

熊敏詮. 2013. 滑动窗口的普通克立格方法在降水量插值中的应用. 气象, 39(4):486-493.

滑动窗口的普通克立格方法在降水量插值中的应用^{*}

熊敏詮^{1,2,3}

- 1 中国科学院大气物理研究所, 北京 100029
2 中国科学院大学, 北京 100049
3 国家气象中心, 北京 100081

提 要: 利用 2009 年全国 2200 个观测站降水量资料, 使用滑动窗口的普通克立格方法对降水量资料进行格点化估计。针对滑动窗口的普通克立格方法在降水量格点化应用中存在的问题, 设计了 3 种技术处理的试验方案。比较了全局搜索与方位邻近方法的误差, 讨论了最大影响半径及屏蔽效应对插值效果的作用。提出了方位邻近法的样本点选择策略, 结果表明, 相对传统滑动窗口的普通克立格方法较常使用的全局搜索法而言, 方位邻近法显著降低了计算资源的耗用, 同时又具有较高的插值精度, 特别是在站点密集地区有突出的优势; 试验结果也表明: 变程为 $4^{\circ}\sim 5^{\circ}$ 的经(纬)线弧长时, 在方位邻近法下, 我国大部分区域有较好的插值效果; 屏蔽效应弱, ϵ 取值为 0.1 时, 降水量插值准确率较高, 随着 ϵ 增大, 插值误差也逐渐增大。

关键词: 滑动窗口的普通克立格方法, 变异函数, 方位邻近法, 插值误差。

中图分类号: P413 **文献标识码:** A **doi:** 10.7519/j.issn.1000-0526.2013.04.011

Application of the Moving Window Ordinary Kriging Method in Precipitation Interpolation

XIONG Minquan^{1,2,3}

- 1 Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029
2 University of Chinese Academy of Sciences, Beijing 100049
3 National Meteorological Centre, Beijing 100081

Abstract: Based on daily precipitation data of 2200 observation stations over China in 2009, precipitation at grid point is estimated by the method of moving window ordinary kriging and three technical experiments are designed in this article. The interpolation errors are compared between the overall search method and direction neighborhood method in the moving window ordinary kriging. The different effects of maximum influence radius and shielding on the interpolation error are also discussed. The sample point selection strategy of the direction neighborhood method is put forward to sort data points. The results show that compared with the overall search method for the traditional moving window ordinary kriging (MWK), direction neighborhood method has the notable advantages of more computational efficiency and high accuracy especially in the areas with dense sites. The results also show that good interpolation effect can be attained within the range from 4° to 5° meridional (zonal) arc length for most areas; and the shielding effect of precipitation is weak. Further more, better accuracy for precipitation estimation is presented when the ϵ value is equal to 0.1, but it is likely to get worse with the increase of ϵ value.

^{*} 公益性行业(气象)科研专项(GYHY201106027 和 GYHY201106010)共同资助
2012 年 1 月 19 日收稿; 2012 年 12 月 4 日收修定稿
作者:熊敏詮,主要从事动力统计预报研究. Email:minquanxiong@hotmail.com

Key words: moving window ordinary kriging, variogram, direction neighborhood method, interpolation error

引 言

由于受到地理位置及观测仪器等因素的影响,降水量的空间变率较大,因此获取精度较高的降水量格点化资料存在一定难度。在众多的插值算法中,克立格方法被认为是具有一定优势的方法。克立格方法是地统计学中重要的统计技术,主要原理是利用变异函数来表示区域化变量的内部关系。此方法在气象中的相关研究主要有:Geard 等(1999)在海平面气压场的 EOF 空间上使用克立格方法对降水量进行了估计,表明估值精度有了提高;Barref (1988)根据日降水量,使用克立格方法对降水化学成分的分布和趋势变化进行了估计;Prudomne 等(1998;1999)使用克立格方法在地形复杂的山区对极端日降水量进行插值分析;Courault 等(1999)对于不同的大气环流型使用地统计学方法做温度插值。魏凤英等(2002a;2002b)使用变异函数揭示了降水场空间变异结构并讨论了地统计学在气象中的应用;高歌等(2007)在 1 km×1 km 格点场上对日降水量开展插值分析,得到普通克立格方法效果好于反距离权重方法;使用普通克立格方法,进行中国热量资源精细化估算(王怀清等,2011)。

根据研究目的和条件的不同,各种各样的克立格方法相继产生(Haas,1990a),如当区域化变量满足二阶平稳(或内蕴)假设时,可用普通克立格法;在非平稳条件下采用泛克立格法;当区域化变量服从对数正态分布时,可用对数正态克立格法;对有多个变量的协同区域化现象可用协同克立格法;对有特异值的数据可用指示克立格法等。当然,应用最为广泛的是普通克立格方法。Hass(1990b)利用观测站点资料对美国全国范围的待估点进行酸沉降估值时,较早地使用了滑动窗口的普通克立格方法(MWK),结果表明,MWK 对研究对象的空间分布估计更有效,而且提高了估值精度。Hass(1996)还进一步扩展了滑动窗口应用对象,在对数正态克立格方法、回归克立格方法和协克立格方法中也应用到了滑动窗口方法。Walter 等(2001)在普通克立格方法中采用了滑动窗口方式估值;Lloyd 等(2002)把滑动窗口方式用于泛克立格方法;Pardo-Iguzquiza 等(2005)在泛克立格方法中使用滑动窗

口方式对 0~22°N 和 22°~44°E 的区域 1999 年 7 月的 250 个测站的月降水量进行了分析。采用滑动窗口方式进行空间数据分析主要可用于气象、酸雨、土壤及高程;研究以上领域的的数据样本都有覆盖范围广和数据量大的特点,是滑动窗口主要应用范畴。

一般认为,虽然降水在较大范围和不同方向上变化起伏大,但是在有限的局部区域中,降水量可以被近似地认为满足平稳和内蕴假设;在此前提下,对局部地区进行点的克立格方法估计会得到更为精确的降水量,因此,滑动窗口的普通克立格方法更适合降水量的格点化。本文在分析滑动窗口的普通克立格方法在降水量估值适用性分析基础上,针对存在的问题设计 3 种不同插值试验方案,比较了全局搜索与方位邻近方法的误差,讨论了最大影响半径及屏蔽效应对插值效果的作用。

1 MWK 方法及其在降水量估值的适用性

1.1 普通克立格方法

设区域化变量 $R(X)$ 满足二阶平稳及本征假设,其数学期望存在且等于常数,其协方差函数存在且平稳:

$$E[R(x)] = m \tag{1}$$

$$\begin{aligned} \text{Cov}[R(x), R(x+h)] &= E[R(x)R(x+h)] - m^2 \\ &= C(h) \end{aligned} \tag{2}$$

变异函数 $\gamma(h)$ 的估计量 $\gamma^*(h)$ 是:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [R(x_i) - R(x_i+h)]^2 \tag{3}$$

式中, $N(h)$ 是被向量 h 相分隔的试验数据对的数目。

变异函数一般选择为球状模型:

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

其中, C_0 为块金常数, C 为基台值, $C_0 + C$ 为总基台值, a 为变程。

1.2 滑动窗口的普通克立格方法

Journel 等(1978)定义和讨论了准平稳过程,指

出在相对大区域的非平稳过程,在子区域中表现出更小的变化趋势和更多的各向同性特点,因此,在子区域中可以近似地认为是准平稳的,这为滑动窗口的普通克立格方法(MWK)的广泛应用提供了理论基础。将大范围区域划分成许多子区域,并在子区域中对待估点使用普通克立格方法估值,这种方法就是 MWK。以待估点为圆心,以最大影响半径所画的圆形覆盖区域称之为窗口,而窗口是随着位置不同的待估点而移动的。

与普通克立格方法不同的是,在实施 MWK 时,需要对每一观测值分别赋予一定的权系数,最后进行加权平均来估计待估点的数值。在无偏条件下,得到普通克立格方程组(侯景儒等,1994),在内蕴假设下,可用 $\gamma(h)$ 表示如下:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} \bar{\gamma}(v_{\alpha}, v_{\beta}) + \mu = \bar{\gamma}(v_{\alpha}, V) \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases} \quad (\alpha = 1, 2, \dots, n) \quad (5)$$

1.3 MWK 在降水量估值中的适用性

降水量具有以下区域化变量的属性:(1)空间局限性,站点降水量是空间点上在某个时间段内观测的降水累计量;(2)连续性,这种连续性可以通过相邻样品之间的变异函数来描述,变异函数是能够对降水进行 MWK 插值的基础,通过变异函数可以得到较高精度的格点值;(3)异向性,降水量在各个方向有不同的分布特点;(4)相关性,降水量在一定范围内、一定程度上表现出空间相关性,对于普遍存在的过程性降水,这种相关性特点是显著的,即使是局地强对流天气,在一定观测密度下或在邻近的格点中,也可以有较好的相关性,当然,超出一定范围后相关性会逐渐减弱或消失。最后,站点的降水量数据具有范围大和随时间变化两大特点,符合 MWK 应用范围。

当使用 MWK 进行降水量插值时,为了取得更快运算效率和得到更为精确的估值,在适用方面应该考虑如下几点问题:

(1)计算资源耗用大是各类克立格方法在气象领域应用中需要克服的难点,也是使用 MWK 进行降水估值中遇到的问题;这是由于各类克立格方法都要进行待估点和样本点之间、样本点之间的变异函数计算,还要进行克立格矩阵变换的计算,都会占

用大量计算资源。虽然 MWK 已经把大区域细分成更小的子区域,客观上减少了参与插值的样本点,减少了计算耗时,但是,对于气象类资料分析还是不够。

(2)MWK 对参与插值的样本并没要求它们具备空间拓扑结构,即在插值过程中只是对窗口内的数据点进行全局搜索,也就是说,窗口内的所有数据点都将参与插值,而没有考虑变量的特性和样本点分布特点。但是,降水量是具有空间相关和各向异性,同时,观测站点也是散乱分布的,若直接使用全局搜索对降水量插值,降低了计算效率,也可能会影响到插值精度。

(3)理论变异函数的参数的计算至今没有好的方法,通常有两种拟合方法:一般是人工拟合法,就是首先给出实验变异函数,根据地质条件和人工经验,选择相应的变异函数模型,最终确定相关参数;另一种方法是自动拟合,使用加权多项式回归、线性规划、目标规划等多种数学手段拟合,但这些方法效果不太理想。上述两种方法在地质采矿的有限数据点的估值计算中可行的,但是对于气象中有上万个格点需要被估计时,其计算量和复杂度很难克服。

(4)气象观测站点相距较远,一般是以每个行政县(区)为单位有一个降水人工观测站,测点相距数十千米到几百千米不等,同时站点分布也不规则,由此,气象站点疏密分布和高精度的地质勘探有较大的不同,当在应用 MWK 时要充分考虑此特点。

2 不同方案的 MWK 降水量插值误差的比较

2.1 试验方案

针对滑动窗口的普通克立格方法在降水量格点化应用中存在的问题,设计了 3 种试验方案。

2.1.1 试验方案 1

使用方位邻近法筛选已知点。将全国 2200 个测站的降水量插值到 $0.28125^{\circ} \times 0.28125^{\circ}$ 网格上,共有两万多个格点需要估值。虽然 MWK 缩小了采样范围,但是在观测较密集的中东部地区,传统的全局搜索法下的 MWK 仍然无法在当前的计算条件下正常运行。本文提出方位邻近法进行数据点选择,具体技术路线如下:将待估点为中心的 360° 圆内平均划分成 32 个方位,每个方位对应夹角是 11.25° ,在不同方位上选择距离圆心(待估点)最近

的观测点作为采样点,若该圆的半径(即最大影响半径)足够大时,就有32个采样点,若圆的半径越小,采样点就越少。在挑选数据点时有4个方面的因素需要被考虑:

(1)数据点的多寡直接影响到估值准确度和计算量;MWK运算复杂度是 $O[n^3]$ (徐建华,2002),过多的数据点,将增加克立格方程的维数,提高计算复杂度;若数据量偏少,较少的信息量也难以准确估值;试验表明,20多个数据点能达到较好的效果,因此平均划分成32个方位是比较理想的选择。

(2)待估点周围的观测站散布分布,各地域间站点疏密度相差较大,方位邻近法尽量使采样点均匀地分布在待估点周围,降低了站点散布分布对格点估值过程的影响。

(3)降水在大范围区域表现出明显的各向异性,而在小区域中,降水的各向异性有所减弱,这也是MWK得以应用的原因之一。

(4)近邻性,选择距离格点最近的站点作为有效数据,在距离较近的情况下,降水量具有较好的相关性;距离过远降水量的相关性降低,会使变异函数模型的可靠性降低。全局法和方位邻近法相比较,全局法是以待估格点为中心,最大影响半径为圆的范围内的观测点都参与插值,由于站点数目决定了克立格方程维数,随着维数的增加,对计算环境要求就很高。本试验中,两种挑选站点方法的最大影响半径设置均为 2° , ϵ ($\epsilon=C_0/C$, ϵ 为无量纲量)为0.1。

2.1.2 试验方案2

最大影响半径试验。最大影响半径是指在搜索样本点时的最远距离,在MWK中就相当于每个“窗口”的大小;如格点 $A(x_1, y_1)$ 和格点 $B(x_2, y_2)$, x 和 y 分别是经度和纬度,假设经向和纬向每隔 1° 的物理距离相等,那么,两点间距离 AB 可以近似用多少度的弧长来表示:

$$AB = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

由于要对每个格点分别建立方程求解待估值,格点所对应的变程和基台值的变异函数是可迁型的。变程 a 反映了待估点和已知点之间的关系,即区域化变量的影响范围。对于可迁型变异函数,以待估点 $z(x)$ 中心, a 为半径的影响范围内的已知点 $z(x+h)$ 而言,距离 $z(x)$ 愈近,则对 $z(x)$ 影响愈大,反之愈小,当 $z(x+h)$ 在影响区之外时,那么就没有影响了,因此 a 的大小关系到插值站点的选取。本文中,在不同方位上搜索观测点时能达到的最大距离,即

最大影响半径,最大影响半径和每个格点的变程不是同一个概念;在方位邻近法下,以距离待估点最远采样点的距离为变程,因此每个格点的变程不仅和最大影响半径有关,还和观测站点的空间分布有密切联系;最大影响半径大致能反映出变程的变化范围,尤其是在站点观测稀疏的地区。观测点位置是以地理坐标系中的经纬度为单位表示,为了便于比较分析,可以用点之间经纬度的间隔表示距离远近;若最大影响半径过小[例如 1° 的经(纬)线弧长],采样点则很少,在站点稀疏的西部地区就往往无样本点可以采集,若最大影响半径过大[例如大于 10° 的经(纬)线弧长],用较远的观测样本参与插值,可能会产生较大的误差。所以,在方位邻近法下,最大影响半径从 $2^\circ \sim 10^\circ$ 的经(纬)线弧长,对全国范围的格点插值误差进行分析。最后,对格点周围32个方位都全部有数据点的情况下进行插值误差比较(对于边界点仍是部分方位有数据点)。在不同的最大影响半径对比中, ϵ 取值固定为0.1。

2.1.3 试验方案3

块金效应(屏蔽效应)。地统计学上用变量 ϵ 表示块金(屏蔽)效应, $\epsilon=C_0/C$, ϵ 越小,屏蔽效应越强,反之,屏蔽效应越弱、块金效应强。 ϵ 取值由0.1~1.0共10个数值进行比较,试验中,使用方位邻近法采样,最大影响半径设为 5° 。所谓块金(屏蔽)效应就是数据点在方位和距离之间的相互影响。块金效应与观测尺度有密切关系,当测站间距大于降水结构范围时,就表现为块金效应;由式(4),表示 h 很小时,两点间观测值的相关性。若已知点间呈现纯块金效应时,则取消了已知点之间的屏蔽效应,即在一个平稳带内进行估值,估计值就是已知数据点的平均数。

2.2 资料及误差检验方法

使用2009年1月1日至12月31日的全国2200个观测站(基本站和加密站)08时24小时实况降水量资料。插值格点格距为 $0.28125^\circ \times 0.28125^\circ$ 。根据国家数值预报中心在预报产品评估中区划方法,把全国划分为8个区(图1):一区:东北,二区:新疆,三区:西北,四区:华北,五区:西藏,六区:西南,七区:长江中下游,八区:华南。

对于全国8个区域的降水量的MWK格点化效果采用交叉检验的方法进行检验,即用10次不重复随机地抽取观测值不参与插值,也就是说,每次有

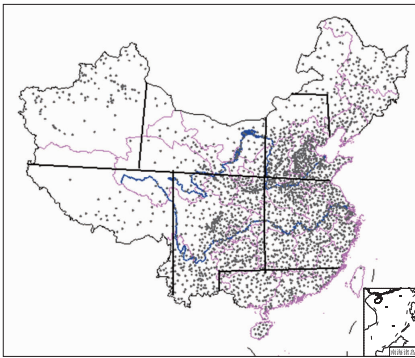


图 1 中国区域划分
Fig. 1 The regional division of China

10%的站不参与插值,只使用 90%的站进行格点插值。获得格点值后,通过双线性方法再插回到没有参与插值的站点上,然后算出其实际观测值和估算值之间的误差,以此评估插值方案的优劣。采用均方根误差、平均绝对误差和平均误差作为评估插值

效果的标准。

2.3 不同方案的误差比较

试验方案 1 中 8 个区域的均方根误差、平均绝对误差和平均误差检验结果如表 1 所示。从表 1 可以看出,在站点比较密集的地区,例如长江中下游和华南地区,方位邻近法插值精度有一定的提高。若从计算耗时分析,在长江中下游地区的格点估值过程中,方位邻近法中参与格点插值计算的站点数大致在 20 左右,而全局法则达到 50~60,按照 MWK 计算复杂度 $O(n^3)$ 分析,单个格点的全局搜索计算量是方位邻近的 21 倍,当数万个格点需估值时,对应有巨大的计算耗时。因此方位邻近法不但提高了插值准确率,而且大幅降低了计算耗时;更重要的是减少了克立格方程维数,降低了对计算机硬件的苛刻要求,当要对较多的格点进行估值时,方位邻近法才能满足普通机器环境下的日常业务需求,因而具

表 1 全局搜索和方位邻近在八个区域插值误差比较

Table 1 The interpolation error comparison between overall search and direction neighborhood in 8 regions

检验方法	站点选择方法	东北	新疆	西北	华北	西藏	西南	长江中下游	华南
均方根误差 /mm	全局搜索	1.609	0.727	0.864	1.424	2.333	3.079	3.381	5.034
	方位邻近	1.609	0.729	0.863	1.420	2.327	3.078	3.197	5.027
平均绝对误差 /mm	全局搜索	0.835	0.366	0.454	0.653	1.402	1.417	1.443	2.687
	方位邻近	0.833	0.366	0.454	0.643	1.401	1.407	1.407	2.667
平均误差 /mm	全局搜索	-0.010	-0.005	-0.001	0.002	-0.055	-0.001	0.034	-0.005
	方位邻近	-0.009	-0.006	0.002	0.002	-0.053	0.001	0.008	-0.001

有更为广泛的应用和推广价值。

图 2 表示最大影响半径取 $2^{\circ}\sim10^{\circ}$ 的经(纬)线弧长和全方位(最大影响半径足够大,在方位邻近法中,32 个方位的空间上都有站点参与对格点的插值,本文称之为全方位,即 32 个站点参与插值)时,均方根误差和平均绝对误差变化曲线(图 2a),Y 轴的左轴表示均方根误差,右轴为平均绝对误差。在 $2^{\circ}\sim10^{\circ}$ 的经(纬)线弧长之间,全国范围内的均方根误差在 2.28~2.282 mm 之间;而在 4° 和 6° 的经(纬)线弧长则分别对应相对小(0.28 mm)和相对大(0.282 mm)值的两个点; 10° 经(纬)线弧长以内的其他点相应的均方根值相近。图 2a 的平均绝对误差变化曲线中,在 2° 的经(纬)线弧长时平均绝对误差达到最小为 1.147 mm,到 3° 时平均绝对误差增加较明显达到 1.153 mm,在 $3^{\circ}\sim10^{\circ}$ 的经(纬)线弧长之间,平均绝对误差缓慢上升。图 2b 是平均误差变化曲线,从图可以看到,平均误差和绝对误差在变化趋势上有较大差异,特别是在最大影响半径取值

较小时,两者呈现出相反的变化:最大影响半径为 2° 的经(纬)线弧长时,平均误差高达 -0.007 mm, $2^{\circ}\sim5^{\circ}$ 的经(纬)线弧长之间则有较大幅度的减小, 5° 的经(纬)线弧长时的平均误差为 0, 5° 的经(纬)线弧长之后,平均误差略增加并保持平缓的变化。当最大影响半径足够大时,均方根误差和平均绝对误差均达到最大,而平均误差接近 0 mm,这说明当最大影响半径逐渐增大时,即参与插值的站点向 32 数值逐渐靠近时,插值误差会逐步放大,同时插值误差在偏大和偏小这两个正负区间摆动的幅度也会更剧烈。

为了比较方便,文中把 8 个区域的检验指标变化曲线放在同一张图中,由于每个区域的降水量相差很大,因此误差值会落在不同的量级,若用统一的 Y 轴刻度,曲线可能以近似直线形式显示,而无法比较分析;所以本文有关分区域比较图的纵坐标没刻度,图形表达的是变化图。图 2c 是 8 个区域的均方根误差变化曲线,西南和长江中下游同全国平均走势相近,其均方根误差数值分别为 3.08 和 3.19

mm;华南地区在 $2^{\circ}\sim 10^{\circ}$ 的经(纬)线弧长的变化范围内, 2° 的经(纬)线弧长时均方根误差偏大,其他影响半径对应的均方根误差变化特点和全国相近,其平均值达到 5.02 mm;东北、新疆两区和西藏呈相反的变化特点,均方根误差值在 $1\sim 2$ mm 之间;西北地区在半径比较小时,曲线走势难以确定,当半径大于 7° 的经(纬)线弧长之后,均方根增大;华北地区则在 6° 的经(纬)线弧长之后,变化趋势明确,均方根误差直线增大。均方根误差是误差离散度的指标,从分区检验比较中得出更有意义的结论,不同区域的变化揭示了空间散乱的站点和插值误差有密切关系,在实践中应具体问题具体分析,这一点不仅是 MWK 对应的问题,也是其他各种插值方法都要面对的难点;当用全国平均值进行检验分析时,往往会掩盖区域的不同特性。

8 个分区的平均绝对误差和全国平均绝对误差变化特点基本一致(图 2d),只是在不同分区中曲线

的曲率变化略有不同。在降水量比较大的区域如:长江中下游、西南和华南 3 个地区在 $3^{\circ}\sim 10^{\circ}$ 的经(纬)线弧长之间变化不同或基本相近;这 3 个地区观测站点较密集,区域内各站点年季降水量差异小,当最大影响半径变化不大时,对各个格点进行插值的站点数变化不大,因而平均绝对误差相差也近似;但是,当最大影响半径明显增大时,平均绝对误差也显著增加,图 2d 所示在全方位时,各个区域的平均绝对误差都有较大的增加。

由于 8 个区域的插值平均误差绘出的图形变化较复杂,将 8 个区的平均误差放在同一张表中(表 2)将有利于分析,从表 2 中可以看出,8 个区域的平均误差有较大的不同,东北、新疆和西藏在最大影响半径为 2° 的经(纬)线弧长是平均误差最大,分别是 -0.009 、 -0.006 和 -0.053 mm;西藏西北的平均误差表现出相反的变化,例如,西藏地区平均误差从 -0.053 mm 逐渐减小到 0 mm;从平均误

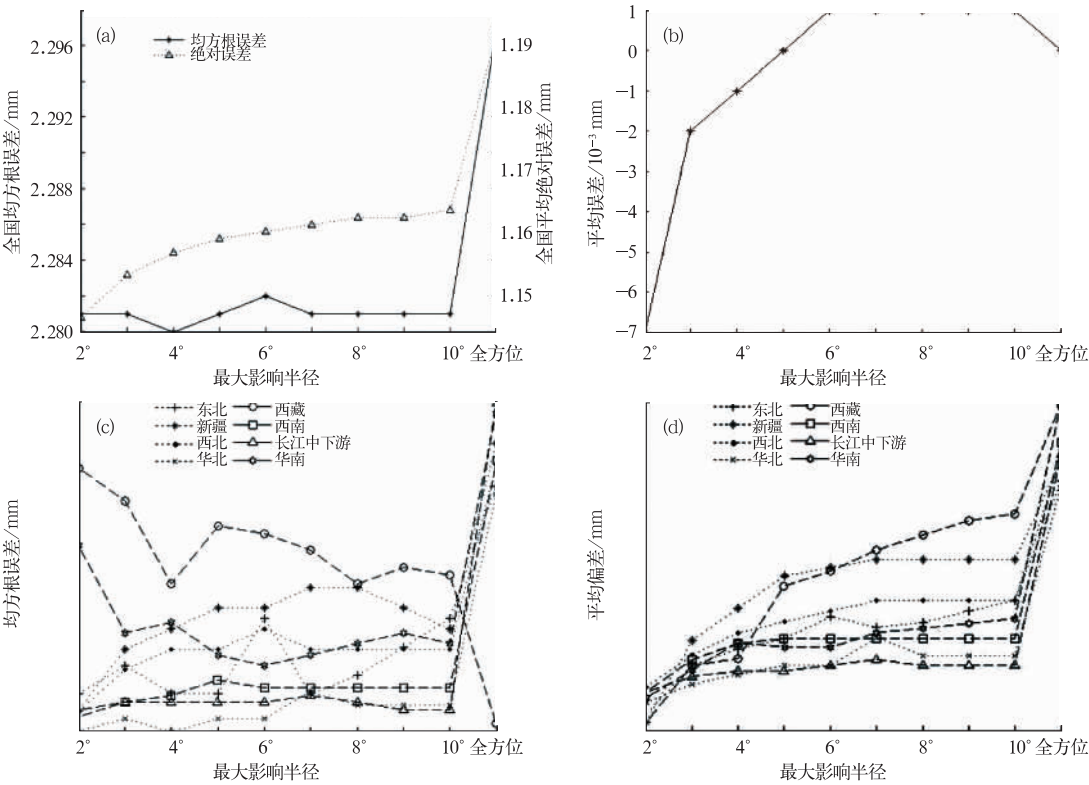


图 2 日降水量插值误差随最大影响半径变化

(a) 全国平均均方根误差和平均绝对误差, (b) 全国平均误差, (c) 8 个区域均方根误差对比, (d) 8 个区域平均绝对误差对比

[图中横坐标以经(纬)线弧长为单位]

Fig. 2 Variation trends of daily precipitation interpolation error according with maximum influence radius (a) national mean RMS error and absolute error, (b) national mean error, (c) comparison of RMS errors of 8 regions, (d) comparison of absolute errors of 8 regions

表 2 日降水量插值的平均误差(单位:mm)在 8 个区域的对比

最大影响半径	东北	新疆	西北	华北	西藏	西南	长江中下游	华南
2°	−0.009	−0.006	0.002	0.002	−0.053	0.001	0.008	−0.001
3°	0.001	−0.002	0.002	0.001	−0.027	0.001	0.008	−0.002
4°	0.002	−0.002	0.003	0.001	−0.015	0.002	0.008	−0.002
5°	0.002	−0.003	0.003	0.002	−0.01	0.001	0.008	−0.004
6°	0.002	−0.003	0.004	0.001	−0.006	0.001	0.008	−0.005
7°	0.001	−0.003	0.004	0.001	−0.003	0.001	0.009	−0.004
8°	0.001	−0.003	0.005	0.001	−0.001	0.001	0.008	−0.004
9°	0.001	−0.003	0.005	0.001	0	0.001	0.008	−0.003
10°	0.001	−0.003	0.004	0.001	0	0	0.008	−0.003
全方位	0	−0.002	0.003	0	−0.006	−0.001	0.007	−0.004

差正负值来看,新疆、西藏和华南的平均误差为负,其他地区为正值;在部分区域平均误差的数值大小和降水量强弱关系不显著,总体来看,西藏和长江中下游平均误差的数值偏大,华北和西南偏小。可见平均误差大小和站点分布、降水分布特点及降水量有关,这些变量之间及其和平均误差的关系较复杂,但是当站点较疏散时,平均误差会明显增大。

试验云案 3 的结果如图 3 所示。从图 3a 和 3b 可以看出,全国均方根误差、平均绝对误差和平均误

差绝对值随着 ϵ 取值的增大呈现单调递增。当 ϵ 为 0.1 时,均方根、平均绝对误差和平均误差分别达到 2.281、1.158 和 0 mm。图 3c 和 3d 分别表示了 8 个区域的均方根和平均绝对误差变化曲线,分区域的变化趋势和全国平均值的单调递增格局保持一致。虽然大多数区域变化趋势相似,但是也有个别区域有些差异:在西藏地区的均方根在 ϵ 取 0.1~0.3 时,有较弱的减小;新疆有较大不同,当 ϵ 为 0.1~0.2 时,新疆平均误差达到 −0.003 mm,随后逐

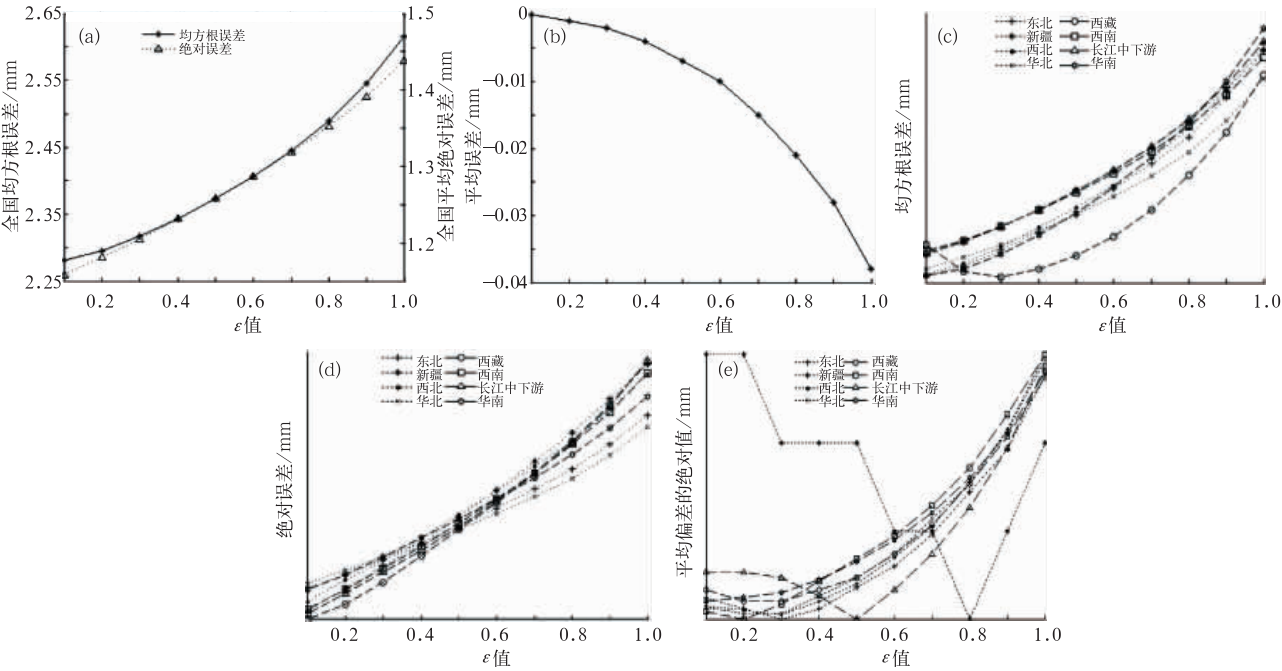


图 3 日降水量插值误差随 ϵ 的变化

(a) 全国平均均方根误差和平均绝对误差变化图, (b) 全国平均误差变化图, (c) 8 个区域均方根误差对比, (d) 8 个区域平均绝对误差对比, (e) 8 个区域平均误差的绝对值对比
(图中横坐标 ϵ 是无量纲量)

Fig. 3 Variation trends of daily precipitation interpolation errors according with ϵ values
(a) national mean RMS error and absolute error, (b) variation of national mean errors,
(c) comparison of RMS errors of 8 regions, (d) comparison of absolute errors of 8 regions,
(e) comparison of mean errors of 8 regions

步趋于 0 mm,当 ϵ 为 0.8 时,平均误差也开始增大。长江中下游地区在 0.5 时达到最小值;平均误差(图 3e)最小值在大多数区域出现在 ϵ 为 0.2 或 0.3,但是与 ϵ 为 0.1 时比较,误差变化幅度较弱。总体来看,均方根误差、平均绝对误差和平均误差都随着 ϵ 值增大而增大,只是在个别地区,由于站点分布和降水结构有较大不同,导致误差变化趋势略有差异。

3 小结与讨论

针对传统 MWK 直接应用到降水量插值中遇到的问题,本文提出了不同的技术处理方案,使该方法在降水中使用更趋合理。通过不同方案的误差比较得到以下结论:

(1) 使用适量的已知点对于插值的精度很重要。当已知点过多时,会使插值准确率下降,因为过多的信息量会掩盖有用的信息,这与降水量空间相关性有密切关系;因此对于每个待估格点,在已有的资料的情况下,寻找适量的已知点是插值要面对的重要问题。

(2) 以待估点为中心,在平均划分的 32 个方位上选择距离最近观测点作为样本点,显著降低了计算耗时、减少了存储开销,同时可获得较高的插值准确度。在观测站点密集地区,方位邻近法具有明显的优势。

(3) 最大影响半径和变程呈正相关关系(以离待估点最远的样本点的距离作为变程),在相同的最大影响半径下,使用方位邻近法时,每个格点的变程是有差别的。在散乱数据点中,变程是数据点的空间相关性的反映;变程大,则相关尺度也大,说明降水量的变化比较缓和;变程小,相关性小,降水量变化较剧烈。以上特性在站点稀疏的西藏和降水局地变化较强的华南地区有较典型的反映。

(4) 变程的变化对各个区域插值精度的影响呈现复杂变化,在降水量大、站点较密集的中东部地区,当最大影响半径选择在 $4^\circ \sim 5^\circ$ 的经(纬)线弧长时,均方根误差、平均绝对误差和平均误差达到较低的数值,从气象学角度能得到较好的解释,一般天气尺度的降水系统 24 小时的移距在 1000 km 左右,其对应的降水覆盖半径也就在 $4^\circ \sim 5^\circ$ 的经(纬)线弧长之间。

(5) 使用衡量变量的空间相关程度的 ϵ 值进行试验的结果表明, ϵ 等于 0.1 时,我国大部分地区的插值准确率较高,当 ϵ 逐渐增大时,准确率也随之降

低,说明降水的块金效应较弱,屏蔽效应强,观测站的降水量有显著的空间相关性。

参考文献

- 高歌,龚乐冰,赵珊珊,等. 2007. 日降水量空间插值方法研究. 应用气象学报,18(5):732-736.
- 侯景儒,黄竞先,等. 1994. 非参数及多元地质统计学的理论分析及其应用. 北京:冶金工业出版社.
- 王怀清,殷剑敏,辜晓青,等. 2011. 中国热量资源精细化估算. 气象,37(10):1283-1291.
- 魏凤英,曹鸿兴. 2002. 地统计学分析技术及其在气象中的适用性. 气象,28(12):3-5.
- 魏凤英,曹鸿兴,徐祥德. 2002. 变异函数在降水场空间特征分析中的应用. 南京气象学院学报,25(6):795-799.
- 徐建华. 2002. 现代地理学中的数学方法. 北京:高等教育出版社,113-114.
- Barrett P E. 1988. Statistical analysis of precipitation chemistry measurements over the eastern united states. Part II: Kriging analysis of regional pattern and trends. J Applied Meteorology, 27:1334-1343.
- Courault D, Monestiez P. 1999. Spatial interpolation of air temperature according to atmospheric circulation patterns in south-east France. J Climatol, 19:365-378.
- Geard B, Eduardo Z, Hans V S, et al. 1999. Estimation of precipitation by kriging in space of the sea level pressure field. J Climate, 12:1070-1085.
- Haas TC. 1990a. Kriging and automated variogram modeling within a moving window. Atmospheric Environment, 24A, 1759-1769.
- Haas TC. 1990b. Lognormal and moving window methods of estimating acid deposition. J Am Stat Assoc, 85:950-963.
- Haas TC. 1996. Multivariate spatial prediction in the presence of non-linear trend and covariance non-stationarity. Environmetrics, 7:145-165.
- Journel A G, Huijbregts C J. 1978. Mining Geostatic. London: Academic Press, 33-34.
- Lloyd C D, Atkinson P M. 2002. Nonstationary approaches for mapping terrain and assessing prediction uncertainty. Trans GIS, 6:17-30.
- Pardo-Iguzquiza E, Dowd P, Grimes D. 2005. An automatic moving window approach for mapping meteorological data. Int J Climatol, 26:665-678.
- Prudhomme C, Reed D W. 1998. Relationships between extreme daily precipitation and topography in a mountainous region. A case study in Scotland. Int J Climatol, 18:1439-1453.
- Prudhomme C, Reed D W. 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: A case study in Scotland. Int J Climatol, 19:1337-1356.
- Walter C, McBaratney A B, Douaoui A, et al. 2001. Spatial prediction of topsoil salinity in the Chelif Valley, Algria, using local ordinary kriging with local variograms versus whole-area variogram. Aust J Soil Res, 39:259-509.