# 绪论

## 研究背景与意义

随着科技的迅猛发展，如今信息数据已经成为人们必不可缺的重要资源。如今计算机的飞速发展与普及，更是加快了新兴的计算机行业与其他传统行业的杂交融合，进而产生了许多新的杂交学科，其中的GIS正是计算机与传统地理行业融合，进而产生的一门学科。

地理信息系统（Geopraphic Information System，GIS）是以采集、存储、管理、显示和分析地球表面与空间、地理分布有关的数据的综合计算机信息系统，是一种分析和处理海量空间数据的技术[]。

地理信息系统作为一门新兴杂交学科在地理信息方面有着不可忽视的作用，特别是在土地管理、灾害预测、环境保护等方面，具有广阔的应用前景。借助于地理信息系统的帮助，气象方面的数据处理和预测有了较多可行且有效的方法，GIS中有很多相关于图形学和地理学的知识，通过数学方程组，计算地理数据、预报数据等，进行气象资料的搜集与计算，从而得出气象GIS数据，最后进行气象GIS信息图像的制作[]。

现在，气象数据已不再是仅在气象局内部使用的机密数据，中国气象数据网与美国的NASA，NOAA等都将国内，甚至世界各个站点手机到的气象数据进行共享，开放了气象数据的商业研究和使用。这为国内工作者提供了较好的数据支撑。而且气象服务并非只能用在科学研究上，特别是在国外，气象服务商业化十分普遍，英国气象服务年产值高达2600亿美元，美国1600亿，日本100亿，而然中国只有6亿美元[]，由此可以看出国内对气象商业化方面发展较为缓慢，对于从事气象方面研究的工作者来看，利用市场的空虚，致力于发展气象服务，在中国将有广阔的发展空间。

目前的气象数据分析，还是需要依赖相关的GIS技术，而GIS中的一些分析显示功能主要是通过一些软件来进行实现。就现在的软件而言，近几年来，不少开源软件应为其相比商业软件有着降低风险、产品质量可靠、更好的支持、有利的许可等优点，在桌面、Web、移动等诸多领域得到了飞速发展和应用[][]。同样在GIS领域中，开源软件也开始发挥着越来越重要的作用[]，从传统的桌面GIS，如ArcGIS，到现在的WebGIS，移动GIS都出现了很多具有代表性的开源软件[][]。比如现在，Google、Yahoo、腾讯、百度等互联网公司退出许多关于GIS服务并对外开放的调用接口[]，有关开源GIS软件应用和开发方面的研究和文献数量正在不断增加。特别是在WebGIS的发展上越来越开放化，面向云端和大数据的WebGIS技术十分让人重视，其原因是利用云计算处理大量的地理数据，采用分布式计算以保证数据的高效性与实时性。但是，虽然地理信息系统和地图技术从Google地图，bing地图和Here地图在WebGIS上取得了非常大的成功，但是这些地理信息系统平台不能满足一些业务应用程序的需求，如成本和离线映射[]。特别是对于气象数据，气象数据中，如大气数据变化可能产生的大气波导现象已经在现在的军事战争中，有着十分重要的作用，所以在WebGIS并不完全成熟的情况下，其他开源的GIS软件十分有研究价值。

DotSpatial是从MapWinGIS开源项目上发展起来的，由MapWindow GIS开源团队以及OSGeo .Net开发者社区负责共同开发[]。DotSpatial开源软件旨在于提供一套免费的、稳定的、开源和独立的类库，能够在.Net、Silverlight和Mono平台上使用[]。DotSpatial已经将其源代码放在了github上，以便开发者可以根据研究源代码深入探讨其底层的实现原理，并在上面做出自己的改进以满足自己的实际需求。同时DotSpatial也提供给开发者原来的类库和插件，如果需要用类似的功能，在开发平台上可以直接调用。Dotspatial使开发人员不借助商业的GIS软件（如借助ArcGIS的arcengine进行二次开发），能够实现基本的GIS操作，如地图显示，数据读取，平移缩放。

## 国内外研究现状

### DotSpatial研究

目前国内外对于DotSpatial这个开源工具，研究气象数据的相关文献还比较少，从2010年到现在出现了仅9年时间，而且作为商用的WebGIS的出现极大地冲击了DotSpatial的研究市场。而且，DotSpatial就目前来说只能实现较为基本的GIS操作，其能够支持的数据格式、空间分析算法、渲染效果和种类、用户帮助文档信息、开发管理手册以及用户数量方面都与着成熟的ArcGIS这样的商业软件有着一定差距。此外DotSpatial是MapWindow的一个子项目，对于DotSpatial的研究一定层度上是基于MapWindow的研究，所以下面将列举一些关于MapWindow以及DotSpatial的研究现状。

国外方面：

Carlos Mario Zapata，Francisco Mauricio Toro等对MapWindow于ArcGIS从进行了多方面的对比[]；Carlos Osoriol，Matthew Over等在分布参数估计的相关部分使用了MapWindow Desktop[]；C. Georgel，L.F.Leon介绍了基于MapWindow Desktop构建的WaterBase项目[]；Rocha JG，Grueau C介绍了在DotSpatial下开发和集成离线和开源替代映射的解决方案工作[]；Hentati I，Trabelsi N，Trili I等在突尼斯莫纳斯蒂尔市地下水管理水文地质信息系统的设计与实现中使用了DotSpatial的工具[]。

国内方面：

方俊介绍了基于DotSpatial实现的地图服务功能[]；马云强，杜婷，毕猛等人对基于DotSpatial的轻量级GIS开发技术进行了研究[]；郭庆胜，马潇雅，王琳等人在基于插件技术的地理信息时空分布与变化特征提取系统的设计与实现中使用了DotSpatial的插件技术[]；肖建军，袁满利用Mapwindow Desktop对大气污染扩散进行了相关分析与研究[]；张弟使用Dotspatial实现了对NetCDF类型文件的跨平台读取以及地图漫游[]；北京灾害预防机构地理信息系统协会的Ye Jiamin，Zhao Hua，Liu Shuai等人使用DotSpatial开源GIS技术进行了地震预测分析[]；同样的，河北灾害预防机构地理信息系统协会的Huang Meng，Pei Biao，Zhang Dian等人也在地震灾难的预测中使用了DotSpatial工具[]；

### GIS对气象数据的研究

对于气象方面的研究，不再只是将其进行单纯的物理研究，而是将其作为地理信息数据中的一部分，使用GIS的分析技术，将其进行空间分析。

国外方面：

美国大气研究中心（National Center for Atmospheric Research，NCAR）将GIS与大气科学相结合，在2001年组织成立由相关领域工作人员构成的GIS研究团队，在2002年和2005年召开了有关GIS如何应用在气象方面的研讨会[]。为了将GIS技术广泛用于气象领域，在2001年，欧洲协作研究开始了COST719计划，由18个欧洲国家参与研究，其目标为：在GIS方面密切合作，建设GIS与气象数据的接口，评估气象数据的可获得性与实用性，鼓励GIS技术应用在气象研究中，紧密联系国家气象服务机构、研究组织和GIS行业[]。

美国国家天气服务中心（NWS）将天气服务与GIS技术结合，在网络上发布天气信息[]。印度中尺度预报国家中心将GIS技术中的可视化技术与空间分析结合中尺度预报模式。Daly使用GIS技术集成DEM模型，显著提高了气候分析和描述能力[]。

同时GIS技术与气象方向结合的趋势使得很多高校投入了大量的科研力量在此方面，并研究出了多款软件。如Winsconsin大学空间科学工程中心研究的VIS5D，是三维气象工具的代表，由三维绘图API开发的可视化软件；马里兰大学气象系研制的GrADS可实现对气象数据的选择性读取、处理、分析、显示，并直接绘图，支持站点数据资料，此系统方便的数据操作交互功能，广泛用于气象科研领域。

由此可以看出，国外将气象与GIS的结合已经有了很长的时间，而且对其非常重视也相对成熟，但是一些软件在气象方面的专业性太强，而且并没有充分用到GIS技术的部分功能，所以，GIS在气象领域的应用还需要继续深入。

国内方面：

我国由1999年7月引进气象站设备，2000年1月我国自主生产的自动气象站产品问世，随后我国的气象站逐渐遍布全国，空间分辨率到达平均10km。为了将GIS技术与气象结合起来，解决气象数据的管理工作，中国气象局与国家测绘局在2007年达成一致，密切将地理气象信息数据资料共享合作，自此我国气象方面迈向新时达。

潘雪萍将气象的温度数据使用不同的插值方法，研究插值方法的优劣[]；罗红磊，黄小燕，苏志等人通过GIS技术分析广西温度参数的空间分布特征[]；甘泉通过GIS软件对气象信息进行集成与可视化分析[]；吴焕萍，罗兵，王维国等人通过GIS的气象要素插值，空间信息分析，三维可视化技术等对GIS在气象中的作用进行评估与判断[]；李江南比对GIS，MapInfo和MapBasic三者的关系，并就GIS在气象的应用进行探讨[]；哈斯塔木嘎阐述了GIS在气象行业的广泛应用与其强大的空间数据分析、处理、展示功能[]。

我国的GIS与气象方面的融合取得了很大的发展空间，但多数工作者的研究都是用GIS对气象数据进行某一方面的研究或应用，大多只是将历史数据进行对比，或者利用GIS的可视化，对某气象数据进行显示，我认为GIS空间分析方面的技术，应该对气象数据有一个更加深入的应用。

## 本文安排与组织结构

本文根据实际情况，介绍如何较好的利用开源的GIS工具——DotSpatial，实现了GIS的部分空间分析功能、可视化以及特殊气象数据文件格式的读取，并对其各个重要的功能模块与算法技术进行了分析和说明，其中包括对基于TIN的等值线算法进行创新，对空间插值算法与KNN进行融合。共有六章，具体内容编排如下：

1. 绪论。简单介绍了GIS的发展，GIS与气象方面的融合，以及当前热门的开源GIS软件的利用，通过国内外对开源GIS软件以及GIS与气象方面融合的研究现状，分析了使用开源GIS软件的好处，以及气象数据与GIS技术融合的宏大前景。
2. 气象分析处理的相关技术。介绍了一个开源的GIS可视化工具——DotSpatial的相关开发方式与背景情况；介绍了目前比较热门Web可视化图表工具——echarts；同时，也介绍了气象数据格式——NetCDF，以及将这个格式的数据转化为可视化图形的方法。
3. 气象数据分析处理工具系统设计。对功能需求做一个简要的概述，并针对需求，设计一个应用的基本结构以及功能模块。并对每个功能模块进一步分析其需要实现的算法或者插件。
4. 基于DotSpatial的插值方法与等值线实现。介绍了在GIS中广泛使用的空间插值算法和等值线生成算法，并加入了部分自己的改进优化。
5. 气象数据分析处理工具系统整体功能实现。本章展示了系统的主要功能，并对一些相似功能进行对比。
6. 总结与展望。对系统的开发与论文的研究进行总结，并对系统与未来的研究方向进行展望。

# 气象分析处理的相关技术

## Dotspatial类库及开发环境

### NET Framework和Visual Studio开发环境

.NET Framework是微软在2002年正式推出的用于Windows的新托管代码编程框架模型，用于构建具有视觉上引人注目的用户体验的应用程序，实现跨技术边界的无缝通信，并且能支持各种业务流程。

图2-1展示了.NET框架的体系结构，构建在操作系统最上层的服务为管理运行代码的公共语言运行时（Common Language Runtime，CLR），其中的代码可以用任何高级语言编写。Runtime提供了多种服务，这些服务有利于简化代码和应用程序的开发，同时也提高了应用程序的可靠性。.NET框架其中涵盖了一套可被用于任何编程语言开发的类库。在此之上的为许多应用程序模板，这些模板为网络站点和网络服务的开发提供了高级组建与服务。.NET框架包含了三个主要成分：公共语言运行时（Common Language Runtime，CLR）、多层次结构的统一类库集合（Base Class Library，BCL）和高级动态服务页面（Active Server Parge，ASP）。

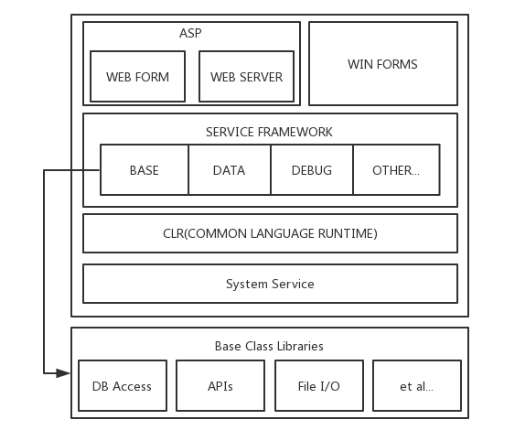


图 2-1 .NET框架体系结构

1. 公共语言进行时：CRL时.NET的运行环境。CLR负责托管程序的执行，线程管理、内存管理、自动垃圾收集、代码安全性检查、异常处理等多种功能。
2. 类库集合：BCL与CRL紧密集成，封装操作系统的底层API。BCL提供字符串管理、文件管理、内存管理、数据库操作、网络操作等常规功能，也能支持多类型的服务与程序，其中包括：控制台程序（Console Application）、Windows窗体程序（Windows Form）、ASP.NET Web程序、WPF（Windows Presentation Foundation）程序等。
3. 高级动态服务页面：ASP是由.NET框架提供的类库构建而成，它提供了一个Web应用程序模型。ASP为建立网络服务提供了一个高级可编译模板，并简化了应用程序的开发。当服务器收到对ASP文件的请求时，它会处理包含在用于构建发送给浏览器的HTML（Hyper Text Markup Language，超文本置标语言）网页文件中的服务器端脚本代码。除服务器端脚本代码外，ASP文件也可以包含文本、HTML（包括相关的客户端脚本）和com组件调用。

Microsoft Visual Studio是VS的全称。VS是美国[微软公司](https://baike.baidu.com/item/%E5%BE%AE%E8%BD%AF%E5%85%AC%E5%8F%B8)的开发工具包系列产品。VS是一个基本完整的开发工具集，它包括了整个[软件生命周期](https://baike.baidu.com/item/%E8%BD%AF%E4%BB%B6%E7%94%9F%E5%91%BD%E5%91%A8%E6%9C%9F/861455)中所需要的大部分工具，如UML工具、代码管控工具、[集成开发环境](https://baike.baidu.com/item/%E9%9B%86%E6%88%90%E5%BC%80%E5%8F%91%E7%8E%AF%E5%A2%83/298524)(IDE)等等。所写的[目标代码](https://baike.baidu.com/item/%E7%9B%AE%E6%A0%87%E4%BB%A3%E7%A0%81/9407934)适用于微软支持的所有平台，包括[Microsoft Windows](https://baike.baidu.com/item/Microsoft%20Windows)、[Windows Mobile](https://baike.baidu.com/item/Windows%20Mobile)、[Windows CE](https://baike.baidu.com/item/Windows%20CE)、[.NET Framework](https://baike.baidu.com/item/.NET%20Framework)、[.Net Core](https://baike.baidu.com/item/.Net%20Core)、[.NET Compact Framework](https://baike.baidu.com/item/.NET%20Compact%20Framework)和Microsoft [Silverlight](https://baike.baidu.com/item/Silverlight) 及[Windows Phone](https://baike.baidu.com/item/Windows%20Phone)。

### DotSpatial的类库与开发

DotSpatial是为.NET 4编写的地理信息系统库。它允许开发人员将空间数据、分析和映射功能整合到其应用程序中，或者为社区提供GIS扩展。Dan Ames博士是Dotspace的项目经理。 DotSaptial可以为.NET提供地图控制和GIS服务能力，包括在.NET的Win窗体或者Web应用中展示地图、打开矢量、网格、栅格和图像形式的文件、呈现符号和标签、漫游中再投影、操作和显示属性数据、科学化分析、读取GPS数据等。

DotSpatial项目组也一直在更新对其的开发与共享，2014年5月1日，开发人员在YouTube提供了一些简单项目的视频；2014年7月25日发布了dotspatial1.7；2015年6月5日更新了开发者基础教学和指导；2015年7月6日TFS的源控制被改写到Git上，2015年9月15日，DotSpatial1.8发布；2016年4月28日1.9版本发布；2016年5月6日，dotspatial迁移到github上，目前正在完善2.0的更新。

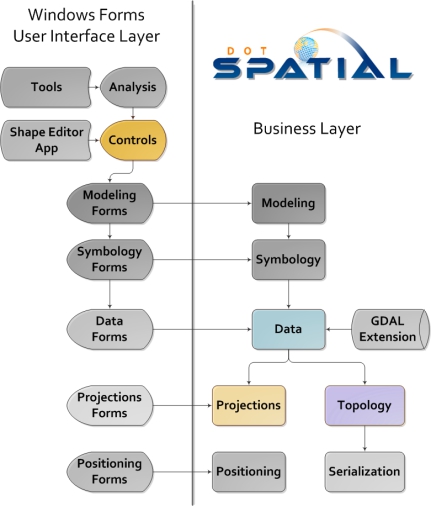


图 2-2 DotSpatial整体架构图

就目前来说，DotSpatial的最新稳定版本为1.9，类库与函数操作以dll程序集的方式提供，现在DotSpatial的相关dll程序集可以从基于.NET框架下的Nuget中获取程序包。

DotSpatial的类库分为用户界面层和业务逻辑层，用户界面层对应用户的操作交互，业务逻辑层负责主要功能的实现，利用这种方式能增加模块的内聚性，减少模块间的耦合性。图2-2为DotSpatial的框架结构图。

DotSpatial中的每个dll程序集都代表了一种GIS功能，常用的dll如表2-1所示：

|  |  |  |  |
| --- | --- | --- | --- |
| 程序集 | 主要接口 | 主要功能 | 主要效果 |
| DotSpatial.Controls | IMap，IMapLayer，ILegend等 | 地图控制 | 控制后台与前端的数据交互 |
| DotSpatial.Data | IRaster，IFeature，IDataSet等 | 数据访问 | 操作不同类型的常见GIS格式数据 |
| DotSpatial.Extensions | [IExtension](ms-its:C:\Users\54943\Desktop\毕不了业了\DotSpatial\DotSpatial.chm::/html/349d6574-e402-080a-830d-3e6e58dd291d.htm)，[IProjectFileProvider](ms-its:C:\\Users\\54943\\Desktop\\毕不了业了\\DotSpatial\\DotSpatial.chm::/html/f55b93ed-de23-9c9e-b5a1-d4cc28e7c123.htm)等 | 插件扩展 | 将外部的插件程序导入到主程序 |
| DotSpatial.Projections | [ICoordinateSystemCategoryHolder](ms-its:C:\Users\54943\Desktop\毕不了业了\DotSpatial\DotSpatial.chm::/html/583dfefa-32dd-8fdb-0e12-04d7be341fc1.htm)，[IProjMatchable](ms-its:C:\\Users\\54943\\Desktop\\毕不了业了\\DotSpatial\\DotSpatial.chm::/html/fa076279-9534-57c7-4e87-242e2268aba7.htm)等 | 投影转换 | 图形在地图中的投影类型与坐标系 |
| DotSpatial.Serialization | 无 | 序列化操做 | 转化对象信息或转换对象类型 |
| DotSpatial.Symbology | ICategory，IColorable等 | 符号化操作 | 对图层进行颜色，样式和字体装饰 |
| DotSpatial.Topology | ICoordinate，ILineString等 | 拓扑分析 | 对应坐标的点线面构建 |

表 2-1 DotSpatial主要程序集

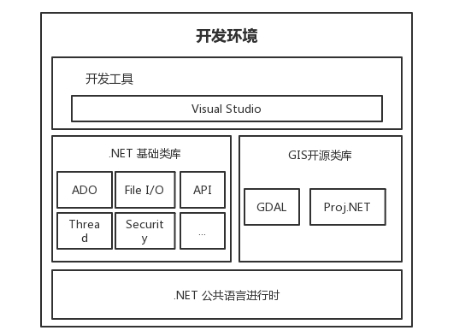


图 2-3 DotSpatial开发环境

其中DotSpatial.Controls、DotSpatial.Data、DotSpatial.Topology这三个程序集非常重要，基本上GIS的主要功能操作都在这三个程序集中。

此外，还有DotSpatial.Data.Rasters.GdalExtension程序集和Proj.Net库，其中GdalExtension封装了GDAL类库，效果是能够读取多种形式的栅格数据，proj.net支持多种空间坐标系的互相转换。

对DotSpatial进行开发的要求比较简单，使用Visual Studio的Winform Framework4.0的程序架构，以及.NET基础类库和包括GDAL在内的一些第三方开源库（如图2-3）。

关于DotSpatial的开发方式则是和winform类似，将DotSpatial.Controls.dll添加到依赖，根据此dll创建DotSpatial的工具栏，将其中的控件拖动到前端页面，对其属性和函数依赖在后台进行编写即可，具体如图2-4所示。

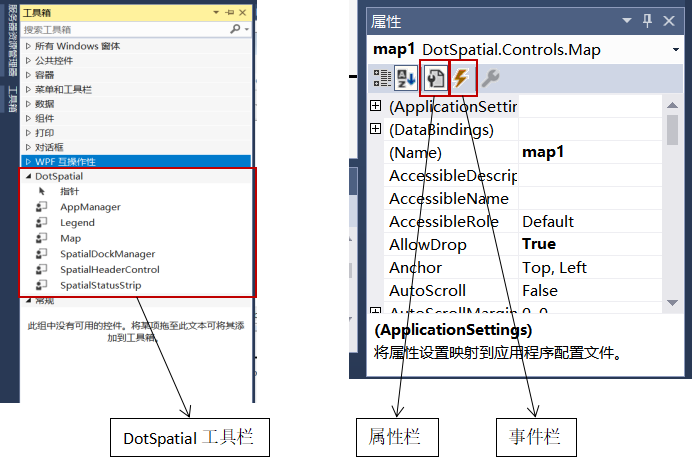


图 2-4 DotSpatial开发操作

上述是一般的开发方法，而DotSpatial的插件机制，允许使用Plugins子目录进行补充开发。通过插件加载的形式，插件将扩展功能提供给核心模块。Plugins处在文件的debug目录下，如图2-5所示。

开发者可以自己随意创建类库文件，将其集成Extension类，并将其输出路径导入到主程序文件夹中的plugins文件夹即可。

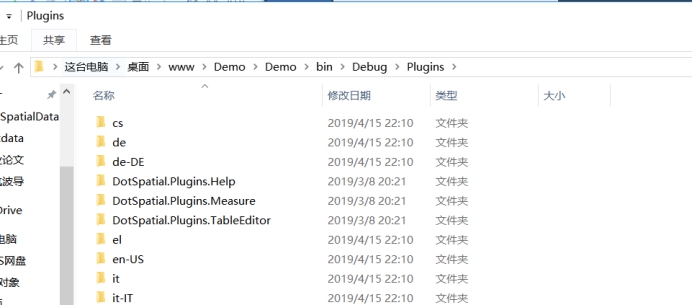


图 2-5 Plugins文件夹

## Web技术

### HTML5与JavaScript技术

HTML是[超文本](https://baike.baidu.com/item/%E8%B6%85%E6%96%87%E6%9C%AC/2832422)标记语言(Hyper Text Markup Language)的意思，[标准通用标记语言](https://baike.baidu.com/item/%E6%A0%87%E5%87%86%E9%80%9A%E7%94%A8%E6%A0%87%E8%AE%B0%E8%AF%AD%E8%A8%80/6805073)下的一个应用。HTML 不是编程语言，而是标记语言 (markup language)，用来制作网页。HTML5表示HTML的第五次重大修改，本次修改直接推动了Web的标准化运动发展，也使Web技术更加受到开发者关注。

HTML5的优点如下：

1. 提高可用性和改进用户的友好体验；
2. 丰富了Web的表现力与渲染力；
3. 强大的可移植性与跨平台特性；
4. 支持离线数据展示；

JavaScript是一种用于html中的脚本语言，通常用在html网页中，用来增强HTML网页的动态效果。JavaScript主要有六大特点[]：

1. 解释型脚本语言：JavaScript是一种解释性语言，以程序段的方式进行编程。基本结构形式与C、C++等语言类似，但不需要编译，而是在程序运行过程中逐行解释。
2. 基于对象的语言：JavaScript使一种面向对象的语言。可以创建于针对对象进行操作。
3. 简单性：JavaScript的变量类型与Python类似，为弱类型。
4. 安全性：JavaScript十分安全，不被允许访问本地硬盘，也不能操作服务器，有效地防止数据丢失。
5. 动态性：JavaScript可以直接由数据驱动，对事件作出反应，事件发生之后，相应的响应发生。
6. 跨平台性：JavaScript是配合HTML语言使用的，所以也拥有HTML的跨平台性，只要浏览器支持，就可以执行操作，与操作系统环境无关。

### echarts图表与json数据格式

ECharts是由百度开发的一个使用JavaScript实现的开源可视化库，拥有很强的跨平台性和可移植性，可以用于PC端和移动端，也可以兼容大部分浏览器，底层由轻量级图形库ZRender构建，提供丰富直观的可视化数据与操作。

Echarts现在已经出现在很多商用的宣传图表中，它有着如下的几个特点：

1. 丰富的可视化类型：Echarts提供了许多常规的图表形式，如常用的折线图，柱状图，散点图等，同时也可以提供三维的图表。
2. 千万数据的前端展示：通过增量渲染技术（4.0+），配合多种细致的优化方式，Echarts能展现千万级的数据量，并且用户依然能体验到流畅的交互操作，
3. 多渲染方案，跨平台使用：Echarts处理移动端的浏览器，还可以在node上进行高效的服务端渲染（SSR），于此同时，社区贡献者也提供了丰富的多语言扩展，如Python的pyecharts等。
4. 绚丽的特效：Echarts针对线，点等基本图形数据，提供了多种吸引眼球的特效。

其中，由于Echarts的语言为js，js的异步加载特性理所当然的应用在其中，echarts对数据集格式的要求比较宽松，自定义的类型、js数组以及json数据可是都是可以利用的数据格式，下面介绍json数据格式。

JSON(JavaScript Object Notation,JS对象简谱)是一种基于 [ECMAScript](https://baike.baidu.com/item/ECMAScript) (欧洲计算机协会制定的js规范)轻量级的数据交换格式，存储和表示数据的文本格式完全独立于编程语言。清晰简洁的结构层次使JSON成为理想的数据交换格式。由于在JS语言中，一切均是对象。任何类型都可以通过JSON表示，尤其是对象和数组。

对象：由“{Key:value,}”这样的键值对结构表示，key表示对象的属性，value表示对应的值。

数组：由“[“asa”,”sasad”,...]”的数组形式表示，由于js是弱类型变量，并没有要求数组中的具体每个值的格式，里面也可以用数组表示。

样例为图2-6所示，其中people为数组名，其中的数组元素是以对象的形式封装的，该对象有firstName和lastName两个属性，“：”后为对象的值。

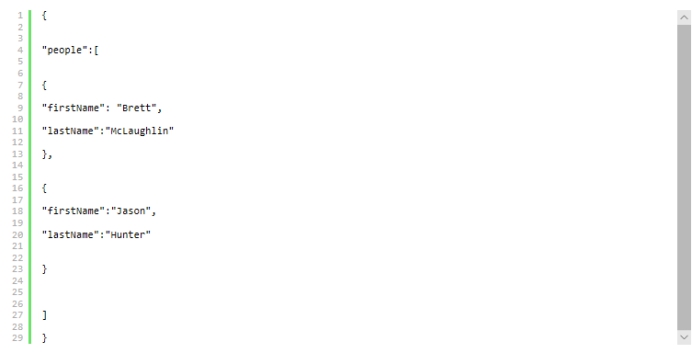


图 2-6 json数据格式

## NetCDF

### NetCDF数据格式

关于.NetCDF的数据格式全称为network Common Data Format，中文译法为“网络通用数据格式”。这种格式的数据已经广泛用于大气科学、水文海洋学等多个领域。

关于NetCDF格式文件存储的数据有三类，分别是变量，维度，属性。变量中存储着一个数组，数组中的数表示该变量的值；维度表示了该文件中存储数据的坐标轴情况，通常的二维数据中会有latitude（纬度）和longitude（经度）两个维度名称；属性则表示某个变量的属性情况，比如该变量的单位。图2-7是一个简单的NetCDF数据格式。

从中可以看到，dimensions为维度，图2-7中有x，y两个维度，x维度的长度为6，y维度的长度为12；变量名为data，data中的数据排列格式为x\*y，也就是说，第一行是(0,12)，也就是当x=0时，y=0 ~ 11的数据情况，图2-7是一个很简单的NetCDF文件格式，其中没有属性。属性一般与变量是紧密相关的，每个变量的属性种类和数目都不一样，常见的变量属性有units（数据单位），long\_name（变量的详细名称），add\_offset（变量的偏移值），scale\_factor（变量的缩放因子），valid\_range（变量的有效数据范围）等，图2-8是我使用NASA的Panoply软件打开MeteoInfor中Sample的NetCDF文件，描述某一年12个月的温度情况的大致数据。

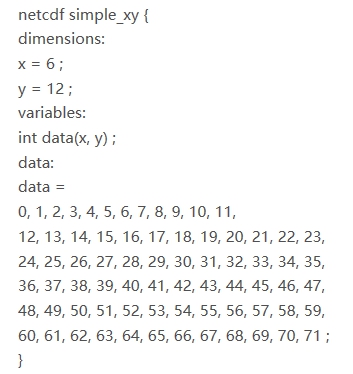


图 2-7 一个简单的NetCDF数据

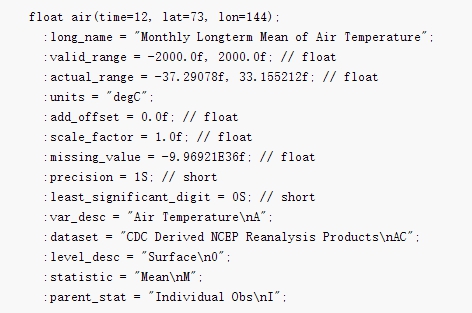


图 2-8 NetCDF属性数据样例

### NetCDF类库

虽然NetCDF的数据貌似很清晰，但是是无法用一般的txt方法进行读取的，在NetCDF的官方网站Unidata（https://www.unidata.ucar.edu/）中有关于读取NetCDF文件的库以及其函数（C语言）。如nc\_inq\_dimname(int ncid,int dimid,char\* name)，通过参数ncid（读取的.nc文件的序号），dimid（文件中维度的序号）得到当前文件某维度序号下，该维度的名称。关于更多函数的功能与使用方式，在网站https://www.unidata.ucar.edu/software/netcdf 中可以查阅。

同时，对于C#调用C语言的dll库，需要使用.NET框架提供的P/Invoke平台调用功能实现。P/Invoke负责C#托管语言与C非托管语言之间的通信。P/Invoke的使用方式也比较简单，在代码中定义类库文件，用DLLImport包装方法，方法的名字与C函数的名字相同，并将netcdf.dll类库加载在文件的bin/Debug文件中，即可调用。如图2-9所示是对nc\_open、nc\_create、nc\_close方法的包装。



图 2-9 包装NetCDF函数

## 本章小结

本章着重介绍了本文气象数据分析处理工具的相关技术，主要包括：开源GIS工具.NET架构下的DotSpatial的使用方法；三维可视化图表的echarts技术以及实现其需要的js语言与json数据格式；NetCDF数据集的格式与读取方法。并在上文中，详细说明了各项技术与气象数据分析处理的关系，以及被选择的原因。

# 气象数据分析处理工具系统设计

## 需求分析

本系统旨在使用开源或免费的工具实现对气象数据的分析处理，包括了对一些地理数据格式文件的读取；对文件中的一些气象数据进行进行绘图；根据本机储存的气象数据进行图表分析处理。

为了使用GIS技术实现对气象数据的分析，需要使用GIS开源软件的技术实现对图像文件数据的读取显示，以及相关基本操作；此外，由于是对气象数据的处理，要了解气象数据的相关格式，根据美国国家航空航天局（National Aeronautics and Space Administration，NASA）和美国国家海洋和大气管理局（National Oceanic and Atmospheric Administration，NOAA）的数据约定，需要能够读取.csv站点数据格式，以及NetCDF数据集；由于站点数据一般布局较为分散，可以看作属性点，这里需要采用GIS的空间插值算法将其区域展示，并使用等值线算法反应区域关系；此外，还可以用图表的形式将站点数据或历史数据进行二维或三维的输出。

本系统根据开源的GIS软件，可进行数据的离线数据分析处理，用GIS的空间算法与Web的丰富可视化为用户提供良好的数据处理分析功能。需要工具为用户提供的功能如表3-1所示。

表 3-1 工具需要的基本功能

|  |  |
| --- | --- |
| 功能 | 效果 |
| 文件读取 | 读取气象数据与地理数据相关格式的文件 |
| GIS基本操作 | 对地图中显示的图形进行缩放，漫游，打印，测量，标记等相关操作 |
| 空间插值 | 使用插值算法对站点数据进行空间插值 |
| 等值线处理 | 使用等值线算法对矢量数据或栅格数据进行处理 |
| 图表展示 | 将得到站点数据进行图形化展示 |

1. 文件读取：对矢量文件（.shp、.dxf等），栅格文件（.adf、.bgd等），站点数据文件（.csv），NetCDF文件(.nc)。
2. GIS基本操作：缩放，漫游，打印，测量，标记等。
3. 空间插值：反距离权重插值（Inverse Distance Weight，IDW）、普通克里金插值（Ordinary Kriging，OK）。
4. 等值线处理：不规则三角网（triangulated irregular network，TIN）、基于TIN的插值算法、基于Grid的插值算法。
5. 图表展示：二维折线图、二维柱状图、三维散点图、三维柱状体；

此外，本工具所需要的工具为表3-2所示：

表 3-2

|  |  |
| --- | --- |
| **工具类型** | **工具使用** |
| 地理数据处理工具 | DotSpatial |
| 图表数据处理工具 | echarts |
| 窗体界面设计工具 | Winform（.NET framework4.0+） |
| 编译语言使用 | C#、javascript、html5、c |
| 集成开发编译器 | Visual Studio |

## 总体设计

### 功能模块划分

根据气象数据处理的需求，可以确定气象数据分析处理工具的设计目标，在总体设计阶段可以将功能模块划分为四个部分，分别是GIS基础操作，文件处理，空间分析，图表分析四个部分。具体模块与功能之间的关系如图3-1，表3-3所示。

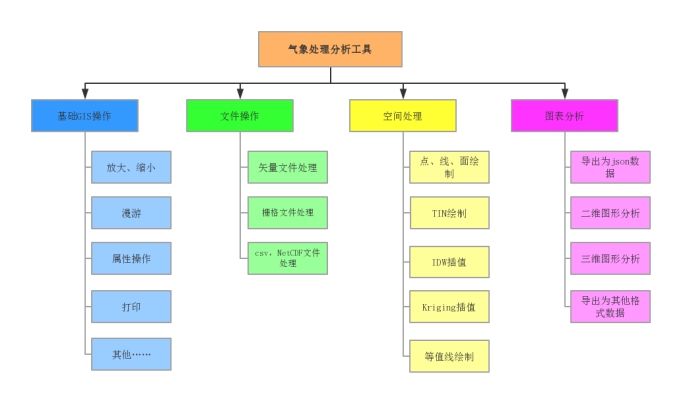


图 3-1 功能模块划分

表 3-3 功能模块描述

|  |  |
| --- | --- |
| **功能模块** | **作用介绍** |
| 基础GIS操作 | 基础的GIS操作主要是对图层进行操作，包括放大、缩小图层；对图层进行拖拽漫游；对图层的属性数据进行增删改操作；打印图层的形状及相关地理数据等。 |
| 文件操作 | 通过GDAL插件的处理，对图层文件的加载格式更有包容性，可以读取矢量与栅格文件，比如Vector中的shp和dxf等，Raster中的adf和bgd等，Image中tif和jpg等。同时通过对csv文件的理解以及对NetCDF的插件处理，可以对csv和NetCDF进行处理操作。 |
| 空间处理 | 空间处理首先是可以绘制基本的点线面，通过这些图形，对之前的图层进行添加；有了这些基本的图形后，可以绘制不规则三角形，可以根据不规则三角形的z值以及等值线光滑算法绘制等值线；空间插值是通过IDW或Kriging算法通过对站点数据的提取，进行插值，从而实现填色图的绘制。 |
| 图表分析 | Echarts可以异步加载数据，但要有json格式的数据集，所以把要分析的数据变成json文件，在使用echarts进行分析的时候，就直接可以调用相关数据成表分析。同时也可以将数据导出为其他形式的文件，比如excel，通过office的工具也可进行统计分析。 |

### 算法设计

系统所使用的GIS算法主要分布在空间处理中的空间插值算法和等值线生成算法。其中空间插值法又分为反距离权重插值法（Inverse Distance Weight，IDW）和普通克里金插值法（Ordinary Kriging，OK）；等值线生成算法分为基于不规则三角形的等值线生成算法与基于网格的等值线生成算法，如图3-2所示。

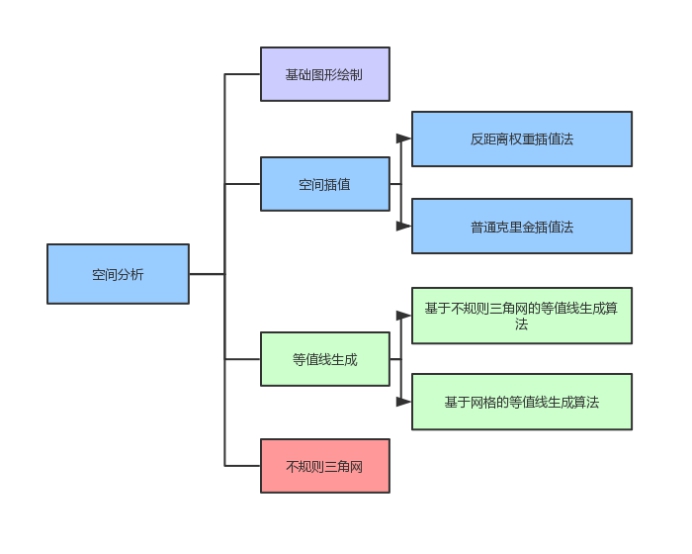


图 3-2 空间分析功能模块

而其中的插值算法需要很多算法辅助实现，IDW算法需要通过KD树（K-Dimension Tree）实现KNN（K-NearestNeighbor）算法确认参与计算的邻近点；而OK算法则是需要使用模型的算法，而每个模型都需要通过拟合得到系数，所以OK算法需要用非线性拟合的方法获取模型参数，同样也要用KNN去确认参与计算的临近点。

而在等值线生成算法中的基于TIN的等值线生成算法中，则需要构建不规则三角网，并进行等值线追踪，在进行等值线追踪后，需要对特殊点所在的等值线做改良处理，处理完成后再对等值线进行光滑插值。

对于插值算法与等值线生成算法，图3-3显示每种算法其中的算法组成，以及表3-4对其功能进行详细描述。

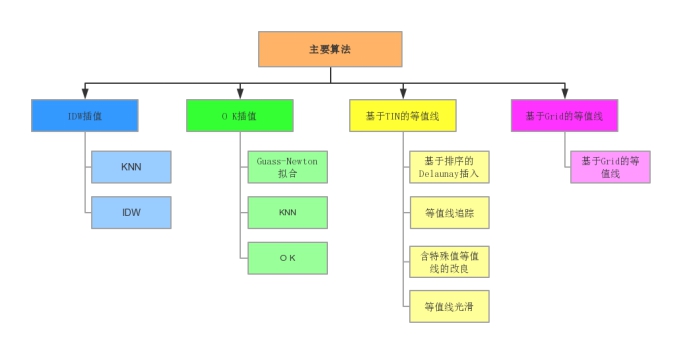


图 3-3 主要算法的结构

表 3-4 算法构成与相关作用

|  |  |  |
| --- | --- | --- |
| **功能模块** | **算法构成** | **算法作用** |
| IDW插值 | KNN | 寻找某个点附近的k个最邻近的点 |
| IDW | 利用权重进行与距离成反比关系的插值 |
| O K插值 | Guass-Newton拟合 | 通过数据对OK法的模型进行参数拟合 |
| KNN | 寻找某个点附近的k个最邻近的点 |
| O K | 通过距离与半方差关系的模型进行插值 |
| 基于TIN的等值线生成 | 基于排序的Delaunay插入法 | 根据点的某个维度排序再进行Delaunay插入构成TIN |
| 等值线追踪 | 通过TIN，在其边上进行点的计算与连接 |
| 含特殊值等值线的改良 | 避免含特殊值的等值线在光滑后畸形化 |
| 等值线光滑 | 在等值线点间进行插值，使其光滑 |
| 基于Grid的等值线生成 | 基于Grid的等值线生成 | 通过格网中4个顶点的特征值与其相互关系构建等值线 |

## 界面设计

### 系统主界面设计

系统主界面的第一栏工具栏用来实现文件的相关操作以及插件的调度，第二栏工具栏主要是利用小图标实现GIS的常规操作，如缩放、拖拽、定位、要素查询、属性显示等操作。

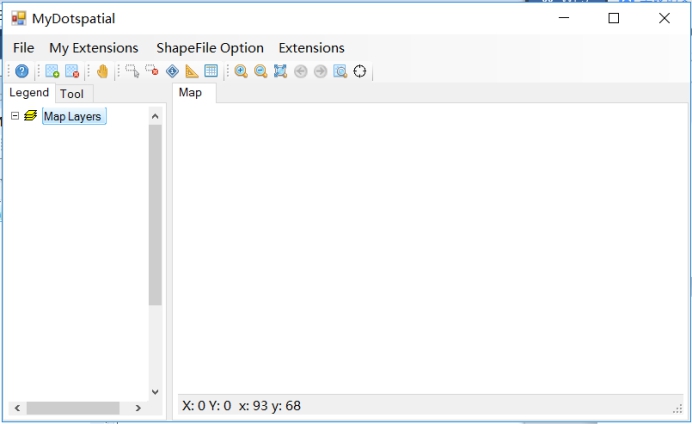


图 3-4 主界面设计1

下面的主题部分分为图层信息，工具栏与地图显示。图层信息主要显示缓存在地图中的图层的名称，属性及操作控制相关图层；工具栏主要是对图层数据进行加工处理，如图表化、空间分析等操作；而地图显示主要是可视化显示图层，如显示shp，raster，img图像等。针对上述构想，对系统主界面的设计如下所示：

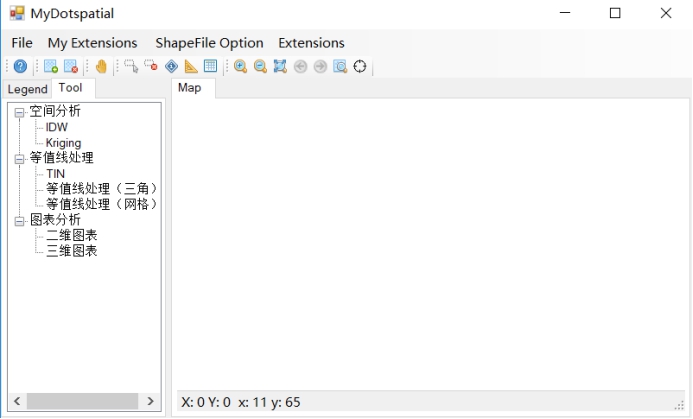


图 3-5 主界面设计2

### 插值界面设计

根据两种插值方法所需要的数据，需求如表3-5，界面如图3-6，3-7所示。

表 3-5 插值数据需求

|  |  |  |
| --- | --- | --- |
| 插值方法 | 输入需求 | 输出需求 |
| IDW | 点要素集，属性值，像元大小，权值，邻近数目 | 栅格文件 |
| Ordinary Kriging | 点要素集，属性值，像元大小，模型，模型参数（可选）、邻近数目 | 栅格文件 |

此外，需要右边的工具栏显示插值算法的含义与其每部分的意义，需要进度条显示在每次插值进行时的进度。

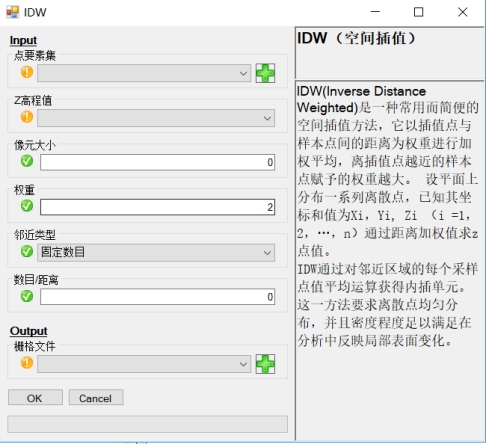


图 3-6 IDW插值方法界面



图 3-7 Kriging插值方法界面

对于插值分析的操作来说，需要得到点要素集，然后对点要素集需要进行插值的属性做一个选择，因为插值后产生的图层为栅格数据，所以需要得到栅格数据中的像元大小，作为输入项，还需要获取参与计算的邻近点数，最后选择一个栅格文件为输出对象。

对于不同的插值方法，有不同的特殊需求，在IDW中，则需要的是认为获取权值信息，在Ordinary Kriging算法中，需要获取使用的模型且直接或间接地获得模型的参数信息。

### 等值线界面设计

等值线分为基于TIN的等值线生成方法与基于Grid的等值线生成方法，由于TIN是基于点要素集，而Grid是基于栅格网格的，所以，等值线的界面需求如表3-6，图3-8，3-9所示。

表 3-6 等值线数据需求

|  |  |  |
| --- | --- | --- |
| 等值线生成方法 | 输入需求 | 输出需求（可选） |
| 基于TIN | 点要素集，属性值，等值线起始值，等值线间距 | 线要素矢量文件 |
| 基于Grid | 栅格文件，等值线起始值，等值线间距 | 线要素矢量文件 |

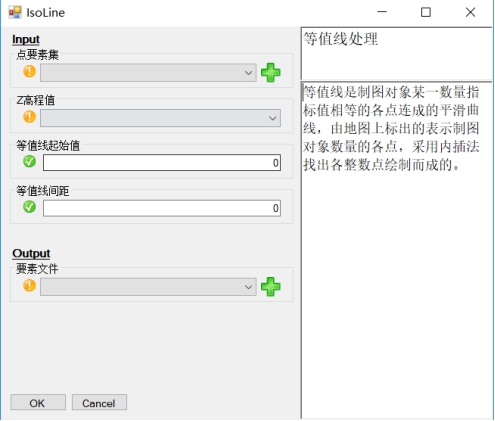


图 3-8 基于TIN的等值线生成方法界面

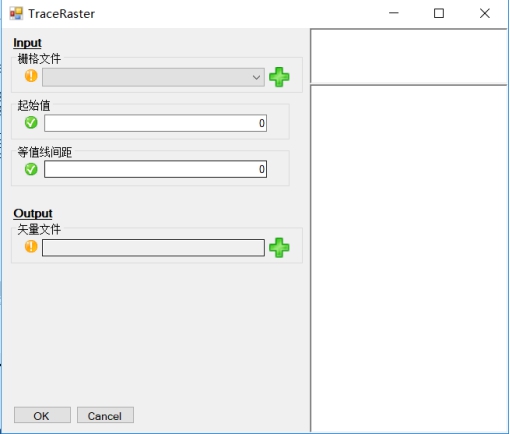


图 3-9 基于Grid的等值线生成方法界面

### 特殊格式数据读取

特殊格式数据主要是指.csv数据格式以及NetCDf数据格式，其中对特殊文件的读取主要是通过了相关文件的格式，以及利用官方发布的对文件读取处理的函数进行处理。

csv文件：第一行为属性名称，其他行为对应的属性数据，行与行之间用换行符分开，列与列之间用逗号分隔。

NetCDF文件：文件的详细数据结构如文中2.3小节所述，这里不再赘述。

根据其性质分析，得到文件的数据需求如图3-10，3-11，表3-7所示。

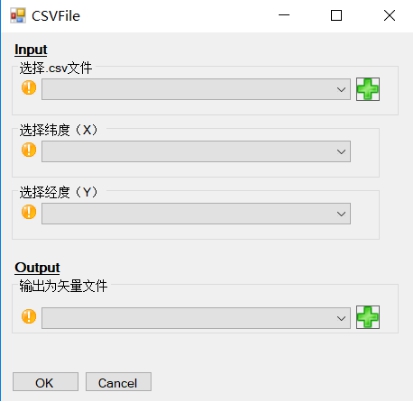


图 3-10 .csv文件读取界面

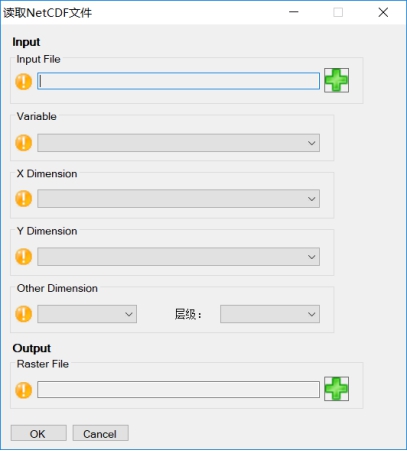


图 3-11 NetCDF数据集读取界面

表 3-7 特殊格式数据读取需求

|  |  |  |
| --- | --- | --- |
| 文件类型 | 输入需求 | 输出需求 |
| CSV | .csv文件，经纬度数据选择 | 点要素矢量文件 |
| NetCDF | .nc文件，变量，经纬度，其他维度，层次（三维情况下） | 栅格文件 |

## 本章小结

本章主要介绍了气象数据分析处理工具的功能模块与算法设计，并对其中所运用的算法与技术进行了分析归纳，详细描述了设计思想、算法设计以及界面设计，对于这些设计条件，进一步填充内容。

# 插值方法与等值线算法设计

## 基DotSpatial的空间插值算法设计

### 反距离权重插值

根据章节3.2.2所述，反距离权重插值的空间分析方法中，包含了IDW算法本身和利用KD树实现的KNN算法，这一节详细描述具体实现过程

IDW：

IDW是一种最简单的插值方法，它以插值点与样本之间的距离为权重进行加权平均，插值点越近的样本点赋予的权重越大，优点是实现简单，当幂值选取得当的时候，插值效果准确；缺点在于反距离权重法主要依赖于反距离的幂值，与任何实际的物理过程都不关联，导致不同空间不同属性的幂值都不相同，无法挑选一个最合适的幂值会导致插值不太准确。

IDW的加权函数如公式（3-1）所示，其中i表示为点要素中的第i个点，p表示反距离使用的幂值（在0.5与3之间取得最佳效果，一般取2），表示第i个点要素到插值点的距离。

|  |  |  |
| --- | --- | --- |
|  |  | (四‑1) |

其中，h表示两点之间的欧式距离，求法如式子（3-2）所示，、表示第i个点要素的横、纵坐标。

|  |  |  |
| --- | --- | --- |
|  |  | (四‑2) |

通过式（3-1）求得每个要素点对于插入点的权重后，根据式（3-3）即可获得该插入点的属性值，其中、分别为第i个点的所占的权重与属性值，

|  |  |  |
| --- | --- | --- |
|  |  | (四‑3) |

KD树：

KD树是一种用于K维上构建数模型的平均二叉树，用于多维空间的数据搜索。KD树每一层分支都将方差最大的维度从中进行二等分。构建二维的KD树的伪代码如下：

|  |
| --- |
| 算法：KD树构建 |
| 输入：包含维度坐标的点集数据Data-Set |
| 输出：KD树 |
| If 点数据集为空，则返回空的KD树  This = KD.root  While（Data-Set != NULL）  计算当前节点this的Data-Set各个维度的方差，比较得出方差最大的维度D  按照K维进行Sort  获取重新排序点集的中间点p，this.point = p，this.dimension = D  按照中心点p的左右点集分裂为两个点集data-set1，data-set2，在当前节点创建两颗子树leftchild，rightchild  使Leftchild.Data-Set = Data-Set1，Rightchild.Data-Set = Data-Set2  This = Leftchild  This = Rightchild  End While |

按照伪代码，有如下数据（2，3）、（5，4）、（4，7）、（9，6）、（8，1）、（7，2），可以构成如图4-1的KD树结构。

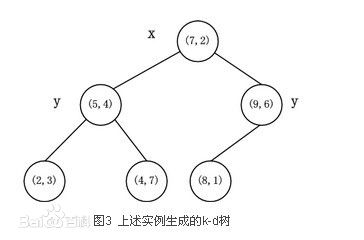


Figure 4‑‑ KD树结构

KNN：

KNN表示矢量空间上的某点的K个相邻最近点，KNN可以在KD树的架构上使用，通过KD树可以有效减少搜寻K个相邻最近点的开销。在KD树上实现KNN的方法如下伪代码所示：

|  |
| --- |
| **算法：KNN** |
| **输入：KD树，查询点thepoint、需要的最近点数目** |
| **输出：K个相邻最近点** |
| **创建一个栈存储K个节点** |
| **this = KD.root**  **创建空间大小为K的链表collection**  **//二叉查找**  **While(this.point != NULL)**  **计算this.point,thepoint的欧式距离distance**  **If(collection.size<K)**  **Collection.add(this.point)**  **Else**  **计算Collection.max与thepoint的欧式距离distancemax**  **If(distance(distancemax>distance)**  **Collection.remove(Collection.max)**  **Collection.add(this.point)**  **EndIf**  **Search(this.leftChild)**  **Search(this.RightChild)**  **//回溯**  **this.parent**  **计算thepoint与this.parent.dimension的欧式距离distance**  **While(distance < distancemax)**  **BackSearch(this.parent)**  **EndWhile** |

将以上三种算法融合，可以形成对反距离权重插值的算法设计，流程图如图4-2所示。

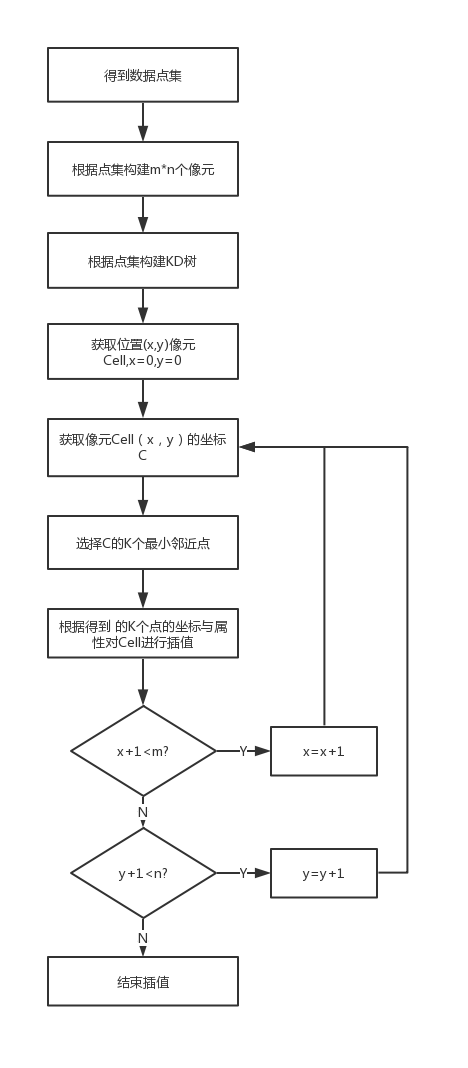


Figure 4‑‑ IDW插值算法流程图

### 普通克里金插值

根据章节3.2.3所述，普通克里金插值的空间分析方法中，包含了OK算法本身、利用KD树实现的KNN算法以及用于拟合空间模型的高斯牛顿拟合法，这一节详细描述具体实现过程。

OK算法：

克里金算法使通过模型对其变成进行拟合，得到适用于该空间的模型参数，再通过确认的模型函数进行插值。优点是由于克里金算法是根据样本点进行最适应模型的拟合，所以再空间插值中使用的较多，效果也较好；缺点在于拟合需要时间，且数据差异较大时，得到的插值效果也较差。

在克里金算法中，有泛kriging，Simple Kriging，Ordinary Kriging等，由于OK算法难度适中且使用广泛，故选择OK算法。

首先要通过模型拟合已知点距离distance和点对的半方差semivariogram，距离为欧式距离，计算公式如4-4所示。

|  |  |  |
| --- | --- | --- |
|  |  | (四‑4) |

式中，表示点i的属性值，表示点i与点j的半方差。

Kriging中，有多个模型可供选取，其中常见的有指数模型，高斯模型，球形模型，如图4-3，4-4，4-5所示。

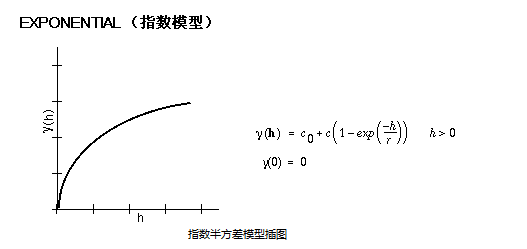


Figure ‑克里金指数模型

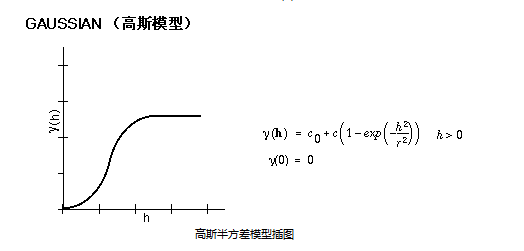


Figure ‑克里金高斯模型

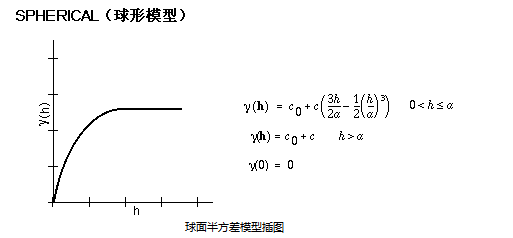


Figure ‑ 克里金球形模型

在上述式子中，h表示点间距，为点间距对应的半方差，c0表示块金值，通常情况下取0，a或r表示变程值，c表示偏基台值，c与a(r)共同描述点间距与半方差的关系。

在通过拟合得到c与a的值之后， 可以通过式(4-5)求点集的相应权重z。

|  |  |  |
| --- | --- | --- |
|  |  | (‑5) |

式中，表示第n个点的权重，表示模型求得的点i与点j的半方差，得到了相应的权重后，根据公式（4-3）可以得到对应位置的值。

Guass-Newton拟合法：

Guass-Newton法式根据jacobian矩阵存储对函数中参数进行的一次偏导，并将其作为步长逼近函数最优解。其基本思想是使用泰勒级数展开式去近似地代替非线性回归模型，然后通过多次迭代，多次修正回归系数，使回归系数不断逼近非线性回归模型的最佳回归系数，最后使原模型的残差平方和达到最小。

|  |  |
| --- | --- |
|  | (‑6) |

泰勒展开的公式如4-7所示,忽略高阶项可以将其看作式4-8。

|  |  |  |
| --- | --- | --- |
|  |  | (‑7) |

此外需要残差residual来计算步数step，残差表示根据当前参数计算出来的值与样本值的插值。残差越大，表示当前参数值与真是只的擦话剧越大，而步数step表示新的属性值与旧属性值的差值，即new parameter = old parameter + step。而step的求解方法如下所式4-8所示。

|  |  |  |
| --- | --- | --- |
|  |  | (‑8) |

Guass-Newton法的优点在于简单易于实现，且求解速度快，缺点在于对初始参数依赖程度太强，且当步长求解太长时导致近似不够准确使得结果不收敛。

Guass-Newton拟合法求解OK算法模型的流程图如图4-6所示。

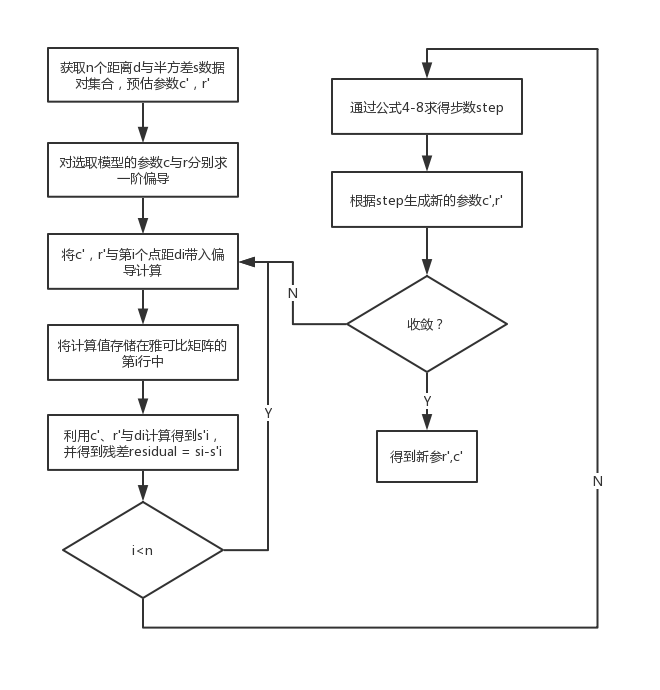


Figure ‑ 基于克里金模型的高斯牛顿拟合流程

将OK算法，KNN与Guass-Newton进行算法设计，流程如图4-7所示。

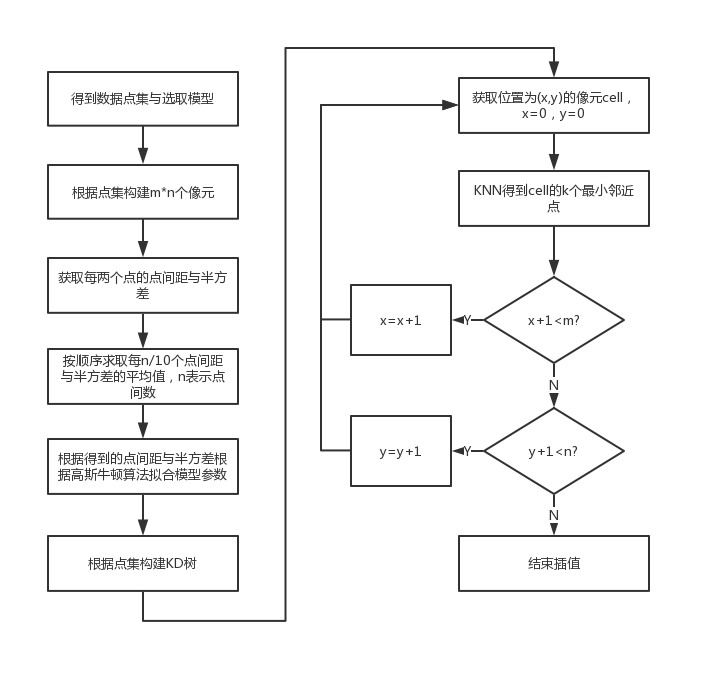


Figure ‑ OK插值算法流程图

## 基于DotSpatial的等值线算法设计

### 基于TIN的等值线生成算法

基于TIN的等值线是根据不规则三角形的特点，对分别并非均匀的站点原始数据进行分析，优点在于，三角网不会要求数据均匀分布，且不会受插值的误差影响，只根据站点关系分析高程值；缺点在于生成的等值线太粗糙，需要进行插值光滑，而插值法不同又会造成等值线较差缠绕的情况，所以还需要对基于TIN的等值线生成算法进一步进行改进。

基于TIN的等值线生成算法分为四步，分别是TIN的生成、等值线追踪、等值线改良、以及等值线光滑。其中1，2，4步是基于TIN的等值线生成算法中原有的，第3步是我的改进。

TIN的生成：

对于不规则三角网的生成，使用最为普遍的是逐点插入法,但逐点插入法会因为点的位置关系，而出现不断进行退化（原来的三角形进行分裂），重组，过分消耗时间，于是有人提出基于排序的逐点插入法[[[1]](#endnote-1)]。其核心思想是对点集先进行x轴的排序，再通过三角形外接圆的定义进行插值，其伪代码如下所示：

|  |
| --- |
| 算法：基于排序的逐点插入法 |
| 输入：顶点列表（vertices） |
| 输出：已确定的三角形列表（triangles） |
| 初始化顶点列表  创建索引列表(indices = new Array(vertices.length))  基于vertices中的顶点x坐标对indices进行sort  确定超级三角形  将超级三角形保存至未确定三角形列表（temp triangles）  将超级三角形push到triangles列表  遍历基于indices顺序的vertices中每一个点  初始化边缓存数组（edge buffer）  遍历temp triangles中的每一个三角形  计算该三角形的圆心和半径  IF该点在外接圆的右侧  则该三角形为Delaunay三角形，保存到triangles  并在temp里去除掉  Break  IF该点在外接圆外（即也不是外接圆右侧）  则该三角形为不确定  Break  IF该点在外接圆内  则该三角形不为Delaunay三角形  将三边保存至edge buffer  在temp中去除掉该三角形  对edge buffer进行去重  将edge buffer中的边与当前的点进行组合成若干三角形并保存至temp triangles中  将triangles与temp triangles进行合并  除去与超级三角形有关的三角形  End |

等值线追踪：

在生成了TIN之后，可以通过TIN进行等值线追踪。根据同一高程的等值线穿过一个三角形最多一次，所以根据三角形的顶点特征，在三角形内构建等值线。

设三角形顶点的Z值分别是Z1，Z2，Z3，

1. Z1，Z2，Z3全大于或小于等值线特征值时，此三角形内无等值线；
2. Z1，Z2，Z3有等于等值线特征值的情况，则对其加或减一个对结果影响较小的，此时应该是。
3. Z1，Z2，Z3有大于和小于等值线特征值时，此三角形内必有等值线，且等值线有如下特征：(Z1-Z)(Z2-Z)>0该边无等值点，否则有；(Z1-Z)(Z3-Z)>0该边无等值点，否则有；(Z3-Z)(Z2-Z)>0该边无等值点，否则有。且上述情况必有两条边有等值线；
4. 边中等值点的X轴位置的求解方法如式4-9所示，Y轴同理。

|  |  |  |
| --- | --- | --- |
|  |  | (四‑9) |

以上4点是等值线追踪的主要思想，以下给出等值线追踪的伪代码。

|  |
| --- |
| 算法：等值线追踪 |
| 输入：特征值Z，不规则三角网triangles，三角网边集edges，三角网点集points |
| 输出：特征值为Z的等值点集line |
| 创建队列line  IF triangles is NULL  Return line = NULL  While triangle == NULL  While  从triangles中任一选择一个三角形triangle  If triangle的顶点值包括Z&&未被搜寻过  根据公式4-9得到triangle含有的等值点p1，p2，将其存到line中  记录p2所在的边e2  将triangle标记为已使用过  Break  Else  将triangle标记为已使用过  EndWhile  根据edges与triangle寻找与e2共边的三角形triangle’  While 不存在triangle’||triangle’已被使用  If triangle’的顶点值包括Z&未被搜寻过  根据公式4-9得到triangle含有的等值点p1，p2，将其存到line中  记录p2所在的边e2  将triangle标记为已使用过  Else  将triangle标记为已使用过  根据edges与triangle寻找与e2共边的三角形triangle’  EndWhile  EndWhile |

等值线光滑：

等值线光滑实质上是对等值线进行插值操作，原理是对一个折现进行多个点的插入后，宏观上呈现为一条曲线。等值线光滑算法有三次B样条函数法，五点光滑法等，因为光滑算法要求过点，结合简单适用性考虑，这里使用抛物线样条函数法。

抛物样条法是以简单抛物线作为曲线拟合离散点曲线的算法，对不在同一直线上的点p1，p2，p3进行如式4-10的表达式描述，并通过3个独立条件如下。

1. 参量t=0时，抛物线以p1为起点，曲线过p1点；
2. 参量t=1时，抛物线以p3为终点，曲线过p2点；
3. 参量t=0.5时，抛物线过p2点，且该店的切矢量等于p3-p1.

|  |  |  |
| --- | --- | --- |
|  |  | (四‑10) |

在上述条件下联立4-10求解方程组，得到式4-11.

|  |  |  |
| --- | --- | --- |
|  |  | (四‑11) |

### 对TIN等值线生成算法的改进

### 基于网格的等值线生成算法

## 本章小结

# 气象数据分析处理工具系统整体功能实现

## 基础GIS功能实现

## 算法功能实现

### 空间插值算法

### 等值线算法

## echarts图表绘制功能实现

## NetCDF数据插件开发实现

# 总结与展望

## 总结

## 研究展望

参考文献

致 谢

1. [] <https://github.com/ironwallaby/delaunay> [↑](#endnote-ref-1)