

Оценка на потоците и обемите от данни при автоматизирано събиране на океанографска информация

Атанас В. Палазов

Институт по океанология, БАН (Варна)

Океанографските изследвания се характеризират с необходимостта да се събират и обработват големи обеми от данни. По тази причина широко приложение са намерили автоматизираните системи както за събиране (Ястробов, 1982), така и за обработка на събраната океанографска информация (Алексеенко и др., 1989). При определянето на необходимите технически и програмни средства за изграждане на системите за автоматизация на научните изследвания в океанологията от съществено значение е да се познават потоците и обемите от данни, които се обработват (Загорский и др., 1987; Краус и др., 1987; Соломатин и др., 1987). Само при това условие може да се постигне ефективно използване на автоматизираните системи.

Основните източници на океанографски данни са научно-изследователските кораби, буйковите станции и бреговите лаборатории (Алексеенко и др., 1989; Ястробов, 1982). От всеки от тези източници може да постъпват различни по характер данни, в зависимост от насочеността на изследванията: хидрологични, хидрофизични, хидрохимични, хидробиологични, геофизични и др. По-съществен обаче е начинът на извършване на изследванията (Алексеенко и др., 1989; Калнини и др., 1983; Ръководство..., 1977), като тук различаваме две основни

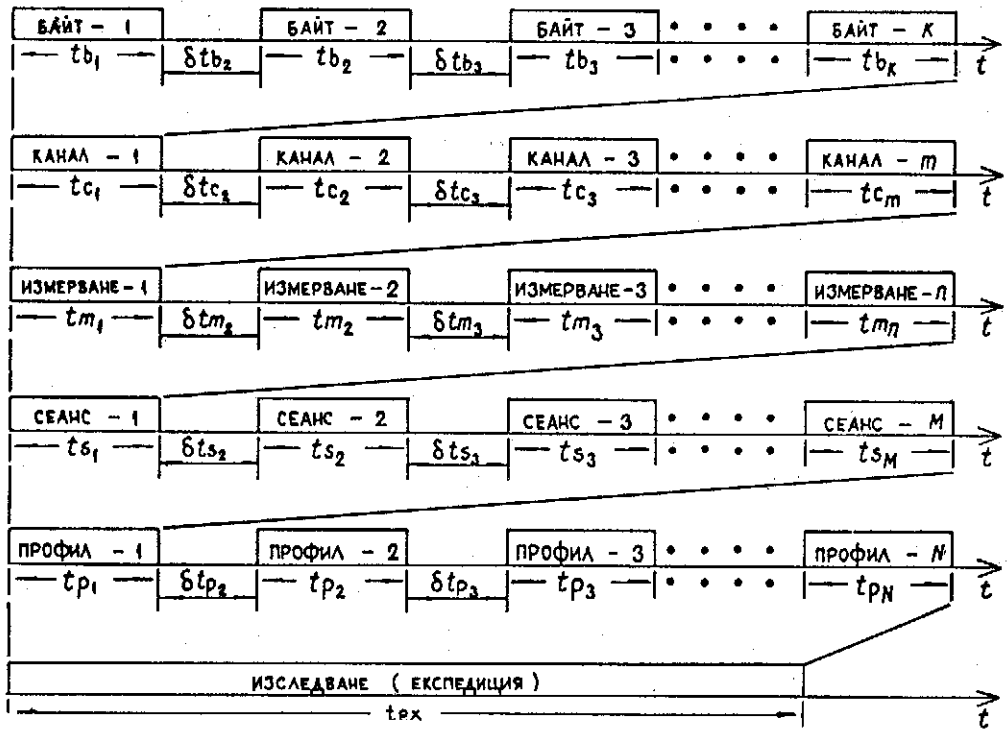
групи: режимни наблюдения и насочени изследвания.

Режимните наблюдения се извършват с цел да се проследят дълговременните изменения на изследваните величини. Обикновено се характеризират с малки потоци и обеми от данни, но има и специфични случаи, при които потоците и обемите са значителни.

Насочените изследвания решават конкретни задачи за разкриване на определени зависимости между изследваните величини или за получаване на представи за моментното състояние на величини и полета както бавно, така и бързо изменящите се. Характеризират се в повечето случаи със сравнително големи потоци от събирани данни.

При оценка на потоците и обемите от данни в научните изследвания, в това число и в океанологията, обикновено се разглеждат конкретни прибори и устройства (Ястробов, 1982; Загорский и др., 1987). В океанологията обаче методиката на изследването влияе съществено върху потоците и обемите от данни, което налага тя да бъде отчетена при оценките, като бъде създадена подходяща методика за оценка.

От гледна точка на автоматизираното събиране на океанографска информация насочеността на изследванията и начинът на извършването им имат второстепенно



Фиг.1. Времениаграма на постъпването на данните от океанографските изследвания

значение. По-съществено в случая е познаването на начина, по който постъпват данните. Независимо от типа на океанографското изследване, може да се приеме, че данните постъпват съгласно времениаграмата, показана на фиг.1. Тъй като в случая се предполага автоматизирано събиране с помощта на компютризирано автоматизирани системи, на времениаграмата е показано постъпването на дискретни данни. По същите съображения е избрана и основната информационна единица – байт. По същество времениаграмата отразява постъпването на данните след преобразуването и кодирането им (напряжение-цифри, честота-цифри и т.н.). Такъв подход е оправдан най-малко по две причини. Първо, съвременните измервателни прибори и системи извършват сами преобразуването и предават на изхода си кодирани величини. Второ, от значение за изследванията е само тази част от сигналите, която се преобразува и кодира, тъй като само тя се из-

ползва по-нататък при обработката. Следователно, когато информацията постъпва в аналогов вид, следва да се отчетат само частта от нея, която ще се преобразува в числов вид, а тази част се определя в зависимост от методиката на изследването.

На времениаграмата (фиг.1) е показана единицата изследване (експедиция), под която следва да се разбира логично завършена част от изследователския процес, въз основа на която се решават определени изследователски задачи.

В рамките на изследването (експедицията) се извършват N броя профили, които представляват технически обособени части от изследователския процес, състоящи се от M броя сеанси. Под сеанс се разбира част от изследването, състояща се от n броя последователни измервания на определени физични величини.

Измерването се състои от последователно събиране на данни от m броя кана-

ли, като по всеки канал постъпват данни, характеризиращи дадена физична величина, стойността на която се описва с мощта на k броя байта.

Количеството на профилите — N , на сеансите — M за всеки профил, на измерванията — n за всеки сеанс, на каналите — m за всяко измерване и на байтовете — k за всеки канал, в общия случай е различно, но на практика много често е постоянно.

Времената, показани на фиг.1, имат следните значения:

t_{bi} — време на постъпване на i -тия байт от определен канал, $i = 1 \div k$;

δt_{bi} — време между края на постъпването на $i-1$ -вия байт и началото на постъпване на i -тия байт от канала, $i = 1 \div k$;

t_{ci} — време на постъпване на данни от i -тия канал от определено измерване, $i = 1 \div n$;

δt_{ci} — време между края на постъпването на данните от $i-1$ -вия канал и началото на постъпване на данните от i -тия канал, $i = 1 \div m$;

t_{mi} — време на постъпване на данни от i -тото измерване на определен сеанс, $i = 1 \div n$;

δt_{mi} — време между края на постъпването на данните от $i-1$ -вото измерване и началото на постъпване на данни от i -тото измерване, $i = 1 \div n$;

t_{si} — време на постъпване на данните от i -тия сеанс от определен профил, $i = 1 \div M$;

δt_{si} — време между края на постъпването на данните от $i-1$ -вия сеанс и началото на постъпване на данните от i -тия сеанс на профил, $i = 1 \div M$;

t_{pi} — време на постъпването на данните от i -тия профил на дадено изследване, $i = 1 \div N$;

δt_{pi} — време между края на постъпването на данните от $i-1$ -вия профил и началото на постъпване на данните от i -тия профил на изследването, $i = 1 \div N$;

t_{ex} — време за извършване на изследването (експедицията),

където k е броят на байтовете; m — броят на каналите на измерване; n — броят на измерванията на сеанс; M — броят на сеансите на профил; N — броят на профилите на изследване.

За определяне на потоците и обемите от данни при океанологичните изследвания

се предлага следната методика.

1. Построяване на времедиаграмата на постъпване на данните по модела, представен на фиг.1.

2. Определяне на характеристиките на времедиаграмата.

3. Оценка на потоците и обемите на базата на построената времедиаграма.

Под поток от данни се разбира количеството данни, което преминава през информационния канал за единица време, а под обем на данни Q — количеството данни, преминало през информационния канал за определен, фиксиран период от време.

Най-характерен момент при научните изследвания е измерването. Това е така, защото именно при измерването се формират стойностите на измерваните физични величини от съвкупността от байтове, получени по измервателните канали. Следователно при измерването се формира основният поток от данни. Изхождайки от измерването, потокът от данни за всеки информационен канал може да се определи по формулата (подобно на С о л о м а т и н и др., 1987)

$$(1) \quad q_i = \left(\sum_{j=1}^m K_j \right) / t_{mi}$$

където q_i е потокът от данни за i -то измерване; K_j — броят байтове на канал, $j = 1 \div m$.

Ако измерванията съдържат еднакъв брой и вид канали, тогава формула (1) се модифицира по следния начин:

$$(2) \quad q = \left(\sum_{j=1}^m K_j \right) / t_{mi}$$

където q е потокът от данни за разглеждания информационен канал. Тук $q = q_i = \text{const}$ следва да се разглежда като частен случай, който обаче се среща доста често.

Характерът на потока зависи от отношението на времената t_{mi} и δt_{mi} ; при $t_{mi} > \delta t_{mi}$ има равномерен поток; при $t_{mi} \approx \delta t_{mi}$ има квази равномерен поток; при $t_{mi} < \delta t_{mi}$ има пулсиращ поток.

Характерът на потока, както и неговата големина са най-важните характеристики,

оказващи влияние върху системата за събиране на океанографски данни.

Удобно е да се въведе ефективен (среден) поток от данни за всеки елемент от изследването, както следва:

$$(3.1) \overline{qc_i} = K_i / tC_i \quad - \text{ за канал;}$$

$$(3.2) \overline{qm_i} = \left(\sum_{i=1}^n qc_i \right) / tm_i, \quad - \text{ за измерване;}$$

$$(3.3) \overline{qs_i} = \left(\sum_{i=1}^n qm_i \right) / ts_i, \quad - \text{ за сеанс;}$$

$$(3.4) \overline{qp_i} = \left(\sum_{i=1}^M qs_i \right) / tp_i, \quad - \text{ за профил;}$$

$$(3.5) \overline{q_{ex}} = \left(\sum_{i=1}^N qp_i \right) / t_{ex} \quad - \text{ за изследване.}$$

От съществено значение е кой от потоците от формули (3) характеризира най-добре информационния канал. По съображенията, изтъкнати по-горе, за такъв следва да се приеме потокът за измерване.

При пулсиращ поток е възможно да се получи нереална оценка на големината на потока, ако се използват формули за ефективния поток. В този случай ефективният поток може да е малък, а всъщност в момента на получаване на данни да е налице голям поток от данни. Затова следва да се отчита и характерът на потока при оценка на неговата големина. Използването на ефективния поток за оценки е допустимо само при равномерен и в някои случаи на квази равномерен поток. Във всеки случай следва да се предвиди възможност за реакция на системата на пулсациите на потока от данни, като се осигури известен запас от пропускателна способност.

На базата на средните потоци може да се получи и оценка за обемите от данни, получени по време на изследването

$$(4) \quad Q = \sum_{i=1}^N \sum_{j=1}^M \sum_{i=1}^n \sum_{p=1}^m K_{ijp}$$

където K_{ijp} е броят на байтовете на канал за съответните измерване, сеанс и профил.

Съществуват характерен клас изследвания, при които каналите съдържат постоянно число байтове, измерванията — постоянно число канали, сеансите — постоянно число измервания и профилите — постоянно число сеанси. Такива изследвания могат да бъдат наречени хомогенни, като хомогенността може да се разпростира до различни нива: профил, сеанс, измерване или канал. При напълно хомогенно изследване, т.е. на ниво профил, потокът и обемът от данни се определят от формулите

$$(5.1) \quad q = (mk) / tm,$$

$$(5.2) \quad Q = NMnmk.$$

Съществен е въпросът на каква времебаза да се определя потокът от данни tm_i или $(tm_i + \delta tm_i)$. Първата характеризира по-прецизно потока, но втората дава по-реална оценка.

Предложените формули позволяват да се оценят потоците и обемите от данни при изследванията на базата на построената времедиаграма. Такава оценка беше направена за информацията, постъпваща от научното оборудване на НИК „Академик“. В състава на научното оборудване

Т а б л и ц а 1. Потоци и обеми от данни, постъпващи от паушните прибори на НИК „Академик“, на базата на експедиционно изследване с продължителност 90 дни

Източник	Поток от данни, В/с	Тип на потока	Обем на данните, МВ
Ехолот	16	равномерен	200
MNS 2000	12	квази равномерен	0,1
Доплеров лаг	1	равномерен	0,01
МТО станция	15	квази равномерен	0,02
Магнитометър	2	равномерен	2
СНСП	4000	равномерен	20000
Клайн 530Т	12000	равномерен	60000
Вълнови измерител	200	пулсиращ	10
Сонда СТД	12	равномерен	20
RCM 4	200	пулсиращ	10

на кораба влизат (Технико-икономическо задание, 1986): ехолот ЕА 200, навигационна система MNS2000, доплеров лаг, метеорологична станция, магнитометър, едноканална система за непрекъснато сеизмоакустично профилиране (СНСП), хидролокатор за страничен обзор Клайн 530Т, вълнови измерител, сонда CDT, записващ измерител на течения RCM 4.

Количеството на измервателните канали и информацията, постъпваща от научните прибори, бяха определени на базата на техническата документация на приборите.

Литература

Алексеев, В., А. Никитин, Б. Платонов. 1989. Автоматизация обработка данных в океанографских исследованиях. Киев, Наукова думка. Загорский, В., И. Пугачев, А. Ярусов. 1987. Информационно-измерительные системы коллективного пользования. Минск, Наука и техника. Идеен проект за създаване на интегрална система за автоматизация на научните изследвания, провеждани на НИК „Академик“. 1987. С., НПЛ „Програма“ — БАН. Калинин, А., В. Калинин, Б. Пивоваров. 1983. Сейсмоакустических исследования на акваториях. М., Недра. Краус,

М., Э. Кучбах, О. Г. Вошни. 1987. Сбор данных в управляющих вычислительных системах. М., Мир. Руководство по гидрологическим работам в океанах и морях. 1977. Л., Гидрометеоиздат. Солома-тин, М., Р. Шертвигис, М. Махшицев. 1987. Выбор микро ЭВМ для информационных систем. М., Высшая школа. Технико-икономическо задание за разработване на интегрална система за автоматизация на научните изследвания, провеждани на НИК „Академик“. 1986. С., БАН. Ястребов, В. (ред.). 1982. Принципы построения технических средств исследования океана. М., Наука.

Резултатите от направените оценки са представени в табл.1. Въз основа на получените оценки е направен изборът на технически и програмни средства за СА-НИ на НИК „Академик“ (Идеен проект, 1987).

Предложената методика и формули позволяват въз основа на техническите характеристики на измервателните прибори и начина на извършване на изследването да бъде направена оценка на потоците и обемите от данни при автоматизирано събиране на океанографска информация.

An estimation of the data streams and the amount of data in automated acquisition of oceanographic information

Atanas V. Palazov

(Summary)

When determining the hard- and the software necessary for completing the systems intended for automation of the scientific research in oceanology, it is of major importance to know the streams and the amount of data which have to be processed.

A time-diagram is proposed of the measured data acquisition, in accordance with the method applied for carrying the research investigations. On the basis of this time-diagram, formulae have derived for estimating

the data streams and the amount of data.

The method for the estimation of the data streams and the amount of data includes the creation of the research time-diagram, the estimation of the time for each element of the time-diagram and the calculation of the streams and the amount of data.

By means of the formulae obtained an estimation was made of the streams and of the amount of data received from the scientific instrumentation of the RV "Akademik".

Постъпила на 30.09.1990 г.