

生物标本地理信息系统 SpecimenMapping 的设计和实现

张尚武 纪力强*

(中国科学院动物研究所 北京 100080)

摘要 针对生物多样性研究领域的需求与现有 GIS 平台的局限,运用 GML 和 SVG 等相关技术设计实现了一个基于 XML 标本地理信息系统——SpecimenMapping。该系统具有导入常见空间数据格式、投影坐标互相转化、管理个人标本数据、导入其他格式的标本数据、图形化显示标本地理信息、分布地生态因子分析和栖息地预测等功能,具有基于开放标准便于功能扩展、方便操作、针对性强等特点。

关键词:可扩展标记语言(XML);地理标记语言(GML);可伸缩矢量图形(SVG);标本;地理信息系统;SpecimenMapping

中图分类号:Q958,Q141 文献标识码:A 文章编号:0250-3263(2006)04-64-07

SpecimenMapping : Design and Implementation of a Specimen-geographical Information System

ZHANG Shang-Wu JI Li-Qiang

(*Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, China*)

Abstract Due to the requirement of the biodiversity research and the limitation of current GIS tools, we designed and implemented a Specimen-geographical Information System based on XML technology with GML and SVG technology, SpecimenMapping, which provides some functions for biological specimen data management, such as importing other format spatial data, converting coordinates from one kind of projection to another, editing the personal specimen information graphically, displaying distribution maps of specimen, and analyzing the ecological factors of the habitat, etc. This system is designed specially for biodiversity researchers, so it's more friendly and easy to use. It is more important that it is based on open standards and more scalable for further development.

Key words: XML; GML; SVG; Specimen; Geographical Information System; SpecimenMapping

随着学科信息化建设的深入,生物标本信息逐步被数字化,一个囊括全国各生物标本馆信息的数据库正在紧张的建设之中。研究如何深入挖掘这些标本信息,尤其是标本的采集地信息,并结合 GIS 的相关技术做更进一步分析,对于物种保护等生物多样性研究显得尤为迫切。然而现有的 GIS 软件功能复杂价格昂贵,更适合具备一定 GIS 专业知识的研究人员使用,这使得生物多样性研究的人力物力成本大幅增加。作为发展中国家的中国在生物多样性

研究领域的投入有限,使得 GIS 技术在我国生物多样性研究领域上的应用滞后于国际同行^[1,2]。随着计算机技术尤其是网络技术和 GIS 技术等相关研究的深入,基于 XML 技术的新兴的图形图像技术标准 SVG 和空间信息的

基金项目 国家科技基础条件平台工作重点项目(No. 2003DEA3N023, 2004DKA30520);

* 通讯作者, E-mail: jli@ioz.ac.cn;

第一作者介绍 张尚武,男,硕士研究生。

收稿日期 2005-12-26,修回日期 2006-04-16

交换格式标准 GML 相继推出,许多 GIS 技术的研究者提出以上述两种标准设计开发基于 XML 的 WebGIS 系统^[3-5]。笔者参与开发的“动物标本信息系统”的 WebGIS 部分正是基于这两种技术实现的。为了进一步支持统计分析等更复杂的功能,笔者设计了具有 GIS 功能的桌面应用程序生物标本地理信息系统 SpecimenMapping 尝试为生物多样性研究者提供一个生物标本信息的管理分析工具。

1 相关技术简介

1.1 可扩展标记语言(Extensible Markup Language, XML) 是 W3C 组织于 1998 年制定的一种通用语言规范,提供一种灵活开放的描述结构化数据的方法。它使用一组标记来描绘数据元素,而每个元素又可封装或简单或复杂的数据。目前,由于具有与平台无关和简单开放等特性,XML 已经被广泛的应用在互联网数据发布和异构数据共享等多个领域^[6]。

1.2 地理标记语言(Geography Markup Language, GML) 是由开放地理信息系统组织 OGC(Open geospatial consortium)开发的基于 XML 的地理信息编码工具,用来表示空间对象的空间数据和非空间的属性数据。它通过提供基本的几何标记、通用的数据模型 和一个创建与共享应用系统框架的机制来实现空间数据的互操作。2000 年 4 月 OGC 推出了基于 DTD(Document Type Definition)的 GML1.0 版本,2001 年 2 月推出了基于 XML Schema(简记为 XSD)的 GML2.0 版本,2003 年 2 月又发布最新 GML3.0 版本。对于实现空间信息的共享和异构空间信息进行的集成,GML 是一个有力的工具。

由于 GML 是基于 XML 的,所以与 XML 编码相似,GML 也是用文本的形式进行地理信息的表示。它建立在 OGC 地理抽象模型的基础之上,用地理特征(features)来描述世界。这里的特征是属性与几何体的序列,属性用名称、类型、属性值来描述,几何体由基本的几何体类型如:点、线、多边形、多点、多线、多多边形等所组成^[7]。

1.3 可伸缩矢量图形(Scalable Vector Graphics, SVG) 是一种基于 XML 的、描述二维矢量图形和矢量/点阵混合图形的置标语言,是一种全新的矢量图形规范。SVG 规范定义了 SVG 的特征、语法和显示效果,包括模块化的 XML 命名空间(namespace)和 SVG 文档对象模型(DOM)。SVG 的绘图可以通过动态和交互式方式进行,在实际操作中,则是以嵌入方式或脚本方式来实现的。SVG 不仅提供超链接功能,还定义了丰富的事件。由于 SVG 支持脚本语言(script),可以通过 Script 编程,访问 SVG DOM 的元素和属性,即可响应特定的事件,从而提高了 SVG 的动态和交互性能。SVG 实现了图形、图像和文字的有机统一。SVG 除了支持 HTML 中常用的标记如文本、图像、链接、交互性、CSS(Cascading style sheets)的使用、脚本外,还提供了大量针对图形、图像、动画的特定标记。这就为实现 GIS 提供了可能^[8]。

2 SpecimenMapping 的设计与功能

2.1 SpecimenMapping 的设计 在软件需求调研中,笔者发现生物多样性研究者需要一个既可以方便访问网络化标本数据库中的数据,又可以管理个人本地标本数据的管理工具,并能通过该工具对这些数据方便地进行比较、生境分析和相关统计。为实现这些功能,笔者设计了如图 1 所示的软件结构。

为了使用现有的地理数据,系统设计了一个空间数据导入和转化组件,用来把常见地理数据格式(如 Arc/Info 的 Shape、PC Coverage 和 ASCII GRID)的空间数据导入到系统之中。

鉴于目前已有网上的标本数据库服务器提供标本信息查询服务,系统也应具有从网络上的标本信息系统获得标本数据的功能。SpecimenMapping 的数据访问客户端组件通过与网络服务器上的“数据访问服务器端组件”通信,完成对网络标本数据的查询。考虑到用户可能会有自己的标本或调查数据,所以这些本地数据的导入,以及对这些数据的编辑管理还需要一个“标本数据管理组件”支持。

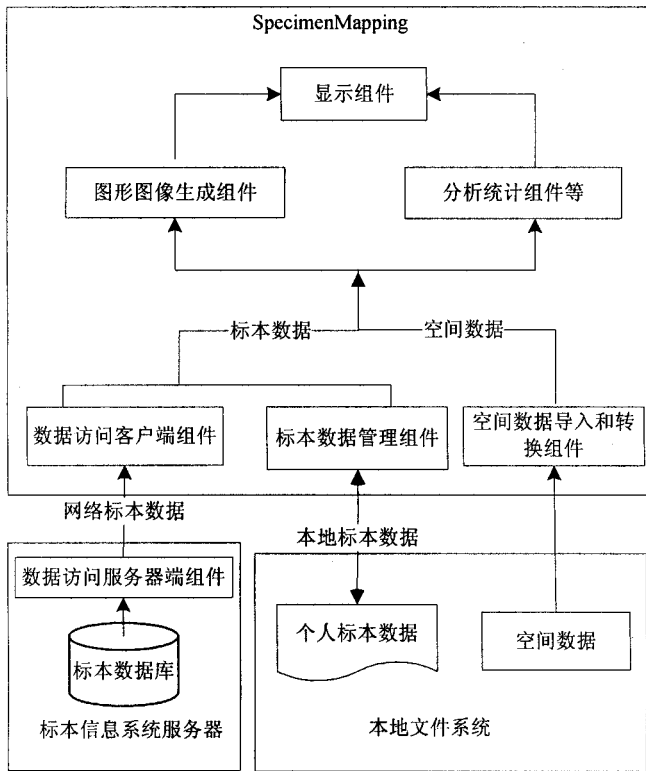


图 1 SpecimenMapping 结构示意图

这些标本数据和空间数据在“图形图像生成组件”支持下生成相应的分布图形或图像,或由生境分析、栖息地预测等分析统计组件经过针对某一生态因子图层进行分析生成统计图表,或针对多个生态因子图层进行叠加处理形成新的分析指数图层,在“显示组件”里显示。

2.2 SpecimenMapping 的功能实现

2.2.1 空间数据的导入与转换

Specimen Mapping 使用的矢量空间数据为自定义的 XML 文件格式,其结构是笔者参照 OGC 的 GML 架构设计的^[9]。作者根据美国环境系统研究所 (Environmental Systems Research Institute, ESRI) 1998 年 7 月发布的 Shapefile 技术白皮书等相关资料设计实现了 Shape、PC Coverage 文件的读取转化模块,用来实现常见空间数据的导入。在导入空间数据的时候,SpecimenMapping 请用户确定一个地物属性作为对地物进行分类的依据,然后对所有要导入的地物进行分类。比如在导入植被图的时候,如果指定“植被类型”作

为分类属性,则在显示和统计的时候会根据“植被类型”的属性值分别着色和统计,其导入过程如图 2 所示。

由于空间数据可能存在投影坐标系统的差异,所以要使不同投影坐标下空间数据叠加显示,就必须把它们投影在同一个坐标系统内。SpecimenMapping 根据实际需要,实现了同一基准面内的“正轴等积圆锥割投影”、“正轴等角圆锥割投影”、“高斯投影”、“UTM 投影”和“墨卡托投影”之间的相互转换。如果源数据未提供投影系统的信息,则需要用户手工指定投影坐标系统。投影变换采用的是综合变换,即先求出原投影点的地理坐标,然后根据该地理坐标求出新投影坐标系统下该点的直角坐标。投影算法参考 European Petroleum Survey Group 所著“Coordinate Conversions and Transformations Including Formulas”一文和杨启和所著《地图投影变换原理与方法》书中的相关章节^[10]。

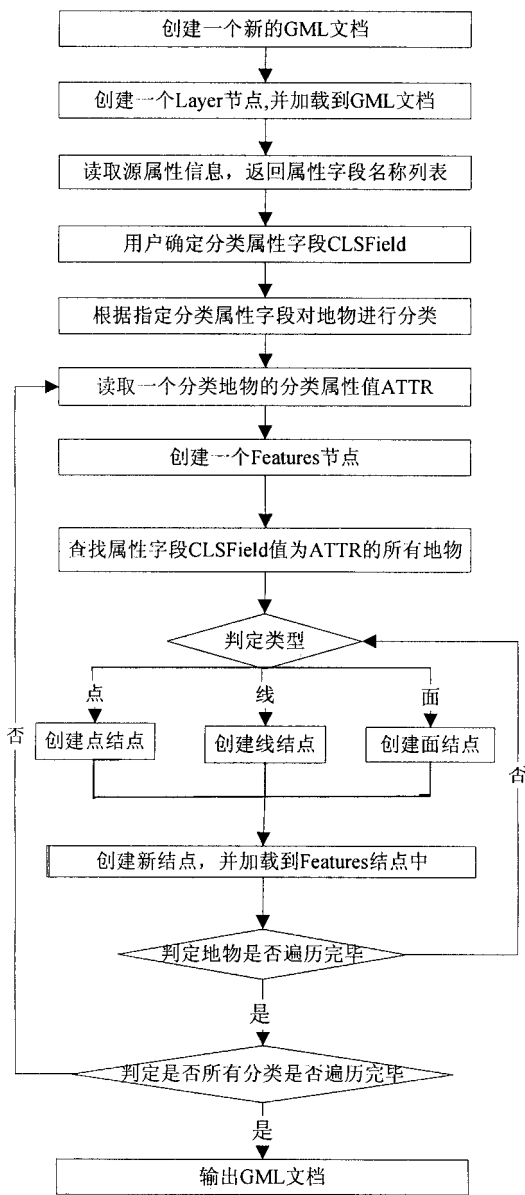


图 2 SpecimenMapping 空间数据的导入

2.2.2 网络标本数据的访问 中国科学院动物研究所生物多样性信息管理研究组开发的标本管理信息系统服务器端装有标本信息查询 WEB 服务,允许用户注册后经数据拥有者授权后读取相应的标本信息。本文笔者结合该 WEB 服务在 SpecimenMapping 中设计了注册组件和网络数据访问组件,使用户可以方便的注册并通过相应接口访问标本管理系统的标本信息,包括采集地经纬度、采集时间、采集人、鉴定

物种等。

为了满足用户的实际需要,SpecimenMapping 提供用户从服务器上下载已注册信息的功能,避免用户因为重新安装 SpecimenMapping 或在别的计算机上安装时需要重复注册所带来的麻烦。

2.2.3 标本数据的编辑与转换 为了方便用户使用,SpecimenMapping 在标本数据的编辑窗口设计实现了表格编辑和可视化图形编辑两种方式,用户既可以在二维表格里直接编辑标本的分布信息(经纬度坐标)和其他相关信息,又可以以编辑图形的方式先确定标本在地图上的位置,再编辑其他相关信息。

因为技术或标准不一致等各种原因,标本信息的记录存在各种差异。比如对标本采集地的记录,在较早的时候都通过记录行政区名称的方式,而在近年来多采用记录经纬度的方式,这给标本的在地图上标记带来了困难。为了解决这一问题,SpecimenMapping 中设置了“标本数据导入向导”功能,引导用户把常用的 Excel 表中的标本数据导入到 SpecimenMapping 中。此外,作者还定义了一个 Excel 模板,用户只需要把自己的标本数据依次填写到格式化的 Excel 表格里,就可方便地导入到 SpecimenMapping 中。

考虑到不同作者对标本信息各字段可能采用不同的名称,在导入数据的时候需要用户自己将源文件中的数据项与 SpecimenMapping 标本数据信息文件的数据列一一对应;另外,由于考虑到行政区划的历史沿革,用户还需要对源文件中的老地名和 SpecimenMapping 所采用的新地名进行匹配。对于没有记录标本采集地经纬度的标本,SpecimenMapping 采用取其记录的行政区中心点坐标方法来标记其位置。该中心点是作者根据“中国县区图”(中国科学院地理研究所资源与环境信息系统国家重点实验室 1996 年 6 月数字化版)各个县区多边形对象的属性“centerid”或“labelid”求得,然后再一一匹配给现有新地名列表中的各个区县。

在背景图层支持下,访问网络数据结合本地数据的显示效果如图 3 所示。

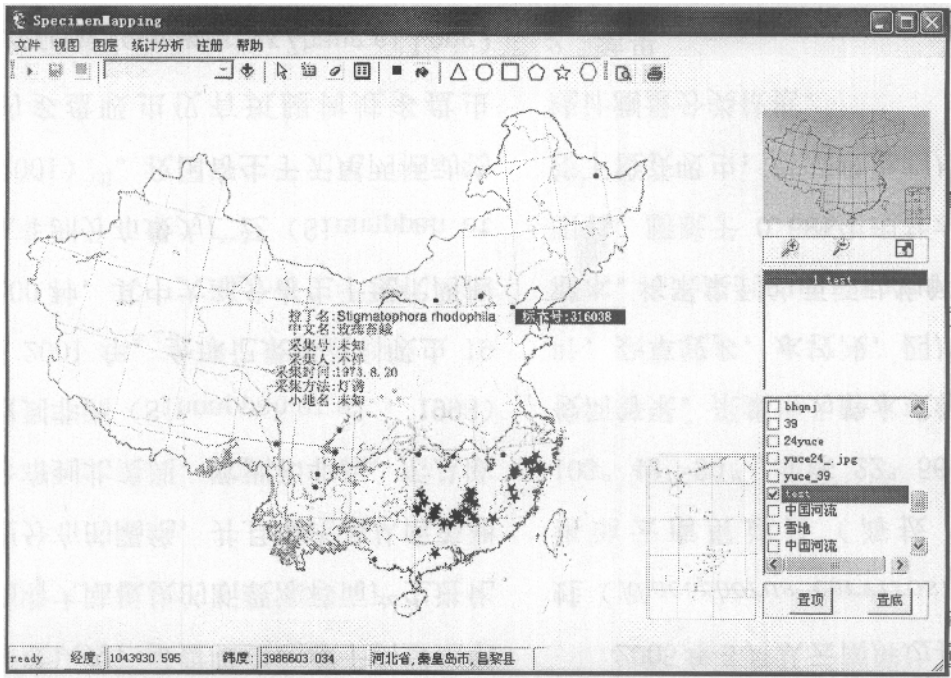


图3 SpecimenMapping 网上与本地数据浏览效果图

红色圆点为网上数据, ★为本地数据, 背景图层是海拔 1 000 ~ 1 600 m, 年均温度 10 ~ 20℃ 的森林斑块和省界、一级河流图。

2.2.4 采集地生境分析和栖息地预测

SpecimenMapping 中的采集地生境分析模块可以查询某一具体标本采集地的生态因子, 例如植被类型、海拔高度、年均温度、年均降水等, 并对其分别做简单的统计分析, 生成统计图表供用户参考。对已知生境条件的物种, 还可以通过设定上述各生态因子条件来预测其在国内的潜在分布。图 4 显示了测试数据的生境因子分析结果, 图 5 显示了栖息地预测的效果。

2.3 空间栅格数据的预处理

由于 SpecimenMapping 还不支持对空间栅格数据的坐标转换, 在 SpecimenMapping 中使用的栅格数据需要进行预处理, 使它们位于同一地理参考系下和同一覆盖范围内。目前 SpecimenMapping 的数据使用的大地坐标系是北京 1 954 坐标系, 投影方式是等积圆锥投影, 投影参数依次是第一标准纬线北纬 25°, 第二标准纬线北纬 47°, 中央经线东经 110°, 覆盖范围的四至依次是最高纬度北纬 54.15°, 最低纬度北纬 14.22°, 最大经度东经 136.42°, 最小经度东经 68.07°, 包括除南海诸岛的中国全境。

目前 SpecimenMapping 使用了 6 个图层的地理数据, 分别是中国科学院地理研究所资源与环境信息系统国家重点实验室 1996 年 6 月数字化的中国植被图和中国地形图, 中国生态系统研究网络 (CERN) 2004 年 12 月出版的《中国陆生系统空间信息图集——气候要素卷》中的年均气温、年均降雨和年均相对湿度 3 个图层^[1], 中国林业科学院根据国家林业局第五次全国森林资源普查所做的全国森林资源分布图。

这 6 个图层的数据需要经过预处理才能导入系统使用。其中, 中国植被图和中国森林分布图转成 1 km × 1 km 的栅格数据, 同时等高线图也被转成栅格数据。由于气候数据中的年均降水和年均相对湿度两个图层的投影中央经线为东经 105°, 所以需要投影变换。所有地理数据都需裁切, 使其都位于 2.3 节所述的经纬度范围内。最后, 这些数据以一个统一的格式输入到 SpecimenMapping 中。

3 结束语

运用 XML 技术, 既满足了 SpecimenMapping

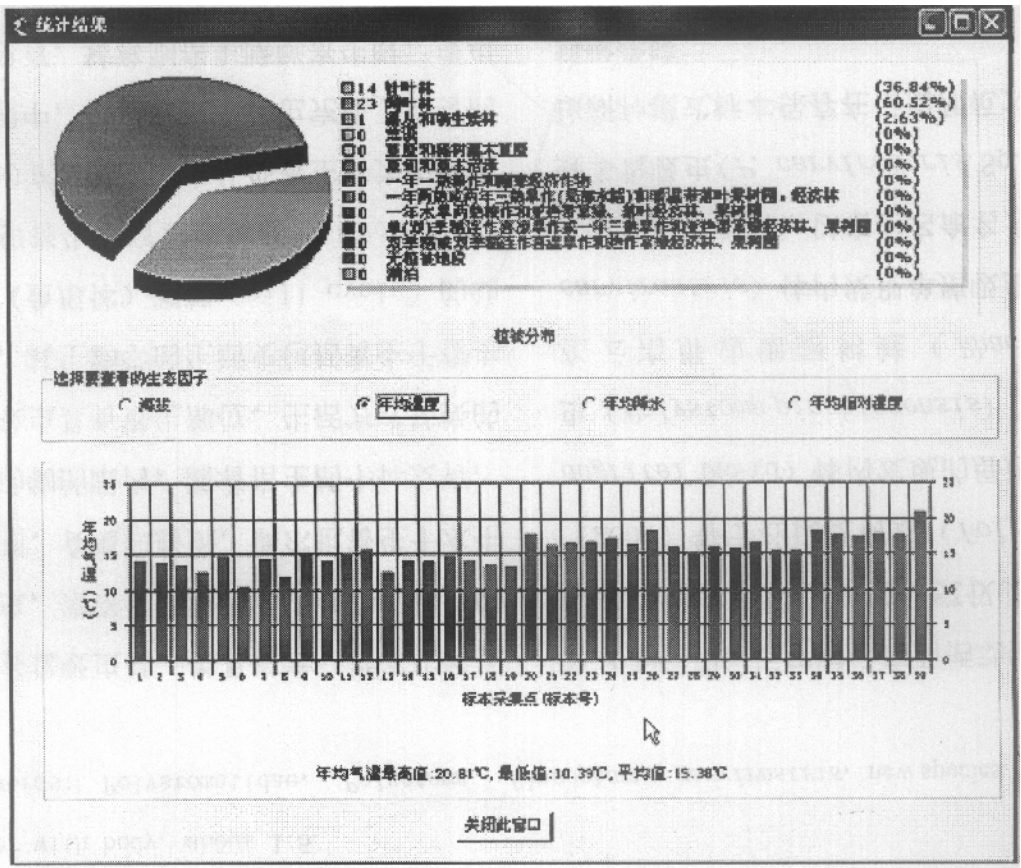


图 4 SpecimenMapping 生境因子获取结果

饼状统计图显示各标本采集点所在植被类型的百分比, 条状统计图显示各模拟标本采集点的生境因子数值。

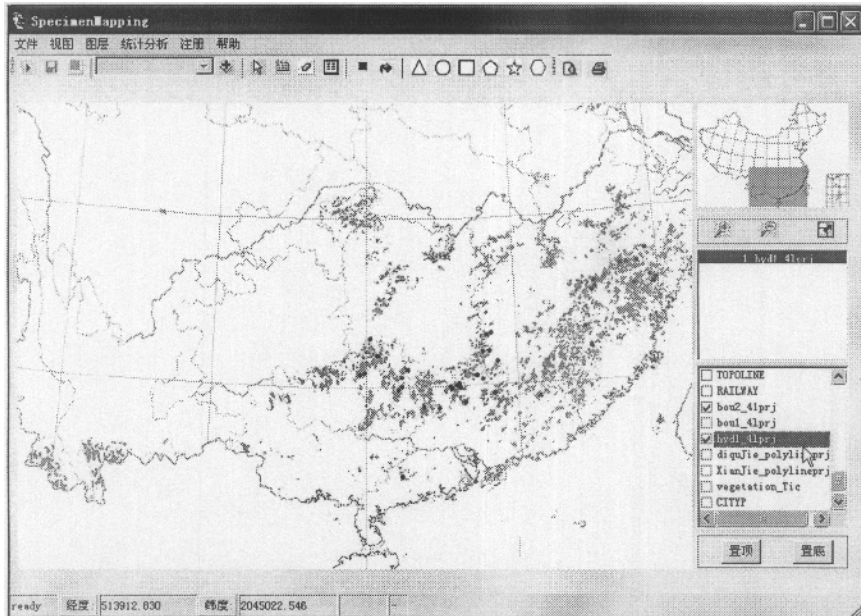


图 5 SpecimenMapping 栖息地预测测试结果
绿色斑块为潜在栖息地, 红色圆点为模拟标本采集地。

对空间数据存储、转化以及分析的需求,又使其具有与其他系统共享资源提高空间数据互操作性的能力。而 GML 和 SVG 技术的使用使得地图的显示非常友好,实现与用户的良好互动。

SpecimenMapping 还存在诸多不足之处:

1) SVG 显示需要第三方插件支持。它在带来便利的同时也对 SpecimenMapping 的性能有所影响。

2) 统计分析,如生境分析和栖息地预测的功能还有待进一步加强和完善。

3) 目前版本的 SpecimenMapping 支持 GML2.0 不支持拓扑关系的描述。最新版本的 GML3.0 对地理信息的定义更加详尽, SpecimenMapping 在今后的开发中会考虑兼容 GML3.0 格式,以利于以后开发更复杂的空间分析功能。

致谢 衷心感谢中国生态系统研究网络综合中心提供的国家尺度的 $1\text{ km} \times 1\text{ km}$ 网格的气候要素数据和中国林科院信息和资源研究所提供的全国森林资源分布图。本文在撰写过程中,中国科学院动物研究所鸟类生态组的孙悦华老师给予诸多指导,在此表示深深的谢意。

参 考 文 献

- [1] 谢焱. GIS 技术在保护区管理中的应用. 见: 中国环境与发展国际合作委员会编. 保护中国的生物多样性(二). 北京: 中国环境科学出版社, 2001, 227~233.
- [2] 周立志, 李迪强, 马勇等. 地理信息系统 GIS 在动物多样性研究中的应用. 动物学杂志, 1999, 34(5): 52.
- [3] 博大勇, 王申康. 基于 GML&SVG 的城市地理信息系统的设计和开发. 计算机应用研究, 2003(7): 113~118.
- [4] 张琦, 陈玉敏. 基于 GML 和 SVG 实现 WebGIS. 测绘信息与工程, 2005, 30(4): 42~44.
- [5] 董伟杰, 史杏荣. 基于 XML 的图形技术在 WebGIS 中的应用. 计算机工程, 2001, 27(7): 174~176.
- [6] 张晓晖, 王艳斌, 赵伟明. XML 手册(第四版). 北京: 电子工业出版社, 2003, 1~2.
- [7] 梁明, 鲍艳, 黄朝华. GML—地理标记语言特征与分析. 西安科技学院学报, 2002, 22(4): 429~433.
- [8] 王兴玲. SVG 与矢量地图的 Web 发布技术. 计算机工程与应用, 2002, 38(10): 1~4.
- [9] 尹章才, 李霖, 朱海红. 基于 XML 的地理信息图示表达模型研究. 遥感信息, 2005(4): 53~55.
- [10] 杨启和. 地图投影变换原理与方法. 北京: 解放军出版社, 1990, 57~59.
- [11] 于贵瑞, 何洪林, 刘新安. 中国陆地生态系统空间化信息研究图集气候要素分卷. 北京: 气象出版社, 2004, 54.