう。このほか日本周辺海域での重力測定としては東京大学海洋研究所や通産省地質調査所でも実施されている。海上重力値の測定精度も着実に向上しており、海上重力測定を開始した当時では10ミリガルの精度を保つことも容易ではなかったが、現在では1ミリガルの測定精度の水準に達してきた。本来この程度の精度では海上の重力値そのものは時間的に不変であるが、古い観測海域をより向上した精度で再測する意義は十分にある。データは単に蓄積すればよいものではなく、膨大なデータがいかに良く管理されているかによってその価値が決定

* 海洋研究室 Marine Research Laboratory

づけられる。そのことは観測データの処理を電子計算機によって実施する場合決定的なものとなる。

2. 海上重力の場合

水路部で実施している海上重力測定業務はその発足当初から電子計算機の使用が前提となっていたから、現在のように小型で高性能の電子計算機が手軽に利用できなかった初期の数年間、航海後の素データの一括処理に要する手間と時間が大きな負担であった。今から十年前の時点で、一観測業務の遂行に電子計算機の利用を前提としたことによって、その後起こるであろう諸問題を的確に見通すことは無理なことであったが、今から振り返ってわれわれは三つの大きな課題を抱えていたということができる。

一つは重力測定の効率的運用とデータ品質の向上のために電子計算機はいかに貢献し得るかということ、二つは観測の自動化をどういう形で具体化できるか、三番目として蓄積される一方のデータをデータの供給者としてデータの利用者にどのようにサービスすべきかということ、の三点である。1および2の課題は本稿の主題から外れるが第3の課題との関連を追って概括しておく。1について水路部で用いている海上重力計 T.S.S.G. (Tokyo Surface Ship Gravimeter)の大きな特徴は、センサー部は垂直加速度に感応する振子であり、その振動周波数を重力値へ変換し、重力値を適当な方式で出力するまでのかなり面倒な計算を含む一切のデータ処理を電子計算機のソフトウェアに任せていることである。ハードウェアとしての T.S.S.G. だけでは重力測定の装置として機能し得ないということは、測深機やプロント磁力計などと比べても大きな相違であった。海上重力測定業務を開始した後に、データの品質にかかわるような重要な改良や開発が、すべてソフトウェアだけの負担で処理されてきたということは特筆すべきことである。2についてはいわゆるミニ・コンピュータの出現が課題解決の突破口となった。この段階に至ってデータの扱い方が個々のデータとしてより、一航海分のデータのようなデータ・ファイルとして扱う考え方が定着し始めた。つまり、データをカード、紙テープ、磁気テープ等のどの記録媒体に乗せるのか、データをどのように出力させるか、測位データや測深データとの結合は実時間処理かそれともそれぞれのファイルを介しての結合か、一連の作業のどの部分は人手で処理されなければならないか、といった観点から全体的なデータ処理の流れの合理性が検討された。現在稼働中の T.S.S.G. はミニ・コンピュータを装備して、データの処理ならびに測定の進行を自動的に制御し、船上での出力はさん孔紙テープの上にエトベス補正を含まない測定されたままの重力値として得られる。

3番目の、蓄積される一方のデータ管理をどのような考え方で具体化させて行くかを考察するのが本稿の主題である。問題点は次のように整理される。

(1) データの時間的、空間的分布にはそれぞれ個々の意味がある。従ってこれらについて適当な順序づけがなされていることがデータ利用上の不可欠の条件である。しかし例えば時間的順序に並べたものが緯度の順序には並べないように一意的な順序づけは不可能である。これをどう処理するか？

(2) データをどのように保存するか？ データ処理はほとんど完全に電算化されているから出力すべきデータをカード、磁気テープなどのいずれにとり出すことも雑作のないことである。しかし数万点のデータ量になると単に磁気テープに乗っているということだけでは問題は完全には解決していない。

(3) 成果をどのように利用するか？ 重力異常図は成果利用のための端緒にすぎない。測地データの解析手法についての研究成果は少なくない。利用すべきデータ量の多さがいろいろな解析を試みる上での大きな障害になっている。

3. データ管理上の問題点

蓄積され続けるデータをどの様に管理すべきかについて、主要な問題点を考えてみる。

(1) データの順序づけ

水路部で測得した重力データの大部分は水路部観測報告(天文測地編)に掲げたような書式で磁気テープ・ファイルとして収納されている。一部のカード・ファイルについても近く統一された磁気テープ・ファイルとして編集される計画である。これらのファイルは観測行動単位のサブ・ファイルの集積であり時間順序に編成されているが、地域的にはサブ・ファイルごとに行動海域に限定されている。従ってサブ・ファイル単位で見れば時間順序に並び、位置的には航路に沿った順序で並んでいる。しかしこれらのサブ・ファイルの集合である重力のデータ・ファイル全体でながめると、時間的にも(「昭洋」と「明洋」が同時期に行動することもある)、位置的にも順序づけが乱れてくることは避けられない。従って最も極端で念の入ったデータ検索としては全ファイルを通して内容を見ることになってしまう。この手数はデータの蓄積に比例して増大する。目的によっては若干のデータの拾い落しは容認できる場合もあるだろうが、ある海域での浅所を調べるといった目的では一点のデータの見落としも許されなくなる。

経度、緯度あるいは時刻といったキーデータを指定して、それによってデータを順序づけることを論理的順序と呼ぶならば、磁気テープ上に並んだデータの順序は物理的に固定されているから、物理的順序と呼び分けるのが適当である。もしキーデータを一つ特定すれば、その論理的順序に物理的順序を揃えることは計算機の力を借りれば容易なことである。しかし磁気テープを別のキーデータでながめれば論理的順序づけはでたらめになる。磁気テープは大容量の記憶媒体としては際立って優れたものであるが、データの順序づけには甚だ硬直的である。

データ利用者から見れば、データ量が多ければ多い程順序づけられた秩序に従って処理する能力は重要である。大容量磁気ディスクはこの問題の解決に大きな役割を果たした。書物を例にとると検索のため一巻の本の最初のページから最後のページまで全部調べあげるのではなく、索引によって必要なページを知り、そのページの中だけを調べればよいという様な使い方ができる。現在水路部で使用している NEAC 2200/500 の磁気ディスクの容量は1台当たり約 8×10^6 字(水路要報の約4,300ページ分)であるが、近い将来更新されるであろう新機種では更に十倍位い容量が増大することが計画されている。この飛躍は従来に増して電子計算機の利用形態がデータ・ファイル主導形に変わりつつあることを物語っている。

(2) 記録の書式

データに対してどのような処理を行うのかということを記述する文書様式については項目を改めて述べるが、ここではデータをカードや磁気テープ上に書き並べたり、プリンターで処理結果を出力したりする場合の、1件分データ内における各項目の並べ方や、数字のけたの割当てについての仕様としての書式をとりあげる。

重力データであれば1件分のデータとは少なくとも日付、時刻、位置、水深、重力の値がセットになったものである。FORTRAN を用いてプログラムを記述する場合、このような1件分のデータの書式は FORMAT 文で与えられる。計算機はこの FORMAT 文を参照しながら、例えばカード上の重力値として与えられた数字を正しく抽出して、重力値として計算処理が進められる。従ってデータ・カード(あるいは磁気テープ)にとって大切なことは、データの並び方やけた数が FORMAT 文で規定するところと一致することであり、そのような事情は COBOL を用いても同じである。ここではこのことを“データはプログラムに従属する”と

言うておく。

データ量が増すに従って書式の不統一はデータ利用の障害となる。現実には重力に限って見ても世界中の各機関で保有するデータ量は膨大なものであり、国際的に標準化された書式はまだ出来上がっていない。もちろん書式の統一の努力はなされるべきであるが、当面の対応策として各データ・ファイルに書式を明示しておくことによって、書式が異なっていることがデータ処理上の大きな障害にはならないというのが最近の考え方のようである。

書式はデータ交換については不可欠のものとして今後も生き残るのであろう。しかしデータ利用の立場から見れば、書式を介して一件一件の記録の中から必要な項目を探し出して行くという様な方法でなく、必要なデータの準備はデータ・ユーザーのプログラムから離れて、データを提供側のサービスとして工夫される必要がある。

(3) 電子計算機用語

FORTRAN を技術用、COBOL を事務用と区別するのは正確でない。前者は計算式記述向きであり、後者は大量のデータ・ファイル処理向きというような特徴で使い分けられている。測地や地球物理のような大量のデータに依存する分野ではCOBOLの特徴がもっと活用されるべきである。実際、これらの計算ではそれに必要なデータを準備する段取りや手続きに要する手間がプログラム作成の中での主要な部分を占める。

FORTRAN でも COBOL でもデータがプログラムに従属する点において本質的な違いはないが、後者ではデータをファイルの単位で処理しようという考え方が進んでいる。データ管理の思想はFORTRAN からは全く感じられないがCOBOL にはそれが認められる。

今一つ電子計算機用語に関して大切なことは、これが計算機に対してプログラムであると同時に、処理されるべき仕事内容の仕様の文書化の役目を果していることである。データが管理されるだけでなくデータがどのように利用されているかの現状は握のしやすさも同時に重要である。文書化は業務内容の厳密な記述であるとともに、読みやすさが要求されるCOBOL にはそのような配慮がなされている。PL/1 のような新しい言語も使われ始めているが、ここでは単にFORTRAN と COBOL の比較で述べた。言語はプログラマーの選択に任されるものではなく業務の内容と進め方から選ばれるべきものであろう。

4. データ管理の進め方

海上重力観測成果が大陸棚海の基本図の中の「重力異常図」として刊行されている以外に、水路部観測報告「天文測地編」によって全データが数値記録として公にされているのは、将来における様々なデータ利用の可能性を予想しているからである。データの持つ潜在的価値はデータ解析技術、他のデータとの組合せの相乗効果として現われてくるものであり、データの蓄積そのことにデータ収集業務の意義がひそんでいる。重力異常図は一つの成果であるが、電子計算機の利用が前提となっている現在のデータ利用者から見れば、機械可読 (machine readable) なデータは不可欠なものである。カードや磁気テープを用い、しかも利用者の都合に合わせた内容や書式で提供できるようにするのがデータ・サービスの目標である。

以下における説明の都合上各種ファイルに名前をつけておく。先ず観測航海の単位で作られるサブ・ファイルを観測別ファイル、これらを用いてメッシュ化された海域に分類し直したファイルをメッシュ化ファイル、データ・ユーザーのために用意されるファイルをユーザー向けファイルと呼び分けることにする。

(1) 航海別ファイルからメッシュ化ファイルへ

重力データは水深などと同じように点で分布している。この点の密度は一般的には大陸棚上でほぼ一様でそ

れより深海底に向かって急激に点の数が少なくなる。一方データを使う場合には海域を指定してその中のデータを全部抽出するとか、あるいはその海域での平均、最大値、最小値等を知りたいというのが通常の利用のしかたである。このような観点から、航海別ファイルは海域を経度、緯度の適当な間隔で規則正しく区分したメッシュに分類し直しておくのがよい。メッシュは東西および南北の二次元の広がりを持つがこれに一連の番号を付ける方法は、例えば Marsden Square No. と呼ばれる方法等で実用的に使われている。このような番号をここではメッシュの参照番号と呼ぶことにする。このメッシュはデータ・ファイル上のデータ集団に対応することになるが、このデータ集団を格納すべき領域をページと呼んでおく。そこでこのページ No. を参照番号に合わせておけばメッシュを指定すれば直ちに必要なページが開かれるという様な使い方が可能になる（このページ No. が書物のページのように自然数として並ぶことは電子計算機にとって重要ではない）。メッシュの切り方は点の密度に応じて細分することも可能である。しかも細分化の細かさは参照番号を見ただけで分かるようにできる。ページ内での順序づけは先着順に並べられて、あるキーデータに基づく論理的順序に並べる必要はないであろう。ページ内の余白は新しいデータを受け入れるために必要である。ページが一ぱいになった時には新しいページを挿入することもできる。ただしこのようなメッシュ化データ・ファイルは磁気ディスク上でないと働かない。メッシュを余り細かく分けるのはデータ管理上望ましくない。従ってファイル上のメッシュを更に細分化した小区域にデータを分類しそこでの平均値を求めるといったような場合にはユーザ・プログラムの負担となる。

(2) メッシュ化ファイルからユーザー向けファイルへ

ユーザーが要求するファイルは多様である。メッシュ化ファイルをそのままユーザー向けファイルとして使うことは不可能ではない。しかしメッシュ化ファイルは膨大なものであり一つ一つのデータごとに検索するのは効率的ではない。計算に着手する前に自分に都合のよいユーザー向けファイルを用意するのが標準的な使い方になるはずである。そしてこの段階では COBOL を使うことになる。例えば日本周辺海域を経度、緯度の 30' ごとのメッシュに区分し、そこでの重力異常値あるいは水深の平均値を与えるようなユーザー向けファイルを作成するのがこの段階である。こうして作られたユーザー向けファイルは FORTRAN で処理できる。

5. むすび

最近の電子計算機利用の新しい動向として「オンライン処理を前提とした分散化処理」が目立ちはじめた。オンライン処理は言うまでもなくデータ通信の利用を前提としたものであり、ミニ・コンピュータの出現は観測システムの末端に位置するセンサーからのデータの入力、ならびに現場において必要なデータ処理の分散化に役割を果たした。この様な進め方はデータの生産量を飛躍的に増大させる反面、手作業を主力とした従来からの作業方式とは極めて整合性の悪いものとなってくる。

「オンライン処理を前提とした分散処理形態」では分散処理という言葉と裏腹に、データ・ファイルを核とした中央集中管理の考え方が前提となっている。このような核となるべきデータ・ファイルを基本ファイルと呼ぶならば、先に例示したメッシュ化ファイルはこの基本ファイルの考え方に近いものである。基本ファイルはデータの流通、処理という角度から見た水路業務の一つのモデルであり、業務の実態が忠実に反映されるべきものである。

基本ファイルは重力データ収集作業と重力異常図の作成作業を形式的にも実質的にも分離する。測量船の行動は「海の基本図」の刊行計画と直接的に結びつかず、水路業務に必要な基本ファイルの整備が行動の目的となる。

水路業務の成果として作られる図誌は観測行動と一対一に結びつく必要はなく、利用者の要望により、あるいは新しい解析技術の開発により生産されることになるであろう。

地形補正をていねいにほどこしたブーゲー異常図,あるいは地殻均衡重力異常図のような成果は,今述べたような基本ファイルの整備を前提としなければ手がつけられないようなデータ依存性の高い作業であり,蓄積されたデータにはこのような密度の濃い計算を可能にするような潜在的価値がひそんでいる.