

一种顾及高程的时空反距离加权插值方法

王彬彬¹, 石丽红^{1,2}, 卢月明²

(1. 辽宁工程技术大学 辽宁 阜新 123000; 2. 中国测绘科学研究院 北京 100830)

摘要: 反距离加权插值法是空间分析中插值的一种常用方法,被广泛应用于各个领域的插值计算中。针对反距离插值法(IDW)中未考虑时间因素与高程因素的影响,本文提出一种顾及高程因素的时空反距离加权插值法(H-TIDW)。该方法首先利用时间序列分解法去除数据的季节性波动,然后综合考虑时空距离与高程因素构建时空插值模型进行预测,最后在预测结果上叠加季节性波动得到最终的预测值。验证结果表明:顾及高程的时空反距离加权插值法将时空信息考虑到距离计算之内,从时间和空间两个维度对PM_{2.5}浓度数据进行分析研究,避免了只考虑空间或者时间而造成大量有价值信息丢失,且插值精度也有一定程度的提升。

关键词: 反距离加权插值; 时空反距离插值; 高程; PM_{2.5}

中图分类号: P215

文献标识码: A

文章编号: 1672-5867(2018)10-0082-04

A Space-time Reverse Distance Weighted Interpolation Method Considering Elevation

WANG Binbin¹, SHI Lihong^{1,2}, LU Yueming²

(1. Liaoning Technical University, Fuxin 123000, China;

2. Chinese Academy of Surveying and Mapping, Beijing 100830, China)

Abstract: Reverse distance weighted interpolation is a common method of interpolating in spatial analysis and is widely used in interpolation of various fields. Aiming at the influence of time factor and elevation factor in inverse distance interpolation (IDW), this paper proposes a space-time inverse distance weighted interpolation method (H-SIDW) which takes into account elevation factors. In this method, the time series decomposition method is used to remove the seasonal fluctuation of the data. Then the spatial and temporal distance and the elevation factor are used to construct the spatiotemporal interpolation model. Finally, the seasonal prediction is superimposed on the prediction result to obtain the final prediction value. The results show that the space-time inverse distance weighted interpolation method takes into account the space-time information into the distance calculation and analyzes the PM_{2.5} concentration data from the two dimensions of time and space, avoiding the consideration of space or time. A lot of valuable information is lost, and the interpolation accuracy also has a certain degree of promotion.

Key words: IDW; spatiotemporal inverse distance interpolation; elevation; PM_{2.5}

0 引言

空间插值常用于将离散点的测量数据转换为连续的数据曲面,即空间插值是用实测的空间数据来估算(预测)未知空间数值的过程,由于空间数据具有数据量大的特点,要获得整个研究区内的数据,通常采用空间插值方法。现有的空间插值方法有泰森多边形法、反距离加权法、梯度距离反比法、样条函数法、趋势面法、面积插值法、普通克里金法等,但没有一种算法适用于所有空间插

值估算未采样区域数据值^[1]。其中,反距离加权插值是GIS空间插值的一种常用方法,具有插值原理简单、插值速度快等优点,因此该方法广泛应用在气象、农业与海洋等领域。

近年来,国内外学者对空间插值方法进行了大量的研究^[2-4],发展了诸多方法,其中也有一部分学者对经典的空间插值进行了一系列的优化改进^[5-7]。各种不同的空间插值方法各有各的优缺点。反距离加权插值法(IDW)是一种使用一组采样点的线性权重组合来确定插

收稿日期: 2017-10-11

基金项目: 国家基础测绘科技与标准计划——面向优化国土空间开发格局的基础地理数据空间分析系统及应用(2017KJ0104)资助

作者简介: 王彬彬(1992-),女,辽宁营口人,测绘科学与技术专业硕士研究生,主要研究方向为地理空间数据挖掘。

值点属性值的估计方法,是基于相近相似的原理对插值点来进行插值。李章林等指出了传统的反距离加权法的优缺点,并在此基础上提出了改进后的指示反距离加权法,提出的新方法充分利用了采样点数据提供的各种有效信息,提高了插值精度^[8];段平等结合自然邻近关系,提出自适应的反距离权重(Adaptive-IDW, AIDW)插值方法^[9];李佳霖等提出一种顾及风向和风速的空气污染物浓度插值方法(DVIDW)^[10];高真等通过分析三维水声数据特点,引入反距离加权插值算法使用人工干预实现水声数据的分层,调整插值参数改进绘制质量^[11];阳文锐等采用地统计学方法中的反距离加权空间插值法,采用风险评价的反距离加权插值法确定污染场地的污染范围^[12]。

本文对 IDW 插值进行改进,提出了时空反距离加权插值方法(TIDW),TIDW 是在反距离加权插值法的基础上,考虑时间变化对插值对象的影响,利用扩展法将时间作为空间插值的第三维计入反距离插值公式。同时在时空反距离插值的基础上顾及高程因素的影响,研究一种顾及高程的时空反距离加权插值方法(H-TIDW)。

1 研究方法

1.1 反距离加权插值

反距离加权插值法(Inverse distance weighting IDW)是最常用的空间插值方法之一,由美国国家气象局提出,是以待插点与实际观测样本点之间的距离为权重的插值方法,离待插点越近的样本点赋予的权重越大,其权重贡献与距离成反比。计算公式为:

$$Z = \sum_{i=1}^m \lambda_i Z_i \quad (1)$$

$$\lambda_i = \frac{d_i^{-n}}{\sum_{i=1}^m d_i^{-n}} \quad (2)$$

其中

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2} \quad (3)$$

式中, Z 为待插点的估算值; Z_i 为第 i 个样本点的实测值; d_i 为第 i 个样本点与待插点之间的距离; m 为参与计算的实测样本点个数; n 为幂指数,它控制着权重系数随待插点与样本点之间距离的增加而下降的程度, n 越大时,较近的样本点赋予更高的权重, n 越小时,权重更均匀地分配给各样本点。

当 $n=1$ 时,称为距离反比法,是一种常用而简便的空间插值方法。当 $n=2$ 时,称为距离平方反比法,是实际应用中经常使用的方法。当 n 取值很大,接近于正无穷时,待插点的估算值等于离待插点最近的样本点的值,该方法退化为泰森多边形法。当 $n=0$ 时,所有参与计算的样本点权重相等,均为 $1/m$,该方法退化为算术平均值法。

1.2 时空反距离加权插值

时空插值方法从约减法和扩展法两种思路来考虑不规则数据集的时空特征,约减法把时间作为独立的维单

独考虑,先进行时间插值,再进行空间插值。扩展法把时间作为等同于空间的第三维考虑。本文采用扩展法思想,将反距离加权插值法扩展为时空反距离加权插值法(TIDW),计算公式如下所示:

$$Z = \sum_{i=1}^m \lambda_i Z_i(t_i) \quad (4)$$

$$\lambda_i = \frac{d_i^{-n}}{\sum_{i=1}^m d_i^{-n}} \quad (5)$$

$$d_i = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (t_i - t)^2} \quad (6)$$

式中: $Z_i(t_i)$ 为第 i 个气象站点在 t_i 时刻的实测值; t_i 为与 t 时刻最接近的时刻; d_i 为第 i 个气象站点与待插点之间的时空距离。

1.3 顾及高程的时空反距离加权插值

顾及高程的时空反距离加权法(H-TIDW)在时空反距离加权法的基础上考虑了高程修正,其权重建立在两点间高程差与距离比值的基础上,其插值模型公式为:

$$Z = \sum_{i=1}^m \lambda_i Z_i(t_i) \quad (7)$$

$$\lambda_i = \frac{\left(\frac{\Delta E_i}{d_i}\right)^n}{\sum_{i=1}^m \left(\frac{\Delta E_i}{d_i}\right)^n} \quad (8)$$

式中: ΔE_i 为参证站 i 与待估点的高程差; n 为权重指数,能显著影响内插的结果,其取值考虑最小平均绝对误差,一般取值越大,插值结果越平滑。

主要流程如下:

- 1) 检验并去除数据季节性波动,本文采用时间序列分解法来去除数据的季节性波动。
- 2) 对去除季节性波动后的数据进行数据预处理,检验是否满足空间插值的平稳假设,对不满足的数据进行变换使其满足前提假设。
- 3) 使用 H-TIDW 模型进行 $PM_{2.5}$ 浓度插值分析。
- 4) 将预测结果与第一步去除的季节项叠加得到最终的 $PM_{2.5}$ 浓度插值结果,并与其他模型对比,采用 MAE、RMSE 作为评价指标来验证本文方法的插值精度。

2 插值试验

2.1 研究区概况

本文以北京地区为研究区域。北京市中心位于北纬 39° ,东经 116° 。全市土地面积 $16\,410\text{ km}^2$,其中平原面积 $6\,338\text{ km}^2$,占 38.6%,山区面积 $10\,072\text{ km}^2$,占 61.4%。全市总人口为 2 092.7 万。近年来,以 $PM_{2.5}$ 和 PM_{10} 为主的大气颗粒物浓度急剧升高,导致北京的雾霾天频发,空气污染严重,严重制约生态文明建设。对 $PM_{2.5}$ 浓度开展监测研究,掌握 $PM_{2.5}$ 空间变化状况,对掌控空间质量分布状况、采取相关防控举措具有一定意义。

2.2 数据来源

本文所采用的数据抓取自网站 www.pm25.in 北京监测站点数据,该数据主要包括 $PM_{2.5}$ 、 PM_{10} 、 CO 、 NO_2 、 SO_2 、 O_3 等,与地理空间数据云网站中的高程数据,监测站点数据包括北京市各个监测站点(35 个)从 2014 年 4 月到 2017 年 3 月的数据,35 个监测站点遍布北京城区及其郊区,其覆盖范围基本可以反映整个北京地区的空气质量状况。

2.3 数据预处理

将北京市各个监测站点(35 个)从 2014 年 4 月到 2017 年 3 月整 3 年的数据,求得月均值后,以月份为单位绘图,结果如图 1 所示。

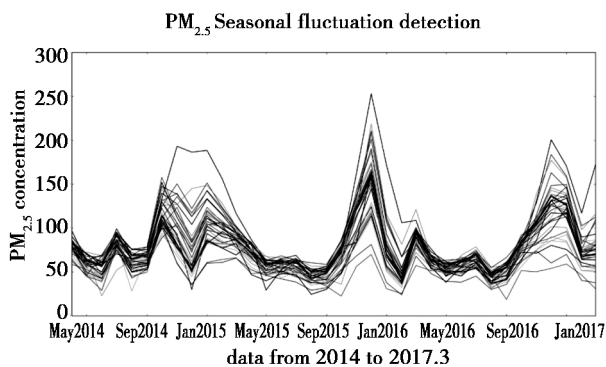


图 1 原始数据图
Fig.1 Raw data

从图 1 中可以看出各个监测站点的数据具有明显的季节性波动。根据时间序列分解法将原始数据分解为季节项、趋势项和随机项,如图 2 所示(图 2 为北京市的一个监测站点,以通州为例)。图中,raw 为原始数据,seasonal 为季节项,rend 为趋势项,residua 为随机项。本文将分解的结果保留趋势项和随机项,去除季节项,去除季节性波动后的数据与未处理的原始数据相比有一个最重要的特点,就是可以及时反映经济或物理现象的瞬间变化,反映序列变化的转折点,为人们制定决策提供比较科学的依据。

空间插值方法建立在平稳假设的基础上,如果数据不满足平稳假设,可通过数据变换使其满足前提假设条件。对去除季节性波动之后的数据进行空间数据探索分析来寻找数据内在的规律性,确定数据是否适合使用空间插值方法。由直方图分析得出峰度为 3.19,偏态为 -0.63,基本满足正态分布假设。其中,峰度(Kurtosis)用来描述数据分布的高度,标准正态分布的峰度应为 3,峰度值越接近 3 表示该数据越接近正态分布;偏态(Skewness)用来描述数据左右的对称性,标准正态分布的偏态值应为 0,如果偏态值大于 0,称正偏态或右偏态,此时大部分数据集中于左边,如果偏态值小于 0,称负偏态或左偏态,此时大部分数据集中于右边。北京地区 $PM_{2.5}$ 浓度 Normal QPlot 图,如图 3 所示。

通过对监测站点的 $PM_{2.5}$ 浓度值进行直方图和正态

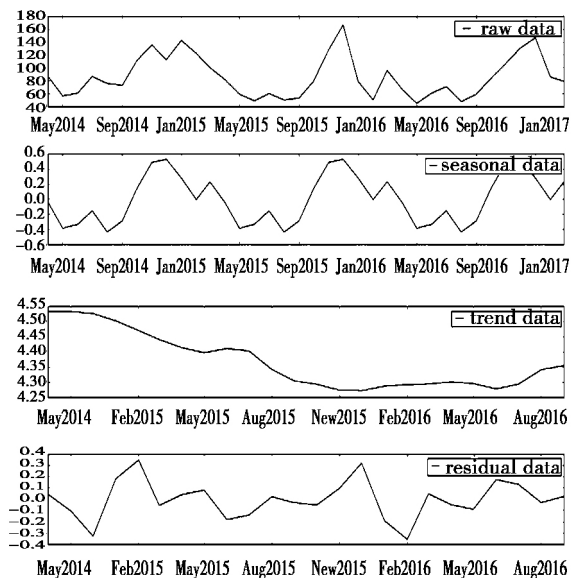


图 2 数据分解图
Fig.2 Data exploded

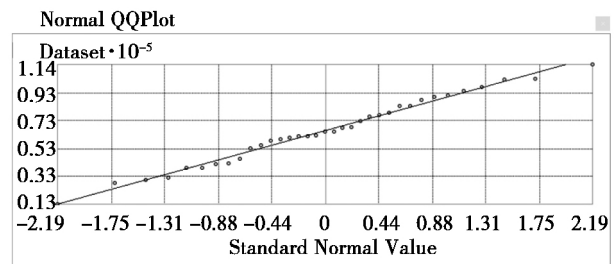


图 3 1 月 $PM_{2.5}$ 数据 QQPlot 图
Fig.3 January $PM_{2.5}$ data QQPlot

QQPlot 图分析得出,数据很接近正态分布且数据的相关性较高,说明该数据进行空间插值有意义。

3 插值结果与分析

本文利用北京市 2014 年 4 月到 2017 年 3 月的 $PM_{2.5}$ 月均浓度数据,然后分别利用反距离加权插值(IDW)、时空反距离加权插值(TIDW)和顾及高程的时空反距离加权插值(H-TIDW)方法进行插值,最后在插值结果基础上叠加之前去除的季节项得到最终的插值结果。为了评估本文研究方法的插值效果,本文将其与反距离加权插值法进行对比,通过交叉验证计算各个方法的平均绝对误差(MAE)、均方根误差(RMSE),并将其作为插值结果的评价指标来对各个方法的插值结果进行评估。其中平均绝对误差反映估计值可能的误差范围,均方根误差反映插值函数的反演灵敏度和极值效应,这两个指标越小代表精度越高。经计算,得到两种插值方法的插值精度与插值精度提升百分比见表 1、表 2(H-TIDW 代表顾及高程因素的时空反距离加权插值法,TIDW 代表时空反距离插值法)。

表 1 插值结果精度分析

Tab.1 Accuracy analysis of interpolation results (unit: $\mu\text{g}/\text{m}^3$)

	MAE	RMSE
IDW	2.364 5	2.670 3
TIDW	2.228 1	2.426 0
H-TIDW	2.219 1	2.423 6

表 2 插值精度提升百分比

Tab.2 Interpolation percentage accuracy (unit: %)

	MAE	RMSE
TIDW - IDW	5.77	9.15
H-TIDW - IDW	6.15	9.24

由表 1、表 2 可知,时空反距离加权插值方法(TIDW)相对于常用的反距离加权插值方法(IDW)精度有所提高,其中,TIDW 方法相对于 IDW 方法的平均绝对误差(MAE)精度提升程度在 5.77%左右,均方根误差(RMSE)精度提升程度在 9.15%左右,误差结果说明了在空间反距离加权插值的基础上考虑第三维时间特征,研究具有明显的时空性质的时空反距离插值方法,插值精度相对高于普通反距离插值。同时,顾及高程的时空反距离插值(H-TIDW)方法相对于反距离加权插值方法(IDW)精度有所提升,其中 H-TIDW 方法相对于 IDW 方法的平均绝对误差(MAE)精度提升在 6.15%左右,均方根误差(RMSE)精度提升在 9.24%左右,顾及高程的时空反距离插值(H-TIDW)方法精度提升相对 TIDW 不明显,主要原因是北京地区的地势总体来说相对平稳,对于考虑高程的插值方法应用于山区和山区、平原混合地形区插值效果明显,实验表明在北京考虑高程的时空反距离加权插值法没有明显提高插值精度。

4 结束语

顾及高程的时空反距离加权插值法(H-TIDW)与时空反距离加权插值法(TIDW)相对于反距离加权插值法(IDW)精度有一定程度的提升,说明该方法不仅可以有效地将时空信息考虑在内,减少了原始数据信息损失,还综合了高程因素对主变量的影响,提高了插值精度。

1) 通过交叉验证试验,说明本文提出的顾及高程的 H-TIDW 与 TIDW 模型预测的精度较高,经过对比本文方

法,MAE 与 RMSE 均低于 IDW 模型,其中 H-TIDW 模型的 MAE、RMSE 分别提升 6.15%、9.24%;TIDW 模型的 MAE、RMSE 分别提升 5.77%、9.15%。

2) 通过对比 H-TIDW、TIDW 与 IDW 模型的评价指标,也可看出,考虑时间因素的 H-TIDW、TIDW 模型要优于 IDW 模型。

本文由于数据限制只考虑了高程因素,还有很多因素没有考虑(如风力、降水等),但也说明该方法的优点,且在高原、山地等地区更能体现出本文方法的优点。

参考文献:

- [1] 彭思岭,邓敏.气象要素时空插值方法研究[D].长沙:中南大学,2010.
- [2] 张靖. GIS 空间内插方法与应用研究[D].呼和浩特:内蒙古师范大学,2008.
- [3] 刘文岭,李伟,刘洋.空间插值法对渤海天津海域海水盐度分布的影响[J].盐业与化工,2009,39(2):44-45.
- [4] 林忠辉,莫兴国,李宏轩,等.中国陆地区域气象要素的空间插值[J].地理学报,2002,57(1):47-56.
- [5] 何立恒,鲍其胜,王庆,等.基于反距离夹角加权算法的地理信息空间内插方法[J].测绘通报,2014(1):33-36.
- [6] 李正泉,吴尧祥.顾及方向遮蔽性的反距离权重插值法[J].测绘学报,2015,44(1):91-98.
- [7] 秦涛,付宗堂. ArcGIS 中几种空间内插方法的比较[J].物探化探计算技术,2007,29(1):72-75.
- [8] 李章林,王平,张夏林.距离幂次反比法的改进与应用金属矿山[J].金属矿山,2008(4):88-92.
- [9] 段平,盛业华,李佳,等.自适应的 IDW 插值方法及其在气温场中的应用[J].地理研究,2014(8):1417-1426.
- [10] 李佳霖,樊子德,邓敏.顾及风向和风速的空气污染物浓度插值方法[J].地球信息科学学报,2017,19(3):382-389.
- [11] 高真,叶学义,周天琪,等.基于反距离加权插值的水声数据可视化算法[J].计算机工程,2015,41(9):266-270.
- [12] 阳文锐,王如松,黄锦楼,等.反距离加权插值法在污染场地评价中的应用[J].应用生态学报,2007,18(9):2013-2018.

[编辑:张 曦]

(上接第 81 页)

2) 对相应 Web 地图算法进行深入研究,给出了其瓦片定位信息代码,使得每张瓦片都具有坐标定位信息文件(*.tfw)。

3) 自主开发了相应软件,利用下载的高分辨率卫星影像,通过投影变换、纠正等工序,制作了满足 1:5 000 和 1:10 000 地形图工作底图,应用于测绘生产。

参考文献:

- [1] 许辉,马晓鹏.基于 Web 墨卡托投影地理信息系统设计与实现[J].电脑编程技巧与维护,2011(8):41-44.
- [2] 程庆.基于配准技术的非合作卫星遥感图像精确定位[D].重庆:重庆大学,2015.

- [3] 崔金红.Google 地图算法研究及实现[J].计算机科学,2007,34(11):193-195.
- [4] 寇曼曼,王勤忠,谭同德.Google Map 数字栅格地图算法及应用[J].计算机技术与发展,2012,22(4):204-206.
- [5] 刘亚东,李青元,谭海,等.开源库 GDAL 及其在影像拼接中的应用[J].数字技术与应用,2010(2):88-89.
- [6] 张世富.基于 Google Earth 的高分辨率影像获取方法探讨[J].电子技术,2013(12):16-18.
- [7] 方昆,邱小波,金宏斌.基于 Google Earth 影像地图数据获取的研究[J].信息技术,2014,38(4):96-99.

[编辑:张 曦]