海洋环境监测数据处理技术流程与方法研究*

路文海[1,2],向先全[2],杨翼[2],付瑞全[2]

(1.中国海洋大学 青岛 266100;2.国家海洋信息中心 天津 300171)

摘要:为适应海洋环境保护的信息化建设,海洋生态环境监测数据采集完成后,需要开展监测数据的统一和规范化处理。文章以各地上报的海洋环境监测数据集为对象,提出了海洋环境监测数据的业务化处理技术流程,包括数据标准化处理、齐全性检验、站位基础信息质控、站位监测参数数据质控和数据输出等 5 个流程,规范了海洋环境监测数据的业务化处理流程,提高监测数据的处理效率和水平。

关键词:海洋环境;监测数据集;处理技术流程;标准化处理;质量控制

中图分类号:P76; X321 **文献标志码**:A **文章编号**:1005-9857(2015)02-0058-05

1 引言

近年来,我国海洋环境监测能力切实得到了加强,业务体系不断创新发展[1]。目前全国已形成了国家、海区、省(自治区、直辖市)、市、县 5 级环境监测网络,已建有海洋环境监测机构 200 多个,沿海 11 个省、市、自治区以及 5 个计划单列市、44 个地级市均建立了海洋环境监测机构,沿海县级海洋环境监测机构建设也取得了明显进展。

随着我国海洋环境监测工作的深入开展,监测工作从最初的单一特征性参数调查,发展至今形成了从监测任务、监测区域、监测站位、监测参数等方面业务化和延续性的监测调查,涵盖海洋环境状况、海洋环境风险、海洋环境监管和公益服务等监测类型,形成包括海洋生物多样性、海水、沉积物、赤潮、海洋溢油、陆源入海、海水浴场等 20 余项监测任务,以及海水水质、水文气象、沉积物质量、生物质量、浮游植物、浮游动物等 40 余种监测要素的业务化监测体系。

海洋环境监测数据的规范处理和管理在海洋监测中是很重要的一个环节[2-3]。随着监测数据的不断获取与更新,已形成了海量的数据积累[4],每年可获得各类海洋环境监测数据超过 200 万个。这些数据包括不同的监测任务和监测要素,出自不同的监测机构,由于各地监测机构上报汇总的监测数据格式和类型多样,且标准化程度不高,这给监测数据的处理带来了一定的困难[5-6]。

为了适应海洋环境保护的信息化发展,海洋生态环境监测数据采集完成后,需要完成监测数据的统一和规范化处理[7-8]。本研究以各地上报的海洋环境监测数据集为对象,提出了海洋环境监测数据的业务化处理技术流程,有别于海洋环境监测实验室内的数据处理与质量控制,适用于对全国海洋环境监测数据接收、汇总、审核、入库等过程的数据标准化处理与质量控制,可作为海洋环境保护标准数据库建设和海洋环境保护评价的前处理工作。

2 监测数据处理流程

通过对各地上报的海洋环境监测数据报表的综合分析,以海洋环境监测数据标准数据集为输出目标,提出了海洋环境监测数据集处理流程,主要包括:标准化处理、齐全性检验、站位基础信息质控、站位监测参数数据质控、数据输出共 5 个流程,各流程内部包含相对应的处理过程和方法,具体流程如图 1 所示。

3 监测数据分析与处理方法

3.1 监测数据集标准化处理

监测数据集标准化处理包括监测任务名称、

* **基金项目**:海洋公益科研专项项目"海洋生态红线区划管理技术集成研究与应用"(201405007);天津市科技兴海项目"天津沿海风险污染源排查和环境灾害风险评价及应用研究"(KJXH2012-28).

组织单位名称、监测区域名称、监测要素名称、监测参数名称及计量单位、数据类型及规范性等内容。

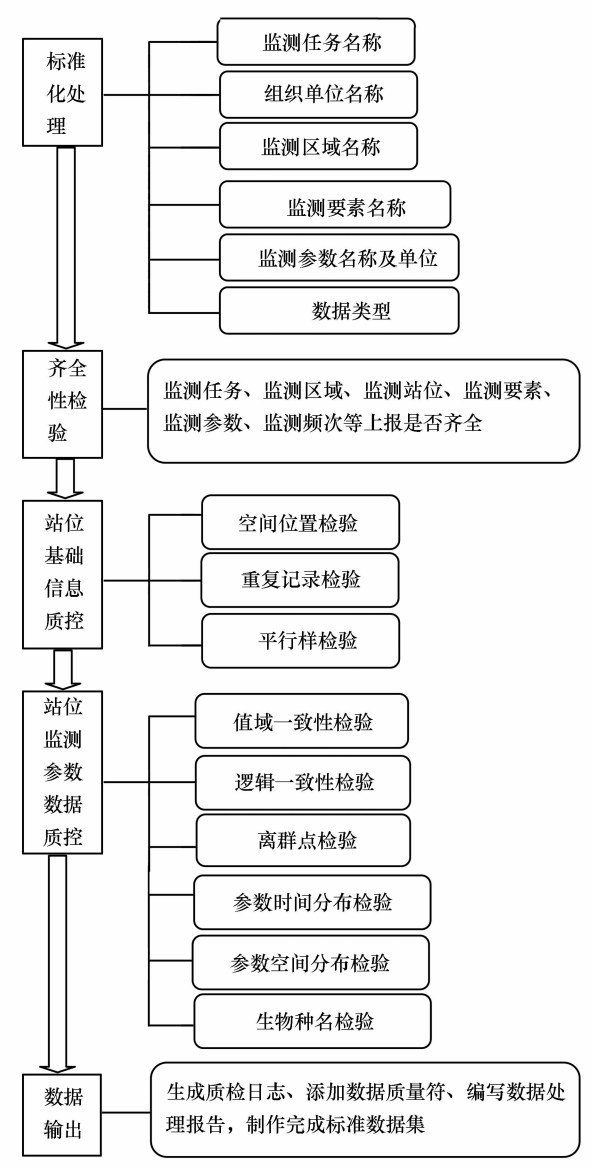

图 1　海洋环境监测数据处理技术流程

3.1.1　监测任务名称的标准化处理

以目前的全国业务化海洋环境监测任务为基础，对上报的监测任务进行标准化命名，如海洋生物多样监测、海洋大气监测，对不同填报的名称进行标准化处理。

3.1.2　组织单位名称的标准化处理

各地上报的组织单位比较混乱，有的上报了监测机构名称，有的上报了其隶属的行政部门名称，不利于监测任务的考核。根据国家海洋环境监测工作任务以及各海区年度海洋环境监测工作方案，目前组织单位主要包括国家海洋局局属单位、3 个分局、11 个沿海省（自治区、直辖市）海洋行政管理部门和 5 个计划单列市海洋行政管理部门，如国家海洋环境监测中心、国家海洋局北海分局、辽宁省海洋与渔业厅、大连市海洋与渔业局，对不同填报的组织单位进行标准化处理。

3.1.3　监测区域名称的标准化处理

由于各地方上报的监测区域不够规范，且很难表现出更多的区域信息，同时考虑到区域统计分析，因此需对监测区域进行规范化命名。监测区域命名结构为：沿海地区/海区＋沿海城市/特定区域＋名称，其中沿海地区/自然海区和名称字段不能省略，沿海城市/特定区域字段若无可以省略。如，辽宁葫芦岛赤潮监控区，广东近岸、福建厦门近岸、东海近海及远海，对不同填报的监测区域名称进行标准化处理。

3.1.4　监测要素名称的标准化处理

每个监测任务里包含了不同的监测要素，且不同的任务可能会监测相同的要素，因此需对监测要素进行规范命名，以便对相同的要素进行统一分析、数据量统计等。以目前的业务化海洋环境监测要素为基础，对上报的监测要素进行标准化命名，如水文气象、海水水质、沉积物质量、浮游植物和浮游动物等，对不同填报的监测要素进行标准化处理。

3.1.5　监测参数及单位的标准化处理

由于每个监测要素需要监测不同的监测参数，如海水水质需要监测化学需氧量、氨氮和溶解氧等。而每个监测参数的名称在写法上有不同的形式，如化学需氧量也可写为 COD，氨氮也可写为氨－氮或 NH_4-N 等，给数据的统计、评价带来一定的不便，因此有必要规范不同监测参数的名称。

另外，每个监测要素的单位也需统一规范。如重金属的锌元素，有的上报其参数单位为 mg/L，有的上报为 μg/L。在数据统一进入标准数据库时，需将单位统一。参照国际标准、国内海洋环境监测调查规范以及各地监测机构的填报习惯等，针对不同的监测任务和监测要素，对每个监测参数的名称及计量单位进行标准化处理。

3.1.6　站位基础信息的数据类型标准化处理

监测数据的类型包括数值型、字符型、布尔

型和百分比等。对站位基础信息如站位编号、经纬度、监测日期、水深和层号等的数据类型进行规范。

(1)站位编号。上报的站位编号大部分为字符型,但也有站位编号为 1、2、3 等,为数据库的统一管理,需统一转换为字符型。站位编号不规范主要有以下几个方面:① 站位编号英文大小写不一致;② 监测机构各自命名;③ 在站位编号上加"临""平行样"和"空白样"等字样。参照目前海洋环境监测站位编号规则,由任务编号、海区编号、类别编号和站位序号顺次排列组成。对站位进行统一编号。对于历史站位编号的确认,可通过核查相关的监测数据、核实年度监测方案、联系地方监测机构等方式,将站位编号统一。

(2)站位的经、纬度。上报的经纬度有两种形式:一个是小数形式,另一个是度分秒形式。为便于计算机的计算方便,目前统一为小数形式。由于经纬度的小数位数不一致,会导致部分空间定位有细微的差别。结合监测任务计划和实际监测情况,统一经纬度的有效位数,目前保留到小数点后 6 位。

(3)监测日期。上报的监测日期格式不一致,主要形式为:"2011－08－20""2011/8/20"、或为时间型等。现统一其形式为"2011－8－20",年份:填满 4 位;监测月份:1—12,月信息小于 10,前位无需补零。注意检查,监测年份是否为该年度;月份是否大于 12;日期是否在该月的自然日以内。

(4)采样深度与层号。部分地方监测机构在该填报"层号"的地方填写了采样深度,同时层号不统一,有的为中文——"表层""中层""底层";有的为英文——"S""M""B"。《海洋监测规范》中对水深和相应的采样层次进行了规范。对层号,统一用英文表示。其中:表层为 S;底层为 B;若只有一个中层用 M 表示,若为多个中层,则分别用 M1、M2、M3 等顺延表示。另需检查层号与层深的匹配情况,若层号为 S(表层),则采样深度应小于或等于 2 m;层号为 B(底层),则采样深度大于 3 m。部分填报机构填写层号时,出现表层填写"B"和底层填写为"D"的现象,可能是按"表层"和"底层"的首拼音字母填写造成的。

3.1.7 监测参数不规范类型的处理

监测参数的不规范类型问题,主要应注意以下几点。

(1)大于号、小于号。某些监测参数如重金属、大肠杆菌数等,其监测参数值上报中含有大于号或小于号。此类数据通常不影响其评价等级的判定,但会影响该类参数最大值、最小值、均值等统计的结果。可研究该参数的理化性质并联系地方监测机构,确认该参数的具体值大小。其缺省解决方法是删除大于号、小于号,以便该参数的统计及评价。

(2)未、无、"－"等字样。结合年度监测任务,联系地方监测机构,确认该监测参数是未被监测,还是低于检出限。未监测用空值表示;低于检出限用"未检出"表示。

(3)空格及其他无效字符。上报的监测数据中常含有空格及其他无效字符,使得计算机在识别、归类等过程中出现异常。可核查监测数据的内容和性质,确认为无效字符后,对数据值前、后含有的空格或其他无效字符进行删除处理。对经纬度空缺,可核查相关的原始上报数据集和年度监测工作方案,或联系地方监测机构;对层号空缺,可根据水深判断,或联系地方监测机构补缺;对某些监测参数值空缺,可结合年度监测任务,联系地方监测机构,确认该监测参数是未被监测,还是低于检出限,再根据判断结果给出规范填写。

3.2 监测数据的齐全性检验

海洋环境监测数据的齐全性检验,是以海洋环境监测方案为依据,检查监测方案中规定的监测数据是否全部上报完整。

首先对国家海洋环境监测工作任务以及各海区年度海洋环境监测工作方案进行分析,对监测工作方案进行信息解析,按空间维度、指标维度和时间维度对监测任务进行细化,空间维度包括监测站位、监测区域、管辖区域等,指标维度包括监测参数、监测要素等,时间维度包括监测时间等。其中监测站位、监测参数、监测时间是空间维度、指标维度和时间维度的最小单元,通过对最小单元的数据量统计,可获得其上一统计单元的数据情况。因此对海洋环境监测方案的解析按监测站位、监测参数和监测时间 3 个方面进行分解。

对照监测方案,检查接收的数据是否存在区

域、站位或频次等有空缺监测的情况。记录缺失的原因：可能由于某些缘故未能进行监测、地方调整了监测方案或地方漏报。仔细核查年度监测任务计划，联系地方监测机构确认。

3.3　站位基础信息数据质量控制

站位基础信息是指海洋环境监测站位的时、空、区域属性等基本信息，如监测区域名称、站位编号、监测日期、经纬度信息等，不包括监测参数值的信息。

3.3.1　空间位置检验

空间位置检验主要针对调查单位在站位信息汇总过程中可能出现的录入错误。将调查站位经纬度转换为十进制的单位后，通过利用 GIS 生成站位图的方式检查站位落点所在位置，看其是否落在规定的监测区域，对于断面上的调查站位，还要检查其是否明显偏离断面沿线。

同时还需检查“相同的站位编号，经纬度不同”和“不同的站位编号，经纬度相同”等数据空间位置精度的问题。

对于该类问题，可通过核查相关的监测数据、核对年度监测任务、联系监测机构确认等方法，予以更正。

3.3.2　站位基础信息一致性的检测

根据站位基础信息一致性检验方法，即监测区域、站位编号、站位经纬度、监测日期等基础信息决定一条数据记录，根据不同的监测任务和监测要素，分析站位基础信息一致性是否符合。

针对站位编号和经纬度不一致的情况，从空间位置检验是否合理，并核实监测方案进行解决。针对监测日期相同且站位编号相同等情况，判断两条记录的监测参数值是否完全一致，若完全一致则认为是重复记录；若不完全一致，可认为是平行样记录，并进一步核实。

3.3.3　数据记录重复的处理

海洋环境监测数据的上报过程中存在很多重复的数据记录，产生这种重复记录的主要有如下原因。

(1)地方上报数据时，重复上报了监测数据集，如 8 月份上报了 5 月份和 8 月份两份数据；年底将全年的监测数据再次上报。

(2)不同监测机构报送的重复数据，如属于上下两级监测机构(省、计划单列市)重复报送。

(3)地方监测机构监测人员填写报表时，将某些记录重复填写。

(4)地方监测机构监测人员填写报表时，将平行样的数据填写。

(5)数据集合并时，将曾经合并过的数据集再次合并。

对于重复的记录数据，在建立环境监测数据库中应做剔除处理。

3.3.4　平行样的处理

平行样数据只作为监测数据质量保证的辅助，在实际统计、评价和监测数据时需区别对待。一般来说，只有少数站位上报的数据是平行样。为了数据量统计、环境质量评价等的需要，对于平行样的记录数据，可将监测参数值进行求平均处理。

3.4　监测参数数据质量控制

3.4.1　值域一致性检验

在海洋环境监测中，每个监测参数有其对应的经验值域范围，通过值域检测规则对填报的监测数据按不同监测要素分别对每个监测参数值进行检验，对于超出值域范围的值，需进一步分析该区域其他站位、其他频次、周边站位的参数值情况，并结合监测任务性质以及超出值域比例，从而判断该参数值的可靠性。

3.4.2　逻辑一致性检验

某些监测参数间存在一定的逻辑关系，即监测参数与监测参数间存在某种相关关系，有些关系具有一定的规律性，根据逻辑一致性检验方法，对于不符合逻辑一致性的监测数据记录，应进一步同监测机构进行核实。

3.4.3　离群点检验

根据离群点检验方法，对于检验出来的离群点记录，应反复检查、核对，并进一步同监测机构进行核实，寻找可能的产生原因。若找不出产生可疑值的原因，则不能随意舍去数据。

3.4.4　生物种名检验

通过与海洋生物种中文名和拉丁名的规范化名称进行核对，可对不规范的物种名称进行规范处理。若与规范名称无法匹对，应进一步同监测机构进行核实。

3.5　数据输出

对文件进行批量检验处理，对于检验结果，

给出合理且足够详细的错误提示，并保存质检日志，使得数据便于修改。

为了区别一个数据是否进行了质检、是否通过质检，以及了解质检的情况，需要对质检过后数据增加一个质量控制符号，简称质量符。综合参考“国标 GB/T 12460－2006 海洋数据应用记录格式”以及“908 海洋化学标准记录格式”等质量符格式。其中，“908 海洋化学标准记录格式”中质量符 2 表示可疑倾向正确，3 表示可疑倾向错误，本研究将这两者综合考虑，记为可疑；另外，“908 海洋化学标准记录格式”中质量符 8 表示痕量，由于与“未检出”有一定的重叠，因此本研究只采用“未检出”。表 1 给出海洋环境监测数据的质量符及说明。

表 1　数据质量符说明

序号	质量类型	质量符
1	未质控	空或 0
2	正确	1
3	可疑	3
4	错误	4
5	大于号	5
6	小于号	6
7	未检出	7
8	缺测	9

一般来说，数值型的监测参数数据，对其质量检验出有问题的只能作为“可疑”处理，不宜随意修改或删除。除非经过专家经验检验，并经监测单位核实，可明确其为错误的，其质量符方可标注为“4”。对于监测站位基础信息，如监测日期、站位编号、经纬度、层号等，检验出有问题的，可根据检验情况，标注其质量符为“4”或“3”等。

按步骤完成监测数据处理流程后，可分年度或季度对处理的文件形成数据处理报告，并制作经标准化处理和质量控制后的标准数据集。

4　结束语

目前，全国业务化海洋环境监测数据逐年积累，但监测数据的处理尚未有一个通用的、规范化的处理技术流程，给监测数据的处理带来一定的困难。本研究从海洋环境监测数据管理的角度，以各监测机构上报的海洋环境监测数据为对象，研究了监测数据集的处理技术流程及方法体系。这一处理技术流程及方法体系的推广将会，进一步规范监测数据的业务化处理流程，大大提高监测数据的处理效率和水平，为海洋环境保护信息化持续健康发展提供高质量的数据保障。

参考文献

[1]　石斌，葛海祥，王荣平. 完善综合监测预警系统提高海洋环境监控能力[J]. 海洋开发与管理，2006，23(3)：39－42.

[2]　许自舟，宋德瑞，赵辉，等. 海洋环境监测数据质量计算机控制方法研究[J]. 海洋环境科学，2009，28(3)：320－323.

[3]　卜志国. 海洋生态环境监测系统数据集成与应用研究[D]. 青岛：中国海洋大学，2010.

[4]　石绥祥，雷波. 中国数字海洋：理论与实践[M]. 北京：海洋出版社，2011.

[5]　陈上及，马继瑞. 海洋数据处理分析方法及其应用[M]. 北京：海洋出版社，1991.

[6]　曹家新. 监测数据的审核及结果的分析与判断[J]. 四川环境，2002，21(2)：19－26.

[7]　夏登文，石绥祥，于戈，等. 海洋数据仓库及数据挖掘技术方法研究[J]. 海洋通报，2005，24(3)：60－65.

[8]　向先全. 基于水信息技术的渤海湾水生态环境特性及模拟研究[D]. 天津：天津大学，2011.