

云南区域站降水资料利用 Surfer 软件 实现 Cressman 插值的研究

张超^{1,2}, 孙绩华³, 段玮⁴

(1. 云南省气象服务中心, 云南 昆明 650034; 2. 成都信息工程大学大气科学学院, 四川 成都 610225; 3. 云南省气象台, 云南 昆明 650034; 4. 云南省气象科学研究所, 云南 昆明 650034)

摘要:针对云南区域站大密度站点的插值问题,结合 Surfer 插值工具,探讨 ECMWF 高分辨网格下云南区域站降水资料的合理插值方案。利用 Surfer 平台自带的插值工具绘制云南区域站降水图,采用 Cressman 插值方法,设计不同的扫描方案,并基于 Surfer 平台绘制同一天的区域站降水图。结果表明,基于 Surfer 软件平台可以间接实现 Cressman 插值;逐级扩大的扫描方案(0.0625°、0.125°、0.34°),相比而言更适用于站点稀疏的地区;而在站点密度高的区域,使用较小的(0.0625°)扫描半径,插值效果更好。总体来看,Cressman 的插值结果要优于 Surfer 软件平台自带的插值方法,尤其是0.125°和0.0625°这两种插值方案,对站点资料的还原度更高。

关键词:计算机应用;客观分析;区域站;相邻距离;Cressman 插值;扫描半径

中图分类号:0241.3

文献标志码:A

doi:10.16836/j.cnki.jcuit.2018.01.015

0 引言

随着科技的进步和发展,计算机逐渐取代人工,开启了天气图的“客观分析”时代。所谓客观分析,通俗来讲就是指在计算机上实现天气图的自动绘制。客观分析虽然是为了满足数值预报的需要而诞生的,但它却有多种用途,包括为预报员提供实况天气图,大气的诊断研究,为统计预报提供因子……,这类基础性的研究工作^[1]。所以,要想让计算机“读懂”海量的气象数据,那么第一步就是对气象数据进行客观分析。

客观分析的方法,比较著名的有多项式插值法、逐步订正法、谱方法、变分法、最优插值法等等^[2-3],这些方法的不断更新,使客观分析的内涵越来越丰富。并且,正因为目前还没有一种得到广泛认同的“最佳”插值方案,所以这项工作一直以来都激励着气象科技人员不断研究和探索。

选取经典的 Cressman 插值方法,利用云南 2914 个区域站某一天的降水资料,基于 Surfer 软件平台,绘制降水实况的等值线图以及经过 Cressman 插值程序运行后,绘制不同扫描半径下的插值场的等值线图。就 Surfer 软件自带的插值工具和 Cressman 算法结果进行比对,并简单探讨以实际数据点的距离和临近数据点的点数^[4]为双重限制的插值方案设计。

1 Cressman 插值算法

Cressman 插值算法由前美国国家气象局局长 Cressman 博士提出,该算法的核心是:根据各临近数据点与插值点之间距离的远近来确定各临近数据点的权重系数。权重系数 w 的公式为

$$w = \begin{cases} 1, & r = 0 \\ \frac{R^2 - r^2}{R^2 + r^2}, & 0 < r \leq R \\ 0, & r > R \end{cases} \quad (1)$$

当数据点与插值点之间的距离 r 为 0 时,说明数据点与插值点重合,权重系数 w 自然为 1;如果数据点所在位置已经超出扫描半径 R 之外($r > R$),则权重系数 w 为 0,即该数据点对插值点的贡献为零;当 $0 < r \leq R$ 时,每个被扫描半径扫到的数据点,都可以求一个该数据点的权重系数。最终插值点上的值 $V_{(i,j)}$ 就是按照临近数据点的贡献大小依次计算得到:

$$V_{(i,j)} = \frac{\sum_{n=1}^n w_n v_n}{\sum_{n=1}^n w_n} \quad (2)$$

总体来说,距离插值点越远的数据点,则权重系数越小,对于插值点的贡献率也越小;距离插值点越近的站点,它对该插值点的影响就越大,表现为权重系数就越高。对于公式(2),不管扫描半径内有多少个数据点,每个数据点的值都按各自的权重系数进行相乘,并累加,再除以所有数据点的权重和,得到的就是该插值

点上的客观分析值 $V_{(i,j)}$ 。

2 区域站分布与扫描半径设置

2.1 云南区域站分布

把云南区域站的经纬度站表信息和具有降水实况资料的站点进行匹配,选取了匹配成功的 2914 个区域自动站。把这些站点按照经纬度坐标,用 Surfer 软件绘制出云南区域站分布图(图 1)。这些区域站坐落在海拔102~4368 m的垂直落差中,地形环境涵盖高原雪山,干热河谷,以及城市,乡镇,村落……,是目前云南区域内密度最大的一套站点资料。

选取欧洲中心细网格(0.125°×0.125°)作为资料的网格化目标,把站点资料插值到格点上。根据欧洲中心细网格资料的区域限制,在 90°E~115°E、10°N~35°N 的区域内,假设 1°=110 km,那么 EC 资料的区域有近 7500000 km²,40401 个格点。如果对应云南 390000 km²的面积,那么大致可以得出云南境内 EC 细网格的格点数为 2100 个左右,需要插值的格点数(2100)略小于区域站的站点数(2914),两者的平均密度基本接近。使用 Cressman 插值方案,如果合理设计扫描半径,有利于使格点的插值场尽可能接近站点的实况场。

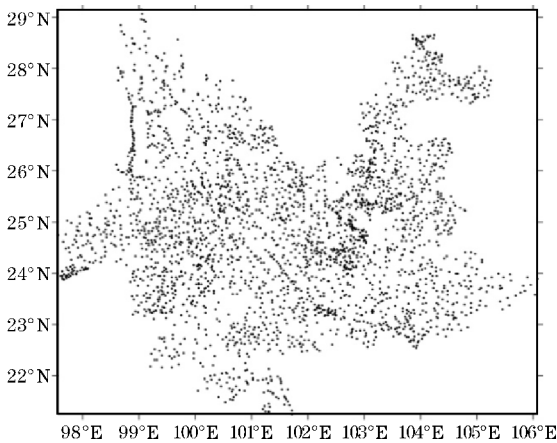


图 1 云南区域站分布

2.2 Cressman 扫描半径的设置原理

关于 Cressman 算法中扫描半径的设置,需要考虑站点与站点间的相邻距离。如果扫描半径过短,有可能只扫到少数站点,失去区域站的站点数量优势;而如果扫描半径设置太长,则可能使得每个格点上的值经过多次平滑处理,削弱了站点值的强中心,或者也可能在原本数值较小的区域计算出大值区,局部失真。

每个站点理论上都与周围离它最近的站点,有一个相邻距离。采用 ArcGIS 的近邻分析工具,分析所有云南区域站的站与站之间的相邻距离,见表 1。结果显示,大约有 61% 的站点,与其相邻站点的距离,在半个格距(0.0625°)的范围之内;有 97% 的站点,与其相邻站点的距离在一个格距(0.125°)的范围之内。同样假设 1°=110 km,那么经过换算后,可以得出不同距离区间内的站点数分布图(图 2)。云南大部分的区域站,相邻距离在 5~10 km。

表 1 各距离区间内的站点数统计

相邻距离/(°)	站点数	站点数累积百分比/%
0.00-0.009	109	4
0.01-0.019	162	9
0.02-0.029	303	20
0.03-0.039	351	32
0.04-0.049	374	45
0.05-0.059	398	58
0.06-0.069	365	71
0.07-0.079	264	80
0.08-0.089	156	85
0.09-0.099	149	90
0.10-0.109	95	94
0.11-0.119	60	96
0.12-0.129	42	97
0.13-0.139	34	98
0.14-0.149	20	99
0.15-0.34	32	100

注:站点总数为 2914 个

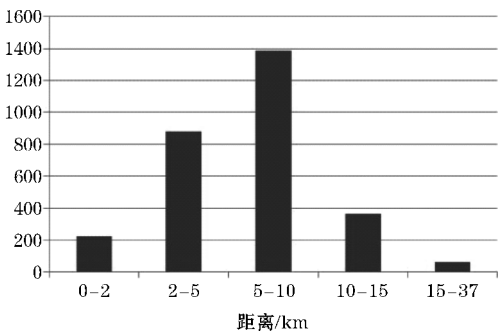


图 2 不同距离区间内的站点数分布

所以,在站点密度较大的地区,应当设置较小的 Cressman 扫描半径;在站点密度较小的地区,应当设置较大的扫描半径。结合插值算法,设置一个逐级扩大的动态扫描半径:即首先在0.0625°(半个格距)的范围里扫描,因为超过一半的站点数,相邻距离处于这个范围;如果扫到的站点数不足 3 个,就把扫描半径扩大到 0.125°(一个格距),在这个半径内,可以确保扫到大约 97% 的站点;如果这时扫到的站点数仍然小于 3 个,就

再把扫描半径扩大到站点与站点间的最远距离 0.34° , 使得扫描半径与站点间相邻距离的最大值相匹配。这就是逐级扩大的 Cressman 动态扫描半径设置方案。

Cressman 插值程序中,是用站点数来控制扫描半径的逐级扩大,在站点密度大的区域,保证每个格点上的值是由 3 个或 3 个以上的站点值求出。当站点数足够,则扫描半径不再扩大。而在站点密度小的区域,扫描半径扩大到相邻距离的最大值 (0.34°) 时,就算只搜索到 1 个站点,对于插值点来说,也足够计算了。

3 基于 Surfer 实现 Cressman 插值的不同方案

3.1 基于 Surfer 绘制降水实况

Surfer 软件自带的插值算法有 12 种,其中经过实践

绘图,比较接近实况的插值算法有:克里金插值法、自然邻点插值法、最小曲率法和线性插值的三角网法。

选取 2016 年 7 月 15 日云南区域站的降水实况资料作为试验数据。受辐合切变影响,14 日 20 时到 15 日 20 时,滇中以西和滇东边缘出现中到大雨局部暴雨天气。据区域自动站监测,出现 1 站特大暴雨(富宁-归朝:272.9 mm)、15 站大暴雨、290 站暴雨、738 站大雨、864 站中雨和 990 站小雨。由此,利用 Surfer 自带的上述四种插值方法来绘制当天的降水实况,见图 3。

在统一网格化规则 ($0.125^{\circ}\times0.125^{\circ}$) 的情况下,四种插值方法都描绘出了强降水的范围和量级。克里金和最小曲率法对强降水中心描绘更好,降水的局地性特征更明显。自然邻点法和线性插值法则是描绘降水区域的连续性更好。

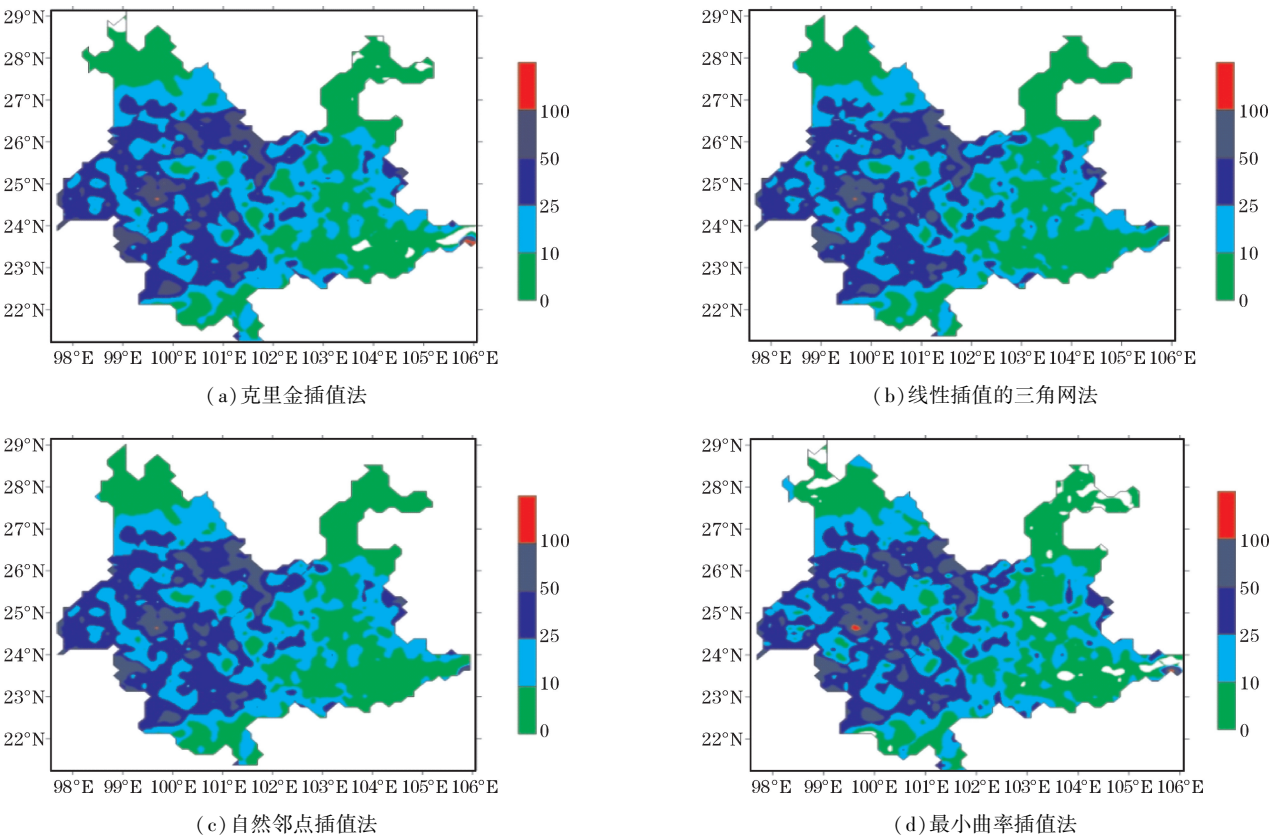


图3 2016 年 7 月 14 日 20 时-7 月 15 日 20 时累积降水(降水量:mm)

如果进行定量的比较,区域站降雨量的数据是 $0 \sim 272.9\text{ mm}$,经过 Surfer 网格化处理以后,克里金插值法的数据是 $-27.2 \sim 136.8\text{ mm}$,线性插值法的数据是 $0 \sim 124.8\text{ mm}$,自然邻点插值法的数据是 $0 \sim 117.5\text{ mm}$,最小曲率插值法的数据是 $-18.1 \sim 190.5\text{ mm}$ 。克里金和最小曲率插值法,数据出现了负值。把上述 4 种方

法的数量值百分比插值区间的边界值列表统计如表 2 所示。求出各插值区间的边界值与区域站之间的平均绝对误差。结果表明,就边界值的误差情况来看,最小曲率法的平均绝对误差 13.1 mm ,在 4 种插值方法中误差最小。

表 2 Surfer 软件自带插值方法对站点资料插值结果的数量值统计(单位:mm)

	克里金插值法	线性插值法	自然邻点法	最小曲率法	站点资料
最小值	-27.2	0	0	-18.1	0
1% 的数量值	-4.8	0	0	-4.4	0
5% 的数量值	0	0.1	0.2	-0.4	0
10% 的数量值	0.1	0.7	0.9	0.2	0.1
25% 的数量值	4.1	4.5	4.9	4.7	4.1
50% 的数量值	13.2	12.6	12.9	14.8	14.9
75% 的数量值	28.4	28.4	28.3	30.6	32.4
90% 的数量值	41.6	43.4	43.2	41.6	49.8
95% 的数量值	49.9	52.7	51.9	49.4	61.3
99% 的数量值	73.2	78.1	74.8	71.4	87.5
最大值	136.8	124.8	117.5	190.5	272.9
平均绝对误差(MAE)	18.9	16.4	17.5	13.1	—

注:平均绝对误差(MAE)的公式为 $MAE = \frac{1}{n} \sum_{i=1}^n |F_i - O_i|$ (F 为各插值百分比区间的边界数据, O 为站点数据)

3.2 基于 Surfer 绘制 Cressman 插值结果

Cressman 算法中,设置不同的扫描半径,最后的插值结果也不相同。下面分别介绍 4 种不同的扫描半径下,Cressman 插值结果的差异:

3.2.1 扫描半径(0.0625°、0.125°、0.34°)

这是理论上最完备的扫描方案,目标是扫到所有的站点。插值程序首先在半个格距的范围里扫描,站点数超过 3,就进入插值程序计算,否则扩大扫描半径为 1 个格距,继续扫描至少 3 个站点,如果仍然不足 3 个站点,再把扫描半径扩大到相邻距离的最大值,此时不管扫到几个站点,都进入插值程序计算,扫描半径不再扩大。

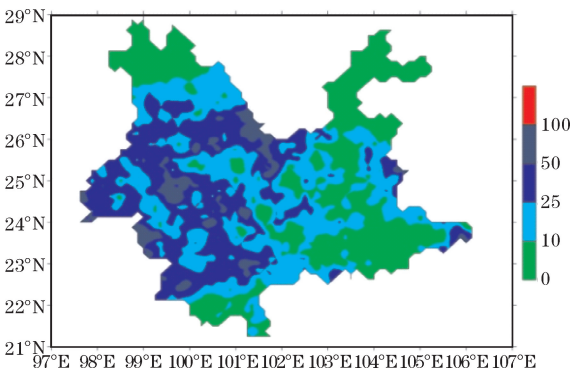


图 4 基于 Surfer 的克里金插值法绘制的 Cressman 插值结果(0.0625°、0.125°、0.34°)

根据网格报告,按照这样的扫描方式,计算出了 2887 个格点数,略多于云南区域内的格点数 2100。说明插值程序在云南区域的边界外围多计算出了一圈虚假的格点,基于 Surfer 软件的克里金插值法对边界进

行白化,留下云南区域内的格点数据。绘图前,Cressman 的插值数据是 0 ~ 137.8 mm,用 Surfer 自带的克里金插值法绘图后(图 4),数据是-3 ~ 137.8 mm。说明经过 Surfer 再次插值并绘成图像以后,Cressman 的插值结果基本没有被改变。所以,基于 Surfer 可以间接实现 Cressman 插值。

3.2.2 扫描半径(0.0625°、0.125°)

在 3.2.1 中,最外层的扫描半径是 0.34°,长度相当于两个半的格距还多。虽然前两步对扫描点数作了限制($n<3$),不太容易使用到最外层扫描半径,但处于云南边界地带的强降水中心就有可能因为扫描半径设置过长,而出现被削弱的情况。例如在图 5 中,滇东南的强降水中心富宁县归朝(272.9 mm),在图 4 中就无法显示出极值。因为半径扩得越大,搜索到的点数越多,意味着强中心被平滑掉的可能性越大。

所以对 3.2.1 进行简化,插值程序首先在半个格距的范围里扫描,站点数超过 3,就进入插值程序计算,否则扩大扫描半径为 1 个格距,继续扫描,此时不管扫描半径内有多少站点,都进入插值程序计算,扫描半径不再扩大。与 3.2.1 相比,少了扫描程序的第三步,即不在最外层的扫描半径里进行扫描。

根据站点间的相邻距离分析,半个格距的扫描半径,能确保扫到 61% 的站点;1 个格距的扫描半径,能扫到 97% 的站点。所以这种方法也能扫到绝大多数的站点。网格报告显示,Cressman 插值计算得到了 2207 个格点数,数据范围是 0 ~ 196.3 mm。经过 Surfer 自带的克里金插值显示到图上以后,数据变为 -11 ~ 190.4 mm。

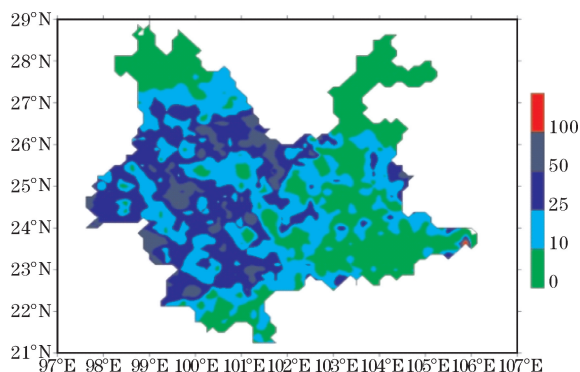


图5 基于 Surfer 的克里金插值法绘制的
Cressman 插值结果(0.0625°、0.125°)

3.2.3 扫描半径(0.0625°)

如果只把扫描半径限定在半个格距的范围内扫描,不管扫到几个站点,都进入插值程序计算,扫描半径不再逐级扩大。这种扫描方法意味着每个格点上,插值程序只扫描一次。

在图6上,出现了4个强降水的中心(降水值>100 mm),另外也有少部分的地方没有降水。总体而言,降水的局地性特征更为明显。根据网格报告,在半个格距的范围内扫描,Cressman 插值计算得到了1348个格点数,数据是0~272.9 mm。单从数值上看,插值结果的范围和实况站点降水量的范围是一致的,说明插值结果具有数值上的可靠性。经过 Surfer 的克里金法插值绘图以后,数据变为-30~261.6 mm。

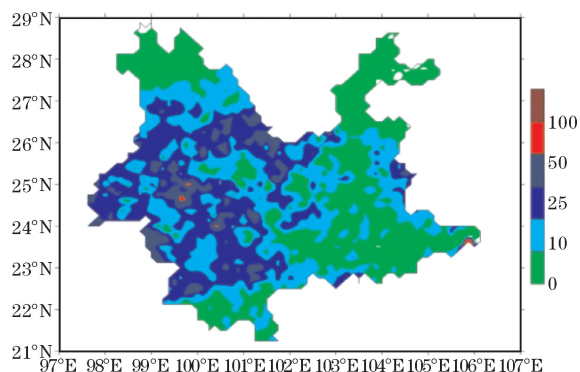


图6 基于 Surfer 的克里金插值法绘制的
Cressman 插值结果(0.0625°)

3.2.4 扫描半径(0.34°)

最后,把扫描半径单纯设置成0.34°,只扫描一次,不管扫到多少站点,都进入插值程序计算。

扫描半径固定为站点间相邻距离的最大值,一方面能扫到最远的一个站点,另一方面也会使得站点密度大的区域计算出来的临近格点值差别不大,降水的局地性特征并不明显,如图7。插值报告显示,有的格点上只扫到了1个站点,而最多的一个格点上,扫描到了87个站点,这些站点全部进入插值程序计算。

Surfer 软件的网格报告显示,Cressman 插值计算得到了2887个格点数。格点的数量与3.2.1相同,说明逐级扩大的动态扫描半径中,插值格点的数量是由最外层扫描半径决定的。另外,数据是0~137.8 mm,经过 Surfer 自带的克里金插值显示到图上之后,数据变成-3~137.8 mm。

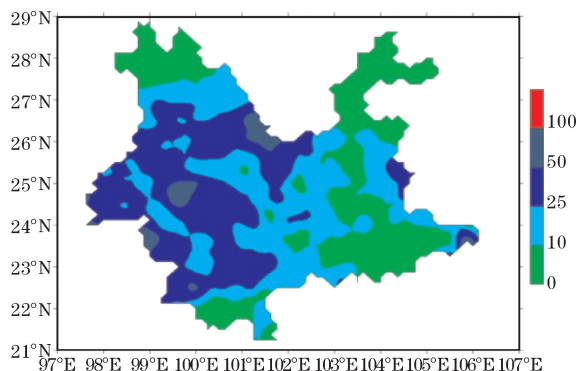


图7 基于 Surfer 的克里金插值法绘制的 Cressman 插值结果(0.34°)

3.2.5 总结

把以上4种不同扫描半径下得出的 Cressman 插值结果进行列表统计,如表3所示。理论上逐级扩大的扫描半径(0.0625°、0.125°、0.34°),具有最优的扫描效果,因为它能扫描到最远的一个站点,并且它比直接用最外层的扫描半径(0.34°),能得到更多数量的极大值,表现在表3中,99%的数量值区间的边界值更高(71.7>64.3),而且边界值的平均绝对误差更小,对于站点资料的还原度更高。

当扫描半径由3层减为2层,最后再减为1层的时候,数量值点数也呈现出递减的态势。而随着数量值点数的减少,插值结果的值域反而逐渐逼近站点资料的值域。

当扫描半径减为1层,半个格距(0.0625°)时,插值结果的值域与站点资料的值域(0~272.9 mm)恰好一致。此时如果再继续缩小扫描半径($R < 0.0625^\circ$),则数量值点数将减少到不再具有参考价值(表3中省略)。

平均绝对误差的结果也表明:直接使用1层,扫描半径为0.0625°的方案,得到的平均绝对误差最小,只有0.7 mm,这说明在站点密度大的区域,使用较小的扫描半径,有利于更好地还原站点资料。

总体来看,以临近数据点的点数控制下的 Cressman 插值方案,其插值结果普遍优于 Surfer 软件平台里自带的插值方法,尤其是0.0625°、0.125°(扫描两次)和0.0625°(扫描一次)这两种插值方案,平均绝对误差更小,对站点资料的还原度更高。最后,经过 Cressman 算法的插值计算,还能得出每一个插值格点上的具体

数值,从而方便对格点插值场进行精细化地诊断、检验。

表 3 不同扫描半径下 Cressman 算法插值结果的数量值统计/mm

	0.0625°、0.125°、0.34°	0.0625°、0.125°	0.0625°	0.34°	站点资料
最小值	0	0	0	0	0
1% 的数量值	0	0	0	0	0
5% 的数量值	0	0	0	0	0
10% 的数量值	0.2	0.3	0.1	0.3	0.1
25% 的数量值	4.8	5.0	4.5	6.1	4.1
50% 的数量值	14.4	15.0	14.8	15.7	14.9
75% 的数量值	29.8	30.7	32.0	30.1	32.4
90% 的数量值	44.0	45.8	48.2	41.8	49.8
95% 的数量值	53.1	54.9	59.7	48.9	61.3
99% 的数量值	71.7	76.4	91.1	64.3	87.5
最大值	137.8	196.3	272.9	137.8	272.9
平均绝对误差 (MAE)	15.3	9.2	0.7	16.7	—

4 结论

基于 Surfer13 软件平台,利用云南区域站资料,进行了 Cressman 插值的研究,得到以下结论:(1)对于云南近 3000 个区域站的分布来说,大部分区域站的相邻距离在 2~10 km,站点在空间分布上的密度大。(2)根据试验,Cressman 插值算法中,插值格点的数量是由最外层扫描半径的长短决定的,扫描半径越长,得到的插值格点数越多,扫描半径越短,插值格点数越少。(3)根据不同插值方案和区域站资料的对比结果,就 EC 细网格(0.125°×0.125°)的插值目标而言,采用 1 次扫描,扫描半径为 0.0625°的扫描方法,插值结果的平均绝对误差最小。(4)如果要扫描到所有站点,可以采用逐级扩大到站点间相邻距离最大值的扫描半径方案(0.0625°、0.125°、0.34°)。相比直接使用相邻距离最大值的扫描半径(0.34°),虽然也能扫到所有站点,但是逐级扩大的扫描半径,插值结果的平均绝对误差更小,对站点资料的还原度更高。(5)逐级扩大的扫描半径方案(0.0625°、0.125°、0.34°),相比而言更适用于站点稀疏的地区;而在站点密度高的区域,使用较小的(0.0625°)扫描半径,效果更好。总体来看,以临近数据点的点数控制下的 Cressman 插值方案,其插值结果普遍优于 Surfer 软件平台里自带的插值方法,尤其是 0.0625°、0.125°(扫描两次)和 0.0625°(扫描一次)这两种插值方案,平均绝对误差更小,对站点资料的还原度更高。Cressman 插值半径的大小,会影响到插值结果的客观真实性。云南干湿季分明,不同季节影响降水的环流背景也有较大不同。干季(冬春

季)以大尺度降水(冷锋降水)为主,Cressman 的插值半径相应更大;湿季(夏秋季)对流性降水明显增多,特别是在对流性降水更强的 8、9 月份,降水的空间尺度更小,降水的局地性特征更明显,这时候的 Cressman 插值半径相应就更小。

参考文献:

[1] 王跃山. 客观分析和四维同化——站在新世纪的回望(I) 客观分析概念辨析[J]. 气象科技, 2000,28(3):1-8.

[2] 王跃山. 客观分析和四维同化:——站在新世纪的回望(Ⅱ) 客观分析的主要方法(1)[J]. 气象科技,2001,29(1):1-9.

[3] 王跃山. 客观分析和四维同化:——站在新世纪的回望(Ⅱ) 客观分析的主要方法(2)[J]. 气象科技,2001,29(3):1-11.

[4] 胡金义,刘轩,林红. 基于 ArcGIS 的 Cressman 插值算法研究[A]. 中国水利学会 2010 学术年会论文集[C]. 2010:132-139.

[5] 沈桐立,温市耕,周慧喜. MOMS 中尺度模式的多元最优插值客观分析方法[J]. 应用气象学报,1996,7(2):195-202.

[6] 冯建设,王建源,崔玉东,等. 气象要素场客观分析软件[J]. 气象科技,2005,33(4):382-384.

[7] 蔡秀华,曹鸿兴. 资料插值的进展[J]. 气象, 2005,31(8):3-7.

[8] 余卫东,李湘阁,王靖. 气象场等值线自动绘制[J]. 气象科技,2002,30(4):222-225.

- [9] 张红杰,马清云,吴焕萍,等. 气象降水分布图制作中的插值算法研究[J]. 气象,2009,35(11): 131-136.
- [10] 冯锦明,赵天保,张英娟. 基于台站降水资料对不同空间内插方法的比较[J]. 气候与环境研究,2004,9(2):261-277.
- [11] Yukio Sadahiro. Accuracy of areal interpolation: A comparison of alternative methods[J]. Geographical Systems,1999,1(4):323-346.
- [12] Van H D, Christine T A, Anne V. Accuracy of areal interpolation methods for count data [J]. Spatial Statistics,2015,14:412-438.
- [13] Hutchinson M F. Interpolating mean rainfall using thin plate smoothing splines [J]. International Journal of Geographical Information Systems, 1995,9(4):385-403.
- [14] Wang Ting, Wan Zhimin, Zheng Weiguang. An improved state space method for force identification based on function interpolation in the presence of large noise[J]. Journal of Vibroengineering,2017,19(2):751-768.
- [15] Huang Jie, Tang Jian, Zhang Meijun, Zhang Xiaoming, et al. An improved EMD based on cubic spline interpolation of extremum centers[J]. Journal of Vibroengineering, 2015, 17(5): 2393 - 2409.

Research on Cressman Interpolation using Surfer Software based on Precipitation data of Yunnan Regional Station

ZHANG Chao^{1,2}, SUN Ji-hua³, DUAN Wei⁴

(1. Yunnan Meteorological Service Center, Kunming 650034, China; 2. College of Atmospheric Sciences, Chengdu University of Information Technology, Chengdu 610225, China; 3. Yunnan Provincial Meteorological Observatory, Kunming 650034, China; 4. Yunnan Institute of Meteorological Sciences, Kunming 650034, China)

Abstract: Aiming at the interpolation problem of large density sites in Yunnan regional stations, combined with Surfer interpolation tool, the reasonable interpolation scheme of rainfall data in Yunnan regional stations under ECMWF high resolution grid is investigated. Firstly, the interpolation tool provided by Surfer platform is used to draw the precipitation map of Yunnan regional station. After that, by using Cressman interpolation method, different scanning schemes were designed and the precipitation maps of the stations on the same day were drawn based on the Surfer platform. The results show that the Cressman interpolation can be realized indirectly based on the Surfer software platform. The scaled-up scanning scheme (0.0625° , 0.125° , 0.34°) is more suitable for the sparsely populated areas than that in the station. In the area with high site density, the interpolation effect is better with a smaller scanning radius (0.0625°). Generally speaking, Cressman's interpolation result is better than that of the interpolation method provided by Surfer software platform, especially interpolation of 0.0625° , 0.125° and 0.0625° program, the restoration of the site data are higher.

Keywords: computer application; objective analysis; regional station; adjacent distance; cressman interpolation; scan radius