



武汉大学学报(信息科学版)

Geomatics and Information Science of Wuhan University

ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目: 顾及空间异质性的自适应 IDW 插值算法
作者: 颜金彪, 段晓旗, 郑文武, 刘媛, 邓运员, 胡最
DOI: 10.13203/j.whugis20180213
收稿日期: 2018-05-30
网络首发日期: 2019-09-05
引用格式: 颜金彪, 段晓旗, 郑文武, 刘媛, 邓运员, 胡最. 顾及空间异质性的自适应 IDW 插值算法. 武汉大学学报(信息科学版).
<https://doi.org/10.13203/j.whugis20180213>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: 10.13203/j.whugis20180213

顾及空间异质性的自适应 IDW 插值算法

颜金彪^{1,3} 段晓旗² 郑文武^{1,3} 刘媛^{1,3} 邓运员^{1,3} 胡最^{1,3}

1 衡阳师范学院 传统村镇文化数字化保护与创意利用技术国家地方联合工程实验室, 湖南, 衡阳, 421002

2 武汉大学 测绘遥感信息工程国家重点实验室, 湖北 武汉, 430079

3 衡阳师范学院 古村古镇文化遗产数字化传承湖南省协同创新中心, 湖南, 衡阳, 421002

摘要: 针对传统的反距离加权 (Inverse Distance Weighted, IDW) 插值法存在以假设空间过程平稳为前提, 插值过程需要用户提供初始化参数, 如分区数目、参考样点个数等, 本文提出了一种顾及空间异质性的自适应 IDW 插值算法。该算法首先根据样点待插值属性数据的数理统计特征自适应设定分类阈值, 将全部样点按照属性值大小分成高、中、低三类, 接着利用机器学习算法 k 近邻法 (k -nearest neighbor, k -NN) 判定待插值点所属类别, 进而根据分类结果自适应为待插值点一阶邻近样点设置相应的权重调和因子, 最终构造出一个集空间相关与空间异质于一体的 IDW 插值算法模型。实验结果表明, 本文提出的顾及空间异质性算法无需用户提供任何经验参数, 其抗粗差的能力较另三种经典 IDW 插值算法更强, 能够有效提高 IDW 插值算法的精准度。

关键词: k 近邻法; 分类; 调和因子; 自适应; 空间异质性

中图分类号: P208

文献标识码: A

空间插值是通过有限的离散采样点来建立某种插值函数关系 $f(x)$, 并将已知采样点范围内的任意位置代入函数关系式计算未知点属性值, 常用的空间插值方法有反距离加权 (Inverse Distance Weighted, IDW)、克里金 (Kriging)、趋势面、样条函数、自然邻域法。其中, IDW 以其插值原理简单、计算简便且符合地理学第一定律而被广泛使用^[1]。

近年来关于 IDW 算法研究较多, 主要集中在幂指数^[2-3]、搜索点数^[4]、搜索方向^[5]三个方面。通过分析现有的 IDW 插值方法, 发现大多 IDW 插值算法均假设空间过程的平稳性^[6], 而由于地表空间异质性的普适存在^[7], 现有 IDW 及其改进算法的精度仍然不够令人满意。基于此, 李佳霖与樊子德分别在 2015 与 2016 年提出基于空间分层异质性的 IDW 插值方法, 算法其中一个核心步骤为根据采样点的位置和属性进行分区。此类方法能够使得插值精度明显提高, 但也存在一些问题: 1) 分层数目的确定需要先验知识, 而且通常会比较困难; 2) 位于异质分层边界区域的插值精度会下降, 尤其是点数较少的层内; 3) 在空间异质分层后的子区域内部并不能消除空间异质性的影响^[8]。

针对现有基于空间分层异质性的 IDW 插值算法存在的问题, 本文将机器学习的监督分类思想应用于 IDW 插值中待插值点的空间局部异质性判别, 提出一种新的 k 近邻反距离加权 (k -nearest neighbor adaptive IDW, KAIDW) 插值算法。该算法首先根据样点属性数据的数理统计特征对样点进行自动分类, 随后利用机器学习中的 k 近邻法在未得到插值结果的前提下判定待插值点的所属类别, 从而自适应的为周边一阶邻近样点设置不同的权重调和因子。该算法在无需人为干预的情况下能同时顾及空间相关与空间异质对插值结果的影响。

1 IDW与 k -NN算法

1.1 IDW插值算法

IDW插值算法以其简便、高效等诸多优点被广泛应用于各大GIS软件中^[2], 如ArcGIS、MapInfo、MapGIS、SuperMap等, 计算模型如下式 (1) 所示:

收稿日期: 2018-05-30

项目资助: 国家社会科学重大项目(16ZDA159)

第一作者: 颜金彪, 博士生, 主要从事时空数据挖掘理论与应用研究。715829216@qq.com

通讯作者: 郑文武, 博士, 教授。13766108@qq.com

$$z(x) = \frac{\sum_{i=1}^m (1/d_i^n \cdot z_i)}{\sum_{i=1}^m 1/d_i^n} \quad (1)$$

其中 $z(x)$ 与 z_i 为待插值点与样本点的属性值， d_i 为样本点与待插值点之间的欧氏距离， n 为幂指数，一般取 1~3^[9]，但多数学者认为 n 为 2 能获得更好的实验结果^[10]， m 为选择的样点个数。

1.2 k -NN分类算法

k -NN 算法以其简单，易于实现，无需估计参数，作为数据挖掘的经典算法而被广泛使用^[11-12]。经典的 k -NN 算法^[13]选取测试样本周边的 k 个近邻对象，应用“投票”规则确定测试样本的类别，即待判定点在邻近空间中的 k 个最相似的样本中多数属于同一类别，则该点将被预测为此类别，如此可实现算法误分类的经验风险最小化^[14]。

本文中 k 值为待插值点一阶邻近点数，从而使得选取的 k 个样点既考虑就近原则，同时又避免由于样点数据发生聚集现象而导致的分类结果严重依赖于某个方向样点的问题。不仅如此，由于 k 值的选取不需要训练，因此能大幅提高算法效率。

2 KAIDW算法

2.1 基本原理

经典IDW算法以假设空间过程的平稳性为前提，通常采用距离平方的反比进行定权，但忽略了待插值点与样点之间空间异质性的存在。本文提出的KAIDW算法以地理学第一定律为核心思想^[15]，同时通过机器学习算法 k -NN判定参考样点与待插值点之间的空间异质性，最终构造出一个集空间相关与异质性于一体的插值模型，降低异质性参考样点的权重，从而提高插值的精度。

2.2 算法的设计与分析

2.2.1 算法流程

KAIDW插值算法主要包括分类，定权，插值三个步骤，详细流程如图（1）所示。

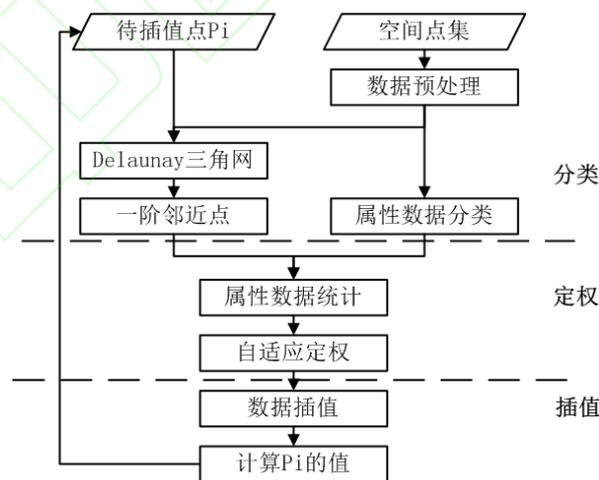


图 1 KAIDW算法流程图

Fig. 1 The flow chart of KAIDW

2.2.2 算法实现

1) 分类

k -NN 算法确定待插值点属性类别的前提为获得样点属性的类别，该过程的主要任务是

将样点连续的属性数据离散化, 获取样点待插值属性数据离散化后的分类结果, 主要过程如下所述:

(1) 数据预处理。由于空间点可能存在压盖的情况, 影响后续狄洛尼三角网 (Delaunay triangulation) 的生成, 因此将压盖位置的空间点保留其中一个。

(2) 样点分类。根据预处理得到的样点, 计算得到全部样点属性值的均值 Z_{mean} 与中误差 Z_{std} , 从而可以获取每个样点属性值分类的阈值—— H_i , L_i , 如下式 (2、3) 所示:

$$H_i = Z_{mean} + Z_{mean} / z_i \times Z_{std} \quad (2)$$

$$L_i = Z_{mean} - Z_{mean} / z_i \times Z_{std} \quad (3)$$

从式 (2、3) 中可以看出, H_i 与 L_i 具备自适应特征, 对于样点中较大的属性值, 则易被式 (2) 检测, 相反属性值愈低则易被式 (3) 发现。通过如上方法, 可以将全部样点按照属性值分成三类值, 如下式 (4) 所示:

$$t_i = \begin{cases} \text{High}, z_i > H_i \\ \text{Middle}, L_i \leq z_i \leq H_i \\ \text{Low}, z_i < L_i \end{cases} \quad (4)$$

(3) 一阶邻近样点的选择。Delaunay三角网广泛应用于空间邻近分析, 如空间数据聚类、空间邻近关系^[16], 故本文应用Delaunay三角网获取待插值点周边一阶邻近点^[1], 使得参考点较为均匀的分布在待插值点的周围。

算法的具体步骤如下:

(1) 移除原始样点中压盖的空间点; (2) 计算全部样点属性值的均值 Z_{mean} 与中误差 Z_{std} ; (3) 选取样点中任意点 P_i ; (4) 计算 P_i 对应属性值 Z_i 的 H_i 与 L_i ; (5) 如果 $Z_i > H_i$, P_i 的类别赋为 High; (6) 如果 $Z_i < L_i$, P_i 的类别赋为 Low; (7) 如果 $L_i \leq Z_i \leq H_i$, P_i 的类别赋为 Middle; (8) 重复步骤 3-7, 直至所有样点均分类完成。

通过上述分类步骤, 如后文实验 2 的降雨数值离散化得到如下图 (2) 所示的类别数据。从图 (2) 中可以发现, 该地区降雨量主要可以分成 3 大区域, 经过地理探测器^[17]软件探测得到 q 统计 (q-statistic) 值为 0.65 且通过显著性水平为 0.05 的分层异质性假设检验, 说明该地区降雨存在较明显的空间分层异质性^[18], 同时由图可以发现该区存在空间局部异质性。因此, 需要算法能够针对层 (stratum) 内与层间做出不同的定权处理, 降低异质性样点数据对插值结果造成的影响。

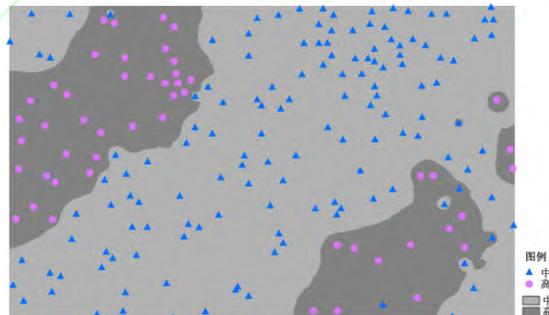


图 2 属性数据分类结果

Fig. 2 The classification result of attribute

2) 定权

常规的 IDW 插值算法一般采用距离倒数的平方定权, 本文中不仅考虑参考样点与待插

值点之间的空间自相关，同时顾及一阶邻近样点与待插值点之间的空间异质性。

该过程的主要任务在反距离平方定权的基础上自适应确定权重调和因子，使得最终的权重不仅考虑到样点与待插值点之间的空间相关性，同时顾及空间异质性，算法过程如下：

(1) 获取待插值点一阶邻近点的分类结果。统计一阶邻近点属性值分类为 High、Middle、Low 各自的数量 k_1 、 k_2 、 k_3 ；

(2) 定权。根据待插值点一阶邻近点中 High、Middle、Low 三种类别点的数量，利用 k -NN 算法“少数服从多数的投票法则”确定待插值点的属性类别。

由于 k -NN 算法得到待插值点的预测值可能与真值不一致，如下图 (3) 所示，未知点实为“低”类（预测值为“高”类），如此将造成参考样点与待插值点异质性的误判。为了削弱误分类对插值结果造成的偏差，KAIDW 算法将三类点各自权重调和因子 a_i 设置成与各类别样点数量成正比例关系，而非将异质参考样点的权重设置为 0，如式 (5) 所示。其中 W 为一阶邻近点中的总参考样点数， d_i 表示待插值点与参考样点之间的欧氏距离，“ C ”为计数符号， t_i 为 High, Middle, Low 其中一类别。

$$a_i = C(t_i) / C(W) / d_i^2 \quad (5)$$

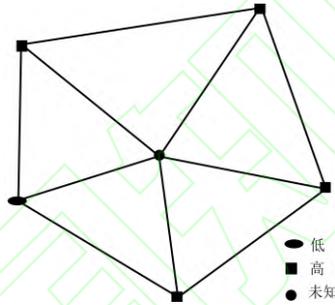


图 3 k -NN 误分类示意图

Figure 3 The misclassification of k -NN

从式 (5) 中可以发现，权重调和因子 a_i 的确定同样是自适应的，无需人工干预，在空间相关性的基础上顾及了空间异质性对插值结果的影响。

3) 插值

(1) 模型表达式

针对参考样点中可能存在空间异质性数据，特别是距离待插值点较近的空间异质点，将使得常规 IDW 插值算法得到的结果将会偏离真值较大。基于此，本文在反距离平方定权的基础上，提出了一种顾及空间异质性的 IDW 插值模型，如下式 (6) 所示。

$$z(x) = \frac{\sum_{i=1}^{k_1} \frac{k_1}{d_i^n} \cdot z_i + \sum_{j=1}^{k_2} \frac{k_2}{d_j^n} \cdot z_j + \sum_{k=1}^{k_3} \frac{k_3}{d_k^n} \cdot z_k}{\sum_{i=1}^{k_1} \frac{k_1}{d_i^n} + \sum_{j=1}^{k_2} \frac{k_2}{d_j^n} + \sum_{k=1}^{k_3} \frac{k_3}{d_k^n}} \quad (6)$$

式中， k_1 、 k_2 、 k_3 分别为待插值点一阶邻近中 High、Middle、Low 三类点对应的个数，三类点数量之和为一阶邻近点数总和 W ； Z_i 、 Z_j 、 Z_k 分别为 High、Middle、Low 三类点各自对应的属性值； d_i 、 d_j 、 d_k 为其对应的欧氏距离， $z(x)$ 为 KAIDW 算法计算得到的待插值点属性值。

(2) 算法特点分析

由于 KAIDW 插值模型不仅考虑了待插值点与样点之间的空间相关性,同时顾及了空间异质性的影响,因此能够得到更高的插值精度,详情如下分析。

首先将式(1)按照式(6)的形式改写得到式(7),其中 k_1 、 k_2 、 k_3 分别代表常规 IDW 插值算法的一阶邻近点中被本文分类算法分类为“High”、“Middle”、“Low”的样本点数量,而 z_i 、 z_j 、 z_k 分别为“High”、“Middle”、“Low”类点对应的属性值, $z'(x)$ 为常规 IDW 插值算法得到的待插值点属性值。

$$z'(x) = \frac{\sum_{i=1}^{k_1} \frac{1}{d_i^n} \cdot z_i + \sum_{j=1}^{k_2} \frac{1}{d_j^n} \cdot z_j + \sum_{k=1}^{k_3} \frac{1}{d_k^n} \cdot z_k}{\sum_{i=1}^{k_1} \frac{1}{d_i^n} + \sum_{j=1}^{k_2} \frac{1}{d_j^n} + \sum_{k=1}^{k_3} \frac{1}{d_k^n}} \quad (7)$$

①设 $k_1 > k_2 > k_3 \neq 0, d_i < d_j < d_k$, 根据 k -NN 算法的“少数服从多数”的原则,那么待插值点的类别即为 High 类,将式(6)分子与分母同时除以 k_1 ,得到 KAIDW 插值模型的变形式 $z(x)$,如式(8)所示。

$$z(x) = \frac{\sum_{i=1}^{k_1} \frac{1}{d_i^n} \cdot z_i + \sum_{j=1}^{k_2} \frac{1}{d_j^n} \cdot z_j \cdot \frac{k_2}{k_1} + \sum_{k=1}^{k_3} \frac{1}{d_k^n} \cdot z_k \cdot \frac{k_3}{k_1}}{\sum_{i=1}^{k_1} \frac{1}{d_i^n} + \sum_{j=1}^{k_2} \frac{1}{d_j^n} \cdot \frac{k_2}{k_1} + \sum_{k=1}^{k_3} \frac{1}{d_k^n} \cdot \frac{k_3}{k_1}} \quad (8)$$

对比式(7)与(8),可以明显看出式(8)中 High 类点的权重比式(7)中 High 类点的权重大。同时从式(8)可以看出,High 类点的权重比 Middle、Low 类点在等距离的前提下权重都要大。反之即使 $d_i > d_j > d_k$,由于 KAIDW 插值模型考虑了空间数据存在异质性的情况,通过调和因子 k_1 、 k_2 、 k_3 ,同样可以使得插值结果的精度要高于传统 IDW 插值算法,从而可以发现顾及空间异质性的 KAIDW 插值算法优于选择同样参考点的 IDW 插值算法,如传统的 IDW、自适应的反距离权重(Adaptive-IDW, AIDW)、四方向反距离权重(Four-IDW, FIDW)(见后文实验 1)等。

②设 k_1 、 k_2 、 k_3 中存在两类点的数量为 0,如上图(2)同一“层”间(区域,下同),即待插值点周边一阶邻近点中不存在空间异质性,那么式(6)将等价于 AIDW 插值算法,其插值精度仍高于常规的 IDW 插值算法^[1]。

③设 $k_1 = k_2 = k_3 \neq 0$,式(6)同样等价于 AIDW 插值算法,通过②、③可以发现,AIDW 插值算法在一定程度上属于 KAIDW 插值算法的一种特例。

④设 k_1 、 k_2 、 k_3 存在其中一种类别为 0,如上图(2)不同层间,其分析形式如上第一种情形,此处不再加以赘述。

通过如上分析可以发现: i) 针对样本空间点数据中存在空间异质性情形,KAIDW 插值算法在同一层间仅考虑空间相关性,在不同层间或局部存在空间异质性区域能够同时顾及空间异质与相关性对插值结果的影响; ii) KAIDW 插值精度并不是在任何插值点上都高于传统的 IDW 插值算法,KAIDW 插值算法的精度能否高于传统的关键在于 k -NN 算法对待插值点属性类别的分类是否准确,这也成为阻碍该 IDW 插值算法精度提高的关键因素,也是下一步重点研究的区域。

3 实验对比分析

本文设计两个不同应用领域的实验来验证 KAIWD 算法的可行性, 将 KAIWD 算法与 CIDW (经典 IDW)、FIDW (四方向 IDW)、AIDW 算法进行对比分析, 其中 CIDW 与 FIDW 参考点数量分别设置为 4~12 个, KAIWD 与 AIDW 待插值点的参考点数量为一阶邻近点, 4 种 IDW 插值算法幂指数均取值为 2, 检查点为除研究区域边界点以外的样点, 实验方案采用留一交叉验证法^[10], 选取平均绝对误差 (Mean Absolute Error, MAE)、均方根误差 (Root Mean Square Error, RMSE)、最大误差 (MAX)、最小误差 (MIN) 4 个指标对实验结果进行分析。

3.1 DEM 插值

本例中将上述 4 种算法应用于 DEM 的插值研究当中, 数据来源于山东黄河大堤的大比例尺地形图中, 研究区域呈现带状分布, 形成中高两低的阶梯, 通过野外采集了 308 个高程点, 平均高程为 42.90 m, 最大高程为 49.01 m, 最低处 36.30 m, 高程中误差为 3 m, 相对中误差为 7%, 实验结果如下图 4 所示。

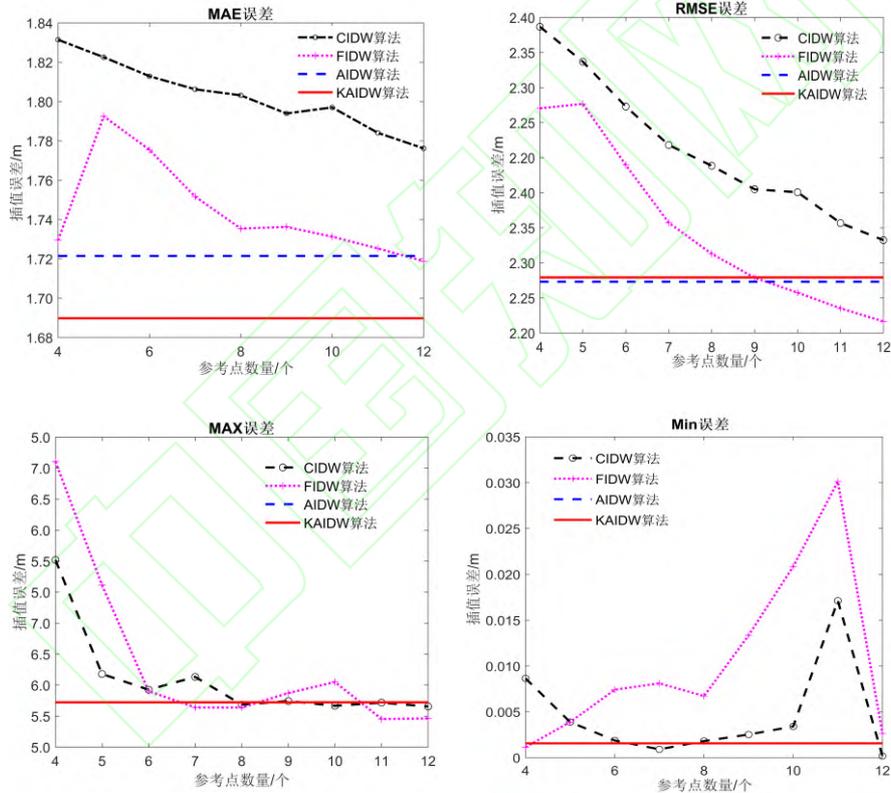


图 4 DEM 插值对比结果

Figure 4 The comparison result of DEM interpolation

从图 4 中 MAE 误差图可以发现: CIDW、FIDW 插值算法 MAE 误差基本都随着插值点数的增多而逐渐下降, 在选择 12 个采样点时达到最小, 分别为 1.776 m 与 1.718 m。AIDW 插值精度比 CIDW、FIDW 基本都要高, 为 1.721 m, 而顾及空间异质性的 KAIWD 插值算法的 MAE 误差在四种算法中最低, 为 1.689 m。

从图 4 中 RMSE 误差图可以发现: CIDW、FIDW 插值算法的 RMSE 随着插值点数的增多基本呈现对数方式下降, 其中 FIDW 下降幅度更快。在参考点数量达到 10 个以上时, FIDW 的 RMSE 为 4 种方法中最低。这是由于该区域高程值变化较为稳定, 高程相对误差仅为 7%, 待插值点周边区域中不存在变化较大的高程点数据, 而 FIDW 插值算法在 4 个方向的样点

数量较为均匀，其 RMSE 误差将随参考点数量增多而稳定下降直至稳定。AIDW 与 KAIDW 插值算法在 RMSE 方面相差较小，约为 2 mm。

从图 4 中 MAX 误差图可以发现：CIDW 与 FIDW 插值算法在参考点数量大于 7 时，MAX 值围绕 AIDW 与 KAIDW 呈现类正弦波动，在参考点达到 11 个以上时，FIDW 略优于其余 3 种插值方法，AIDW 与 KAIDW 插值算法得到的 MAX 值结果基本一致。经分析发现，KAIDW 的 MAX 误差发生的待插值点的真实高程通过图（1）所示的“分类”步骤结果为 High 类点，但该点周边一阶邻近点中 High 类点的权重调和因子仅占 1/6（即 5 个 Middle 类点，一个 High 类点），从而造成该点 MAX 误差相比于 FIDW 值要略大一些。

从图 4 中 MIN 误差图可以发现：CIDW 与 FIDW 的 MIN 值随着样点数的增多呈现震荡形式，且 CIDW 较 FIDW 小。AIDW 与 KAIDW 插值算法的 MIN 值保持一致，这是由于待插值点一阶邻近点中不存在异质点，KAIDW 与 AIDW 定权一致。KAIDW 与 CIDW 插值算法相比，当样点的数量为 7、12，其 MIN 值超过 CIDW 约 1mm，这是由于该插值点附近地形平坦，空间自相关性较强，当参考样点的个数达到合适的条件将使得 CIDW 法 MIN 较小。

从上述图(4)的对比分析来看：1) 不存在一种算法的插值精度在任何数据分布上 MAE、RMSE、MAX、MIN 的 4 个指标都占据绝对的优势；2) 顾及样点分布均匀性的 FIDW 插值算法精度优于 CIDW；3) AIDW 插值算法的精度通常要优于 FIDW；4) 待插值点一阶邻近点中不存在空间异质性，KAIDW 插值结果与 AIDW 算法保持一致，反之在待插值点属性类别分类准确的前提下，那么其插值精度将明显高于 AIDW。

3.2 降雨数据的插值

为进一步验证 KAIDW 插值算法的适用性，将其应用于气象数据的插值应用研究当中。数据采用 SIC97 数据（Spatial Interpolation Comparison exercise 1997, SIC97）中 192 个站点 1986 年 5 月 8 日降雨观测数据，数据来源于 [https://www.researchgate.net/profile/Gregoire_Dubois/p-ublication/281292076_Spatial_Interpolation_Comparison_97_\(SIC97\)_dataset](https://www.researchgate.net/profile/Gregoire_Dubois/p-ublication/281292076_Spatial_Interpolation_Comparison_97_(SIC97)_dataset)，192 个站点当天平均降雨量为 19.8 mm，中误差为 12.1 mm，相对中误差约为 61.1%，最大降雨量为 51.7 mm，最小降雨量为 1.7 mm，实验结果如下图 5 所示。

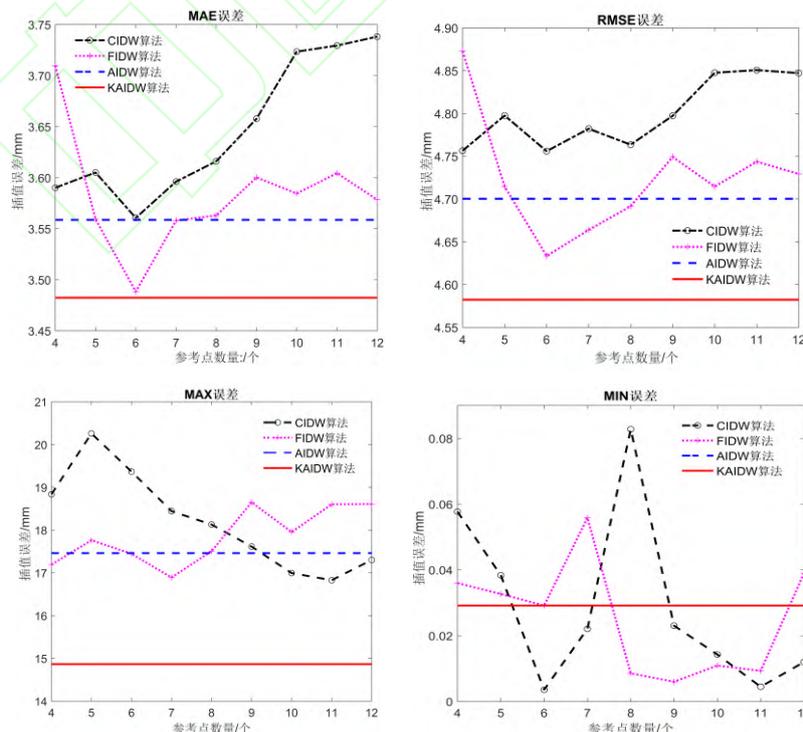


图5 降雨插值对比结果

Fig.5 The comparison result of precipitation

从图（5）中可以发现：KAIDW 插值算法在 4-12 个点中 MAE、RMSE、MAX 误差均完全优于 CIDW、FIDW、AIDW 三种插值算法，其中 MIN 误差 KAIDW 插值算法与 AIDW 保持一致，CIDW 与 FIDW 的 MIN 值围绕 KAIDW 上下波动，且 4 种插值方法在 MIN 上相差较小，极差为 0.79 mm。基于该实验结果的分析，可以发现 KAIDW 得到的实验结果优于其余 3 种插值算法。除此之外，通过对比 DEM 与降水实验的数据，瑞士 192 个站点的降水值相对中误差为 61.1%，而 DEM 实验高程数据的相对中误差仅为 7%，根据离散方差大其空间异质性强的原理^[19]，实验（2）数据的空间异质性更为明显，但 KAIDW 算法在实验（2）的实验结果优于实验（1），从而说明顾及空间异质性的 KAIDW 算法在一定程度上较另 3 种 IDW 插值算法的抗差能力更强。

4 结语

本文提出了一种顾及空间异质性的自适应 IDW 插值算法，该算法简便、易于程序实现，通过与 3 种 IDW 插值算法进行对比分析后发现：1）KAIDW 算法具备良好的自适应特征，用户不需指定任何初始化参数，降低用户使用 IDW 算法的知识储备要求；2）应用机器学习监督分类思想的 KAIDW 算法能够有效的对空间异质性数据进行处理，使得其抗大粗差能力更强，能够有效的提高插值精准度。

然而 KAIDW 算法并不是在任何插值点上都是有效的，算法的适用条件在于 k -NN 算法在插值场景中对待插值点属性类别分类的准确率，因此未来关于 KAIDW 算法的研究工作主要集中在 2 个方面：1）将多种机器学习算法引入到插值算法的研究当中，如 SVM（支持向量机），进一步提高待插点属性分类的准确率；2）根据可靠性更高的判别结果进一步加大同类样点间的权重，从而降低异质样点对最终插值结果的影响。

参 考 文 献

- [1] Duan Ping, Sheng Yehua, Li Jia, et al. Adaptive IDW Interpolation Method and Its Application in the Temperature Field [J]. *geographical research*, 2014,33(08):1417-1426(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,2014,33(8):1417-1426)
- [2] Lu G Y, Wong D W. An Adaptive Inverse Distance Weighting Spatial Interpolation Technique [J]. *Computers & Geosciences*, 2008,34(9):1 044-1 055
- [3] FAN Zide, Li Jialin, DENG Min. An Adaptive Inverse-distance Weighting Spatial Interpolation Method with the Consideration of Multiple Factors [J]. *Geomatics and information science of Wuhan University*, 2016,41(6):842-847(樊子德,李佳霖,邓敏.顾及多因素影响的自适应反距离加权插值方法[J]. 武汉大学学报·信息科学版,2016,41(06):842-847)
- [4] ZHANG Jinming, YOU Xiong, WAN Gang. Experimental Research on Optimization of DEM Interpolation Parameters [J]. *Acta Geodaetica et cartographica sinica*, 2014,43(02):178-185(张锦明,游雄,万刚. DEM 插值参数优选的试验研究[J]. 测绘学报,2014,43(02):178-185)
- [5] Růhaak W. A Java Application for Quality Weighted 3-d Interpolation [J]. *Computers & Geosciences*, 2006, 32(1):43-51
- [6] FAN Zide, GONG Jianya, LIU Bo, et al. A Space Time Interpolation Method of Missing Data based on Spatio-temporal Heterogeneity [J]. *Acta Geodaetica et cartographica sinica*, 2016,45(04):458-465(樊子德,龚健雅,刘博,等.顾及时空异质性的缺失数据时空插值方法[J]. 测绘学报,2016,45(04):458-465)
- [7] Wang J F, Christakos G, Hu M G. Modeling Spatial Means of Surfaces With Stratified Nonhomogeneity [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2009, 47(12):4167-4174

- [8] LI Jinlin,FAN Zide,DENG Min.Residual Inverse Distance Weighting Spatial Interpolation Method based on Spatial Heterogeneity Subregion[J].*Geography and geo-information science*, 2015, 31(5):25-29(李佳霖,樊子德,邓敏.基于空间异质分区的残差 IDW 插值方法[J]. 地理与地理信息科学,2015,31(5):25-29)
- [9] Lam N. Spatial Interpolation Methods: A Review [J]. *The American Cartographer*,1983, 10(2):129-150
- [10] DECLERCQ F A N. Interpolation Methods for Scattered Sample Data: Accuracy, Spatial Patterns, Processing Time [J]. *Cartographer and Geographic information systems*, 1996, 23(3):128-144
- [11] XIONG Yajun,LIAO Xiaonong,LI Ziming,et al.Application of KNN Data Mining Algorithm to Haze Grade Forecasting in Beijing[J].*meteorological monthly*, 2015,41(01):98-104(熊亚军,廖晓农,李梓铭,等. KNN 数据挖掘算法在北京地区霾等级预报中的应用[J].气象,2015,41(01):98-104)
- [12] LI Yang,FANG Binxing,GUO Li, et al.Supervised Intrusion Detection based on Active Learning and TCM KNN Algorithm[J].*Chinese journal of computers*,2007(08):1464-1473(李洋,方滨兴,郭莉,等. 基于主动学习和 TCM-KNN 方法的有指导入侵检测技术[J].计算机学报,2007(08):1464-1473)
- [13] Cover T, Hart P. Nearest Neighbor Pattern Classification [J]. *IEEE Trans.inf.theory*, 1967, 13(1):21-27
- [14] LI Hang.Statistical Learning Method [M].*Beijing:Tsinghua University Press*,2012:40(李航,统计学习方法[M].北京:清华大学出版社, 2012:40)
- [15] Tobler W R.A Computer Movie Simulating Urban Growth in the Detroit Region [J]. *Economic Geography*, 1970, 46:234-240
- [16] YANG Wei, Ai Tinghua. The Extraction of Road Boundary from Crowdsourcing Trajectory Using Constrained Delaunay Triangulation[J]. *Acta Geodaetica et cartographica sinica*, 2017,46(02):237-245(杨伟,艾廷华. 运用约束 Delaunay 三角网从众源轨迹线提取道路边界[J]. 测绘学报,2017,46(02):237-245)
- [17] Wang J F, Zhang T L, Fu B J. A Measure of Spatial Stratified Heterogeneity [J]. *Ecological Indicators*, 2016, 67:250-256
- [18] Wang J F, Haining R, Liu T J, et al. Sandwich Estimation for Multi-unit Reporting on a Stratified Heterogeneous Surface [J]. *Environment & Planning A*, 2013, 45(10):2515-2534
- [19] WANG J F,LIAO Y L,LIU Z.The Tutorial of the Spatial Data Analysis[M].Beijing: Science Press,2010:60(王劲峰,廖一兰,刘鑫.空间数据分析教程[M].北京:科学出版社,2010:60)

An Adaptive IDW Algorithm Involving Spatial Heterogeneity

YAN Jinbiao^{1,3} DUAN Xiaoqi² ZHENG Wenwu^{1,3} LIU Yuan^{1,3} DENG Yunyuan^{1,3} HU Zui^{1,3}

¹ National-Local Joint Engineering Laboratory on Digital Preservation and Innovative Technologies for the Culture of Traditional Villages and Towns, Hengyang normal university, Hengyang 421002,China

² State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

³ Cooperative Innovation Center for Digitalization of Cultural Heritage in Traditional Villages and Towns, Hengyang normal university, Hengyang 421002, China

Abstract: In this paper, we propose an adaptive IDW algorithm involving spatial heterogeneity to solve the problems existed in the classical IDW. The first problem is that classical IDW algorithms are heavily dependent on the spatial stability. Another one is that the initial parameters are determined by the users empirically, such as the number of stratum or sample points. The algorithm (KAIDW) put forward in this article can take both spatial correlation and heterogeneity into account simultaneously without the needs of parameters input for users. The details are as follows. Firstly, KAIDW sets the classification threshold adaptively for each sample point according to the statistical characteristics of the sample data and then divides the reference points into high, medium and low categories. Secondly, the k-nearest neighbor algorithm is used to

determine the category of the interpolation point. According to the classification result, different weight adjustment coefficients are adaptively determined for the first-order neighboring samples of the point to be interpolated. Finally, an IDW interpolation algorithm model integrating spatial correlation and heterogeneity is constructed. In order to validate the effectiveness of the algorithm, two different practical applications are adopted. By comparing with 3 classical IDW algorithms, we find that the KAIDW can effectively improve the accuracy of the IDW interpolation algorithm without the user providing any empirical parameters.

Key words: k-nearest neighbor; classification; adjustment factor; adaptive; spatial heterogeneity

First Author: YAN Jinbiao, doctoral candidates, specializes in the theory and application of spatial-temporal data mining. Email: 715829216@qq.com

Corresponding author: ZHENG Wenwu, Ph D, professor. E-mail: 13766108@qq.com

Foundation Support: Major program of National Social Science Foundation of China (No. 16ZDA159)

