

海洋科学数据管理与可视化平台^①

李海涛¹, 关煜¹, 黄海广^{2,3}

¹(青岛科技大学 信息科学技术学院, 青岛 266000)

²(中国海洋大学 信息科学与工程学院, 青岛 266100)

³(温州市海洋渔业船舶安全救助信息中心, 温州 325000)

摘要: 海洋科学数据作为我国海洋科技创新发展的战略资源, 是一个可持续发展的要素. 然而海洋科学数据涉及学科众多、类型复杂、来源多样、格式迥异等特点, 数据处理和可视化处理难度大. 平台通过引入 WebGIS 技术, 针对海洋科学数据特点, 结合海洋科学数据业务, 实现海洋数据的一体化管理、分析与可视化处理.

关键词: 海洋科学数据; 数据管理; 质量控制; 数据解析; 可视化技术

引用格式: 李海涛, 关煜, 黄海广. 海洋科学数据管理与可视化平台. 计算机系统应用, 2017, 26(9): 62-68. <http://www.c-s-a.org.cn/1003-3254/6057.html>

Visual Management Platform for Marine Scientific Data

LI Hai-Tao¹, GUAN Yu¹, HUANG Hai-Guang^{2,3}

¹(School of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266000, China)

²(College of Information Science and Engineering, Ocean University of China, Qingdao 266100, China)

³(Wenzhou Ocean and Fishery Vessel Safety Rescue Information Center, Wenzhou 325000, China)

Abstract: As a strategic resource of technological innovation and development of the oceans in China, marine science data should be the element for sustainable development. However, marine scientific data involve many disciplines, complex types, diverse sources, and different formats, and hence it is difficult to process the data and visualization. In terms of the characteristics of marine scientific data, the platform achieves the integration, analysis and visual processing of marine data management by introducing the WebGIS technology, combined with marine scientific data services.

Key words: marine scientific data; data management; quality control; data analysis; visualization

科学数据是人类在认识自然、发展科技的活动中产生和积累的, 随着信息技术的进步与发展, 科学数据在科研活动中的作用日益提升, 已经从支撑科研活动的基础资源提升为支撑国家科技创新发展的宝贵资产和战略资源^[1-3].

然而传统的海洋科学数据管理平台以文字、图片等资料为主, 不能较好地表达海洋信息的空间位置及可视化展示, 这是其重要缺陷之一. 且海洋科学数据涉及学科众多, 具有收集零散、类型复杂、格式各异等特点^[3], 如何保证这些数据能够准确、有效地存储到数据库中, 也是海洋科学数据管理与可视化平台建设的

关键^[4]. 本研究通过引入 WebGIS 技术, 研究了海洋科学数据的管理与空间可视化方法, 验证并实现该平台的可用性.

1 系统概述

海洋科学数据管理与可视化平台是基于实现对海洋专项调查获取的海量珍贵的海洋科学基础数据和处理数据管理、分析、可视化, 采用 WebGIS 结合多种近海要素对空间数据进行分析处理和数据专题的可视化, 依靠先进的一体化读取解析、数据质控机制与整合分析的集成研究, 提供科学、可靠、有效的海洋数

① 基金项目: 青岛市创新创业领军人才(15-07-03-0030); 科技部 2015 年科技支撑计划(2015BAG20B00)

收稿时间: 2016-12-08; 采用时间: 2017-03-20

据,大幅度提高数据的管理水平和使用效率.

2 系统设计

2.1 系统总体设计

海洋科学数据管理与可视化平台功能是基于系统的业务数据的浏览、处理分析和 Skyline TerraGate 所发布的空间数据和服务^[5]进行系统编码的. 总体设计主要包括四个部分, 海洋科学数据服务, 海洋科学数据管理与可视化, 数据访问及数据库. 其总体设计如图 1 所示.

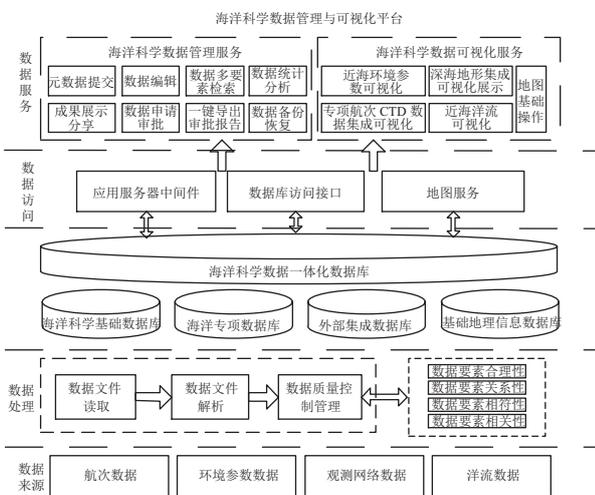


图 1 系统总体设计

2.2 系统功能设计

2.2.1 海洋科学数据管理

数据管理是利用计算机技术对数据进行有效的收集、存储、处理和应用的过程. 目的主要是针对海洋科学数据特点, 结合海洋科学数据业务, 构建一个以集成化、科普化为主, 分学科、专业化为辅的数据资源管理服务. 平台功能主要包括数据提交, 可通过文件导入、文件夹导入等上传形式, 并进行文件解析与质量控制, 对原始数据进行质量分析、异常提示、记录存档等处理; 数据多条件检索, 可按照时间范围、空间范围(矩形选、多边形选、圆形选)、学科分类、要素种类、要素大小等条件组合进行查询; 数据申请审批, 满足在线数据申请和离线数据申请两种服务方式, 高级用户对申请的数据进行审核, 并一键生成审核报告, 方便用户下载打印; 成果发布分享, 对采集到的海洋数据成果及产品进行展示与宣传; 对按照条件查询出的数据, 可以进行简单的数据统计, 也可以进行比较复杂的

运算分析, 比如相关性分析等, 统计结果将以表格、文字、图片的方式多方位的展示和描述; 数据备份恢复, 系统能支持设定自动备份间隔和备份位置, 系统进行定时的自动数据库备份任务. 其主要通过 Ajax 请求与系统服务端用户数据进行数据通信, 并返回相应数据及状态.

2.2.2 海洋科学数据可视化

平台利用 Skyline 服务端所发布的各种数据服务接口和功能处理服务, 提供海洋科学数据的可视化展示, 包括近海环境参数可视化, 近海站位调查中的环境参数数据具有横向分布和纵向分布两个特点, 依据此特点需要进行剖面可视化展示; 深海地形集成可视化, 把生物信息, 具体包含多媒体图片、音频、视频等信息叠加到海底地形上面进行展示; 近海洋流可视化, 将洋流数据按照时间顺序动态可视化展示; 专项航次 CTD 数据集成可视化, 将航次数据点分布到地图上, 可按照时间序列动态播放航次轨迹, 模拟航次运行过程等操作. 海洋科学数据可视化将海洋信息影像在地图上展示, 可供浏览分析, 其目的是以海洋科学数据为支撑, 具备二三维一体化与动静结合展示能力, 通过先进的可视化技术, 将海洋描绘的更加生动形象, 使其适用于专业人员更好的研究海洋.

2.3 数据库设计

平台系统数据库主要包括海洋科学基础数据库、海洋专项数据库、外部集成数据库及基础地理信息数据库. 针对海洋科学数据复杂, 并且具有地理空间特性, 通过对数据进行基础信息和空间信息的分离, 将基础数据和空间数据按照数据库的表空间独立进行存储, 再通过表的主键和外键等约束, 将基础数据和空间数据进行二者联合, 设计海洋科学数据的一体化数据库, 如图 2 所示.

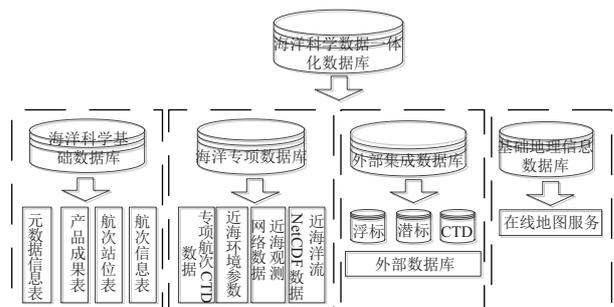


图 2 数据库设计

1) 海洋科学基础数据库. 本平台与海洋数据业务相关的数据直接以数据表的形式在 MySQL 数据库中存放, 通过数据访问接口进行访问. 相关的业务数据主要包括元数据信息、产品成果发布信息、航次站位信息等.

2) 海洋专项数据库. 平台的海洋专项数据库包括航次、环境、观测及洋流等可视化专题信息, 其中包含空间地理信息坐标和空间相关属性等信息, 采用 ArcGIS Engine 对空间数据进行插值计算, 并生成相关图层.

3) 外部集成数据库. 包括浮标库、潜标库、CTD 信息库等外部集成数据库. 各个数据库中的表信息与平台数据具有集成相关性, 利用其外部的表数据与平台数据进行结合, 能够更详细的反映数据的真实状态及信息.

4) 基础地理信息数据库^[6]. 平台的地图为影像基础空间数据类型, 基础空间数据采用必应卫星地图, 采用球面墨卡托投影类型, 包含 1-19 级的显示级别. 由 TerraBuilder 制作基础地理信息地图 MPT 文件, 包括底图和中文注记. 利用 TerraGata SFS 发布在线地图, 不需要在系统数据库中进行组织.

3 系统实现关键技术

3.1 异构海洋科学数据一体化读取解析技术

3.1.1 数据特点简介

本平台针对近海科学专题数据的不同数据类型与格式对采集的数据进行分类整理, 主要为专项航次 CTD 数据、近海环境参数数据、近海观测网络数据及近海洋流 NetCDF 数据. 已整合的数据资源类型及格式如表 1 所示.

表 1 数据类型及格式表

数据类型	格式	文件后缀
专项航次CTD	ASCII文本文件	.txt/.CAL等
近海环境参数	Excel表格式文件	.xls等
近海观测网络	Excel表格式文件	.xls等
近海洋流NetCDF	NetCDF数据文件	.nc等

其中专项航次 CTD 数据记录主要包括站位、时间、经纬度、水深、温度及盐度等相关信息. 近海环境参数数据记录主要包括站位、水深、经纬度、五项营养盐(SiO₃、NO₃、PO₄ 等)、pH 值、叶绿素等相关信息. 近海观测网络数据主要包括浮标号、风速风

向、气温气压、流速流向、浊度叶绿素、平均波高平均波周期等相关信息. 近海洋流 NetCDF 数据主要包括经纬度、时间、流速、流向等相关信息.

3.1.2 一体化组件式读取解析设计^[7,8]

异构海洋科学数据一体化读取解析组件类图基于工厂模式进行设计, 如图 3 所示, 对于同一接口, 可以无限扩展实现层, 针对不同的数据类型, 通过建立模块的形式, 例如.txt 格式数据对应文本文件解析模块, 各类文件数据又对应各自的解析模块. 一体化组件式读取解析设计在各组件的基础上进行分类, 形成模块, 然后在每个模块内进行真正的读取解析. 并且根据解析数据的格式, 可以无限扩展解析模块, 遵循设计模式的“开闭原则”, 从而达到代码复用、增加可维护性的目的, 其一体化组件式读取解析设计如图 4 所示.

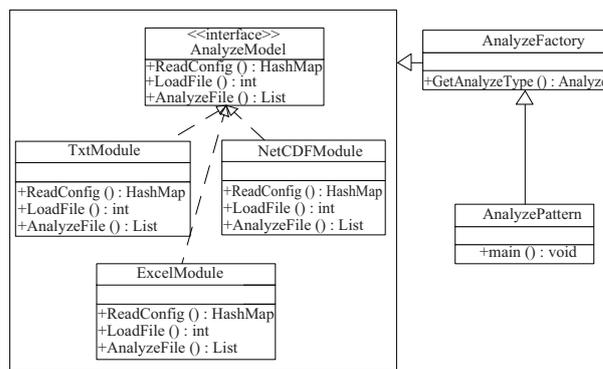


图 3 异构海洋科学数据一体化读取解析类图

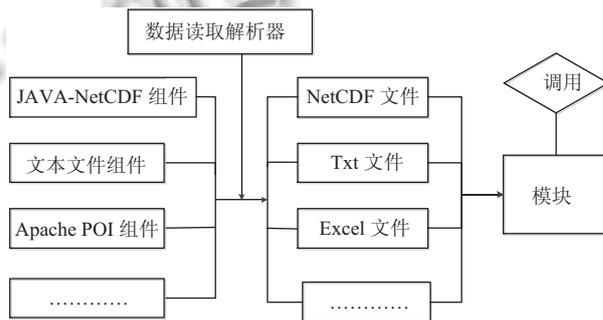


图 4 一体化组件式读取解析设计

3.1.3 数据读取解析设计流程^[9,10]

为解决多种类、多维度的异构海洋科学数据的读取解析问题, 运用面向对象的思想对海洋科学数据文件中的数据进行对象化, 以洋流 NetCDF 数据为例, 设计了包括文件、维度、变量和其类型的读取解析流程, 如图 5 所示.

- 1) 导入 nc 文件, 判断文件是否符合格式要求, 并提取文件后缀名, 进行文件类型判断.
- 2) 创建实体类 NC, 并读取 nc 文件.
- 3) 读取解析维度 Dimensions 和属性 Attributes, 分别得到(D1, D2...Dn), (A1, A2...An)并创建实体类添加到 NC 中.
- 4) 读取解析数据 Data, 获取数据 Data 中属性值总数 count 为维度 Dimensions 总数和其对应的值 Value 数之和. 取其中一个维度, 获取总数 Dcount.
- 5) 判断 Dcount 和 count 的大小, 依次获取 (D₁[0]...D_n[0]), 及值 V[0]数据, 映射等价于 value=f(D₁[0], D₂[0], ..., D_n[0]), 创建实体类并添加到 NC 中.
- 6) 从实体类 NC 中获取全部数据.

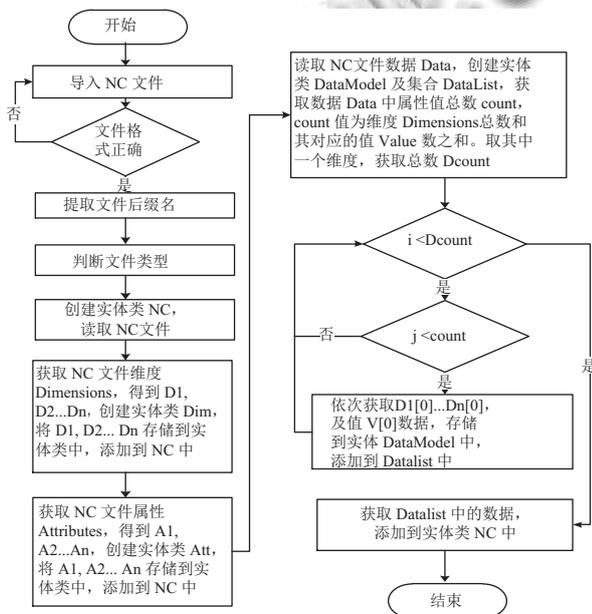


图5 数据读取解析设计流程

3.2 多源异构海洋科学数据的质控机制与方法技术^[11]

3.2.1 海洋科学数据质量控制方法研究

海洋科学数据质量直接影响着海洋事业的科学性, 准确可靠的数据源是海洋科学研究和海洋综合管理的依据. 因此在数据入库之前, 都需要对数据质量进行控制. 在入库前要保证原数据的正确性, 入库后能保证数据前后的一致性.

1) 数据要素合理性检查

根据海洋科学数据的类型及自身的物理特性、变化范围, 对其进行范围检验. 可以根据各科学数据专家

经验, 相关资料文献, 历年来数据波动记录, 历史资料统计值, 确定海洋科学数据的正常取值范围等, 如果超出该范围则认定该数据为异常, 否则为正常数据.

$T_{\text{Min}} \leq T(X) \leq T_{\text{Max}}$, 其中, $T(X)$ 为海洋科学数据的要素值, T_{Min} 和 T_{Max} 为此要素值的合理范围区间内的最小值和最大值.

$TX_{\text{Min}} \leq TX \leq TX_{\text{Max}}$, 其中, TX 为某年中某月的平均要素值 $TX = \sum Xi(t)/N$, TX_{Min} 和 TX_{Max} 为某年中某月要素值的波峰值和波谷值.

2) 数据要素间关系性检验

海洋科学数据中的数据要素之间存在着一定的联系, 并且具有一定的规律性, 利用这些数据的联系性和规律性, 进行检验数据要素间的关系性.

3) 数据要素相符性检验

依照海洋科学数据的标准格式及数据标准, 对满足某一特定属性的海洋科学数据进行相符性检验. 取值范围应严格按照海洋科学数据的标准来进行检验, 且数据值应在固定的取值范围和默认值的范围之内, 否则为数据异常. 另外, 还可以对时间记录和时间标识记录等进行递增量差值的检验, 满足是否为某一确定值.

4) 数据要素相关性检验

海洋科学数据中某些要素数据值的区间范围变化, 通常是不间断且有规律的, 通过确定同一要素相邻时刻之间的差值与此要素的统计参数特定值, 对海洋科学数据进行校验, 超过这一特定值的数据即为异常数据.

$A \geq |Z(t) - Z(t-1)|$, 其中, A 为此要素值的统计参数, $|Z(t) - Z(t-1)|$ 为相邻两时刻的要素值的绝对值.

3.2.2 海洋科学数据实现质量控制设计流程

由于海洋科学数据的数据量庞大, 并且是多源异构的数据, 因此利用计算机的高速度和准确性等特点, 结合海洋科学数据质量控制方法, 设计如下的流程图, 如图6所示.

1) 将一体化读取解析后得到的数据, 进行质量控制.

2) 质量控制分为4个部分, 数据要素合理性检验、数据要素间关系性检验、数据要素相符性检验及数据要素相关性检验, 这4种检验为循序渐进, 并且以资料文献、专家经验和数据特征等信息为参照.

3) 根据质量控制的检验方法, 后台自动加载海洋

科学数据的类型及质量控制方法,对文件数据进行质量控制。

4) A.当数据判断为异常时,系统界面会标识出异常的数据,当点击某一异常数据时,界面给出相应的异常信息提示.用户可对该数据进行人工审查,若审查不通过,则该数据异常,不通过;而对于不可处理及多次质量控制均标识异常的数据,可进行人工审查通过. B.当数据判断为正常时,系统界面上会标识出正常数据。

5) 存储数据库,最终数据全部通过,存储到数据库中。

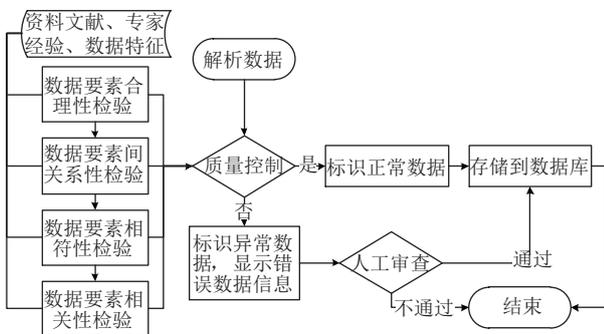


图6 海洋科学数据质量控制流程

3.3 海洋科学数据可视化技术

3.3.1 海洋科学数据可视化特点

海洋科学数据可视化是指通过可视化图像实时动态展示数据动态变化过程^[12]。当今,海洋数据需要处理的数据量越来越大,时空关联性强,数据多源异构,为了能更好的分析利用海洋数据,从海量数据中获取到其中的规律,为相关科研部门提供可视化展示,使其转化为通俗易懂的图形图像,以此达到直观表达展示海洋现象及特征,快速获取有效信息的目的,因而海洋数据可视化的实现是非常必要的。

3.3.2 海洋科学数据可视化流程设计及表达^[13]

针对海洋科学数据可视化特点及业务需求,设计了可视化实现流程,如图7所示.其中对数据的处理主要是完成对海洋专项数据的提取,通过插值、转换、平滑及绘制等操作,将数据转换成图层,实现可视化展示的基础;可视化展示主要是对数据的多维表达及分析结果的展示,支持对海洋科学数据的点、面可视化,分析结果主要是以统计图表、断面图、动态洋流等形式展示。

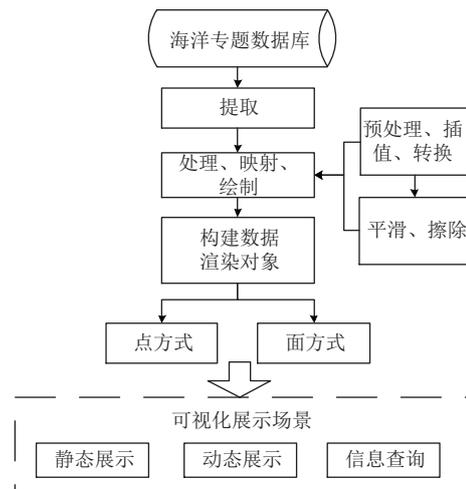


图7 海洋科学数据可视化流程设计

平台根据海洋科学数据的类型和学科要素的属性不同,以地图交互及曲线显示为主,地图交互基于海洋数据的地理位置信息在地图上进行相应展示^[14],包括近海环境参数断面图,近海洋流等;生成曲线的数据来源于数据库中所有可视化海洋要素,包括温度、盐度、及近海观测网络数据对比图等,如图8所示。

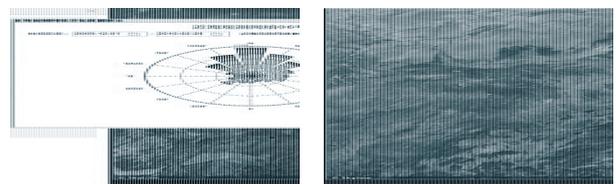


图8 海洋科学数据可视化表达

3.3.3 海洋科学数据可视化设计流程

依据可视化流程设计,以近海环境参数为例,设计如下的流程,如图10所示.近海环境参数包含 pH、溶解氧、叶绿素、CTD、五项营养盐等,且环境参数数据具有横向分布和纵向分布两个特点,可依据各类参数数据特点进行断面图可视化,即在三维空间视图下查看和比较各个参数的分布特点。

1) 从数据库中读取环境参数数据,存储到集合 A,每一行记录包括站位、纬度、经度、日期、时间、水深、水层、参数值。

2) 循环读取集合 A. ①通过选择排序法,找到其参数值的最大值和最小值. ②依据水层深度划分行数据,

$$f(x) = \begin{cases} 0 & 0 \leq x \leq 3 \\ x & x-2 < x \leq x-3 \text{ 且 } x \% 5 = 0 \end{cases} \quad (1)$$

其中, x 代表层数, 每 5 m 一层, 划分到 70 m 层, 其中 0-3 m 为 0 m 层数据, 大于 3 m 且小于等于 8 m 为 5 m 层数据, 以此类推, x-2 至 x+3 之间的数据均划为 x 层数据, 以增加 x 层的数据量, 划分结果存储到哈希表中。

3) 循环处理各个水层数据. ①根据水层数据创建要素类 SHP 文件, 将数据集合写入到 Feature 中, 判断插值字段, 确保插值字段的数据均为可用数据. ②运用克里金插值算法对散点采样值估算相应区域, 从而获得这个区域的连续表面. 插值方差, 就是指实际参数值 z_v 与估计值 z_v^* 两者偏差平方的数学期望:

$$\sigma_E^2 = E[(z_v - z_v^*)^2] \quad (2)$$

而插值点的 z_v^* , 通过 N 个离散点获得:

$$z_v^* = \sum_{i=1}^n \lambda_i Z(X_i) \quad (3)$$

对参数值采用球形模式, 球形模型公式:

$$\gamma(h) = \begin{cases} 0 & h = 0 \\ C_0 + C(\frac{3}{2} \times \frac{h}{a} - \frac{1}{2} \times \frac{h^3}{a^3}) & 0 < h \leq a \\ C_0 + C & h > a \end{cases} \quad (4)$$

对水层的点要素图层进行插值. ③对插值后的栅格数据集进行重分类, 根据参数的最大值和最小值获得各个的参数的数值范围, 以此为基础计算重分类的方式. ④将插值后的栅格数据集转换为矢量图层。

4) 等值面, 等值线的后期平滑处理. 由栅格数据得到的矢量数据, 通常边界会出现很多方格子或者锯齿, 因此需要进行平滑处理^[15]。

设计思路: 平滑处理分成两部分, 一是将面要素转换成线要素(FeatureToLine), 二是将面要素转换成点要素(FeatureToPoint)。

设计流程: ①等值面经过面要素转换成线要素(FeatureToLine)得到等值线. ②等值线进行平滑线处理(SmoothLine), 采用 PAEK 算法, 得到新的等值线_\$. ③新的等值线_ \$转换成面要素(FeatureToPolygon), 得到新的等值面_ \$. ④等值面经过面要素转换成点要素(FeatureToPoint), 并得到等值面中心点. ⑤平滑处理的最后一步, 将等值面_ \$与等值面中心点进行空间连接(SpatialJoin), 生成最终的等值面_Final. 如图 9 所示。

5) 擦除, 将大于 0 米的水层(低于水层深度的部分)擦除. 依次循环操作, 最终生成各个水层的断面图。

6) 利用 TerraExplorer 将各个水层深度的图层加载到 FLY 工程文件上, 并根据 Skyline 客户端, 呈现

在地图上。

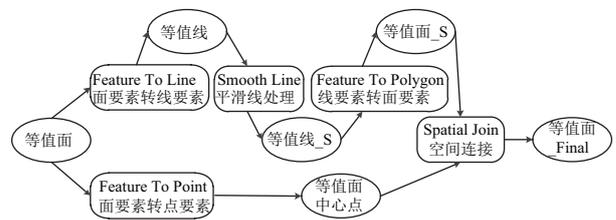


图 9 平滑处理流程

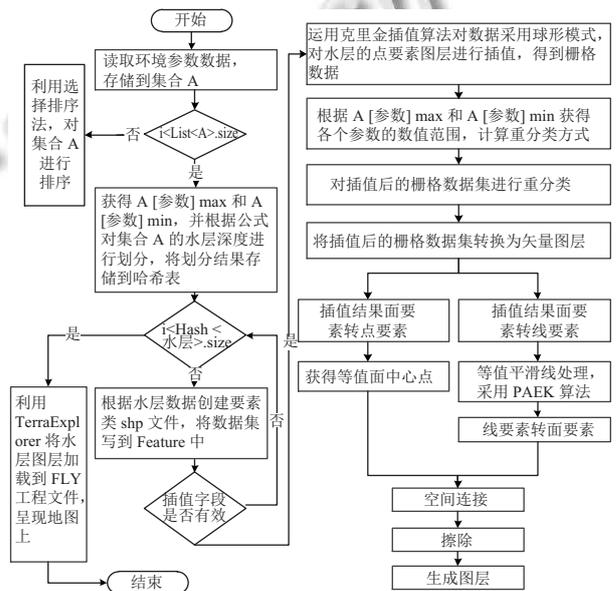


图 10 海洋科学数据可视化流程

4 结语

通过计算机技术和海洋科学数据业务的结合, 能够高效的对海洋数据进行管理, 使海洋数据管理工作更加全面专业, 同时由于海洋科学数据格式的多源异构, 对数据的读取解析、质量控制也成为提高业务效率的关键。

海洋科学数据管理与可视化系统, 将海洋科学数据的常见格式类型进行读取解析集成, 实现对多种格式的数据解析与质控, 并基于 WebGIS 技术实现对海洋科学数据的管理、分析和可视化, 对提高海洋数据查询效率、改善数据展示友好性具有一定意义, 为海洋相关部门管理决策提供科学的支持。

参考文献

1 阎保平, 桂文庄, 罗泽. 我国科学研究信息化的发展与启示. 科研信息化技术与应用, 2010, (1): 10-19.

- 2 中国科学院科学数据库资源整合与持续发展研究报告写作组. 中国科学院科学数据库资源整合与持续发展. 第九届科学数据库与信息技术学术讨论会论文集. 桂林, 中国, 2008.
- 3 徐超, 李莎, 王东晓, 等. 南海海洋科学数据库数据共享与服务系统的设计与实现. 科研信息化技术与应用, 2011, 2(1): 10-17.
- 4 刘振民. 基于 WEBGIS 的海洋环境信息共享研究[硕士学位论文]. 青岛: 中国海洋大学, 2005.
- 5 梁吉欣, 左小清. Skyline 在 Web 三维 GIS 中的应用研究. 昆明理工大学学报(理工版), 2009, 34(2): 1-4.
- 6 胡英杰, 吴健平. Skyline 和 ExtJS 的企业级三维 GIS 设计. 测绘科学, 2010, 35(6): 247-249.
- 7 涂建光, 边馥苓. 基于设计模式的组件化 GIS 软件开发方法. 武汉大学学报·信息科学版, 2005, 30(1): 77-81.
- 8 梁庆发, 梁志宇, 郑国珊. 基于 GIS 的多源异构数据整合方法研究. 测绘与空间地理信息, 2015, 38(6): 144-146.
- 9 王想红, 刘纪平, 徐胜华, 等. 基于 NetCDF 数据模型的海洋环境数据三维可视化研究. 测绘科学, 2013, 38(2): 59-61.
- 10 马伟霞, 田丰林, 纪鹏波, 等. 海洋多源异构数据转换系统的设计与实现. 计算机工程与设计, 2014, 35(8): 2917-2922.
- 11 温玉波. 海洋环境观测数据的质量控制研究. 农业网络信息, 2014, (2): 35-38.
- 12 孔倩倩, 韩勇, 李文庆, 等. 海洋标量数据多维多模式动态可视化系统设计实现. 微计算机信息, 2011, (5): 177-179.
- 13 宋丽丽, 康林冲, 王漪, 等. 基于 B/S 的海洋环境数据可视化与服务. 海洋科学, 2016, 40(7): 124-131. [doi: 10.11759/hyxx20150624001]
- 14 孟娟. 海洋数据平台数据可视化查询与展示子系统设计与实现[硕士学位论文]. 青岛: 中国海洋大学, 2013.
- 15 陈星亮. GIS 在海洋环境评价预警中的应用与研究[硕士学位论文]. 青岛: 青岛科技大学, 2016.