

利用 K-邻域频数法处理多普勒雷达原始资料^①

王 俊

(山东省气象科学研究所, 济南 250031)

提 要

利用 K-邻域频数法对 714SDN 型多普勒天气雷达原始资料进行预处理, 分析了参数的选取对处理结果的影响, 得出了一些重要结论。根据应用中出现的问题, 对该方法作了补充完善, 使得补测效果更加有效。

关键词: K-邻域频数法 资料预处理 多普勒雷达

引 言

多普勒雷达探测的原始数据中包含不同尺度的风场信息, 要从中提取出有用的风场数据, 需要剔除原始数据中的奇异点、补足缺测的数据点, 以及消除小尺度和微尺度大气运动产生的速度脉动变化, 还要退速度模糊等。

补缺测点简单常用方法是插值法, 如文献[1]采用拉格朗日插值公式补缺测点。沿切线方向进行一维线性平滑可以消去速度脉动现象^[2], 周建芬等^[3]则利用变分原理对雷达径向速度资料进行预处理, 数值试验表明可以有效地消除径向速度资料中的小尺度运动和“噪声”。白洁等^[4]为了满足 VAP 法反演风场的需要, 提出一个二维区域滑动低通滤波方法对资料进行预处理, 经过滤波后, 中-β 尺度以上的风场信息相当完整的保留下来, 滤掉了小尺度以下的速度脉动变化。梁海河等^[5]提出了 K-邻域频数法来消除风场中的“噪声”, 在此基础上, 提出了“双径向-双切向”退模糊技术。

K-邻域频数法物理统计意义比较明确, 能够有效地处理风场信息中的“噪声”。本文将这一方法应用于 714SDN 型多普勒天气雷达原始资料的预处理, 分析了参数的选取对处理结果的影响, 并根据应用中出现的问题, 对该方法作了补充完善, 使补测效果更加有

效。

1 原 理

K-邻域频数法的要点是: 把 $\pm$ Nyquist 速度范围分成不同的 P 个等距区间 N_p ($p = 1, 2, \dots, P$), 区间间隔为 Δd , 另外窗口中心点的值 $V_{ij} \pm \Delta d/2$ 列为第 $p+1$ 个区间 N' , 然后统计在 $M \times N$ 窗口内的点分布在不同区间出现的频数 χ_p ($p = 1, 2, \dots, P$) 和 χ , 令 $\chi_{\max}$ 是 χ_p ($p = 1, 2, \dots, P$)、 χ 中的最大值, 取 $\chi_{\max}$ 所对应的区间的中点值为 V_p , 那么, 窗口中心点新值 V'_{ij} 可以表示为:

$$V'_{ij} = \begin{cases} 0 & V_{ij} \neq 0, \chi \leq k_1 \\ V_{ij} & V_{ij} \neq 0, \chi > k_1 \\ V_p & V_{ij} \approx 0, \chi_{\max} \geq k_2 \\ V_{ij} & V_{ij} \approx 0, \chi_{\max} < k_2 \end{cases} \quad (1)$$

式中, k_1 是剔除噪声阈值, k_2 是补值阈值, “0”代表无回波。

K-邻域频数法的处理结果与参数 P 、 M 、 N 、 k_1 和 k_2 有关。文献[5]没有从理论上解决统计窗口的大小与系数 k_1 和 k_2 的关系。因此, 在实际应用过程中, 应根据实测资料通过实验来确定 k_1 和 k_2 的取值。

714SDN 多普勒雷达在体扫过程中, 不同的扫描仰角具有不同的 Nyquist 速度。9 层扫描模式中, 最低层的 Nyquist 速度为

^① 科技部社会公益研究专项资金项目(2001DIE20104)、山东省气象局科技创新项目和 2004 年省局重点攻关项目共同资助。

$13.0\text{m}\cdot\text{s}^{-1}$, 2、3层为 $19.5\text{m}\cdot\text{s}^{-1}$, 上面6层为 $26.0\text{m}\cdot\text{s}^{-1}$ 。如果统计窗口大小固定, 则平均数 $\bar{\chi}$ 对不同的层就有很大的差别, 在补测时采用固定的 k_1 和 k_2 是不合适的。为了克服这种情况, 我们采用大小变化的统计窗口, 最低层采用 5×5 , 中间两层采用 5×7 , 上面6层则用 5×9 , 即在切向保持5点不变, 径向上有变化。这样, 一方面, 各层的平均数差别不大, 采用固定的 k_1 和 k_2 是合理的; 另一方面, 由于714SDN多普勒雷达的径向上的分辨率为250m, 9点格距为2km, 完全可以满足一般中尺度分析的需要。

另外, 在补缺测值时, 存在这样一种情况: 统计窗口内的点在不同区间的频率分布表明, $\chi_{\max}$ 对应的区间为 $(V_{ij} - \Delta d/2, V_{ij} + \Delta d/2)$, 此时 $V_{ij} \approx 0$, 虽然 $\chi_{\max} \geq k_2$, 根据式(1), 补测后的值仍为0, 即没有填补。进一步分析窗口内数据点的分布, 有两个明显特征, 一是窗口内的最大值或最小值与 V_{ij} 相差较大, 二是次最大频率 $\chi_{\max 1}$ 所对应的速度区间内的数值与 V_{ij} 相差较大且出现的频率也比较大。这表明缺测点较多, 是一小片区域, 但小于统计窗口的大小, 而统计窗口内的非零值占有比较大的比例。为了更加合理的填补这类缺测点, 采用以下做法: 令 $\chi_{\max 1}$ 所对应区间的中点值为 V_{p1} , 则窗口中心点的新值 V''_{ij} 为:

$$V''_{ij} = \begin{cases} V_{p1} \cdots V_{ij} \approx 0, \chi_{\max 1} \geq k_{21} \\ V_{ij} \cdots V_{ij} \approx 0, \chi_{\max 1} < k_{21} \end{cases} \quad (2)$$

2 结果分析

2.1 参数 M 、 N 、 k_1 和 k_2 对处理结果的影响

统计窗口 $M \times N$ 的大小影响 k_1 和 k_2 的取值以及最终的处理效果。下面利用实测资料分析 $M \times N$ 的大小与 k_1 和 k_2 的关系。

利用2003年5月15日体扫资料分析不同大小的统计窗口、及不同 k_1 取值时能剔除的孤立点数, 结果如图1。可以看出, 在取相同 k_1 时, 随着统计窗口的减小, 被作为孤立点剔除的点数迅速增多; 当统计窗口大小不变时, 随着 k_1 的增大, 孤立点数也不断增多。

从上述结果来看, 除孤立点时, 应仔细选择统计窗口的大小, 然后选择合适的 k_1 。另外原始体扫资料的质量对于 k_1 的取值有重要影响, 对于714SDN雷达, 统计窗口分别取 5×11 、 5×9 、 5×7 、 5×5 , 当Nyquist速度为 $13.0\text{m}\cdot\text{s}^{-1}$ 时, 合适的 k_1 取值为3、2、2、2; 当Nyquist速度为 $19.5\text{m}\cdot\text{s}^{-1}$ 时, 合适的 k_1 取值为6、5、4、3; 当Nyquist速度为 $26.0\text{m}\cdot\text{s}^{-1}$ 时, 合适的 k_1 取值为7、6、5、4、3。在上述配置下, 剔除的孤立点数相差不大, 效果也比较好。

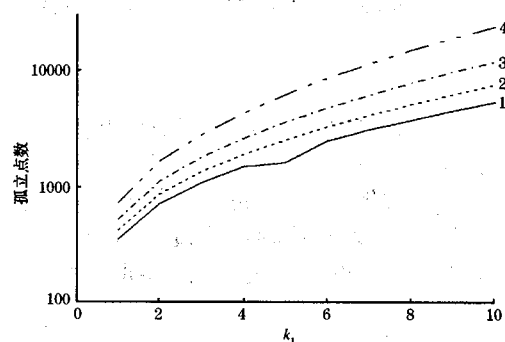


图1 不同大小统计窗口时可剔除的孤立点数与参数 k_1 的关系

曲线1对应的统计窗口为 5×11 , 曲线2对应 5×9 , 曲线3对应 5×7 , 曲线4对应 5×5 。

表1是取不同大小的统计窗口时可填补的最大缺测点数。结果表明, 在补缺测点时存在最大可补测点数, 即 k_2 有大于1的最小值 $k_{2\min}$, k_2 小于 $k_{2\min}$ 时, 能填补的缺测点数不变; k_2 大于 $k_{2\min}$ 时, 随着 k_2 的增大, 填补的缺测点数逐渐减少(图2), 开始时变化比较缓慢, 当 k_2 增大一定大小时, 可填补的缺测点数迅速减少。从实验结果来看, 最大补测点数目较少, 占总数的比例低(表1), 一般不超过5.0%, 因此在实际应用中可以将满足条件的点全部填补, 而不必担心影响整体数据结构。

此外, 随着统计窗口的减小, 最大填补点数逐渐减少, 但减少的幅度并不大, 这表明填补的点数对统计窗口的大小不敏感。

表中个例分析表明, 对不同的降水过程, 可填补的最大缺测点数有差别, 天气系统越剧烈, 填补的点数越多。

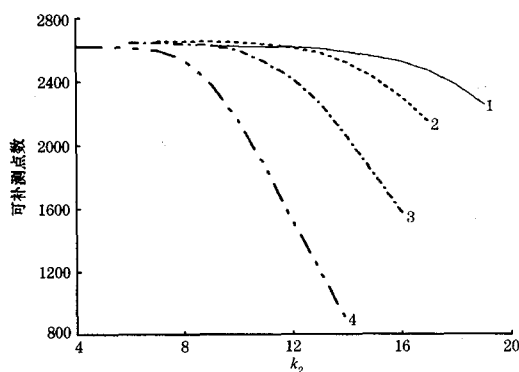


图2 不同大小统计窗口下可补测点数与参数 k_2 的关系

说明同图1

表1 统计窗口的大小与可填补的最大缺测点数之关系, Nyquist 速度为 $19.5\text{m}\cdot\text{s}^{-1}$

时间	$M \times N$	最大补测点数	百分比/%	$k_{2\min}$
2002.3.20 小雨	5×11	4087	3.00	7
	5×9	4055	2.94	6
	5×7	4005	2.90	5
	5×5	3911	2.83	3
2002.5.15 中雨	5×11	4772	3.46	8
	5×9	4702	3.41	6
	5×7	4613	3.34	5
	5×5	4465	3.24	4
2002.8.23 冰雹	5×11	1582	5.02	8
	5×9	1436	4.56	6
	5×7	1388	4.41	6
	5×5	1293	4.10	4

2.2 修改后的 K-邻域频数法补测效果分析

利用修改后的 K-邻域频数法对资料进行预处理,为便于对比,原方法称为方法 A,修改后的方法为方法 B,统计分析结果如下。表 2 是 2002 年 3 月 20 日 13 次体扫资料预处理结果的平均值,体扫资料包括 6 层;表 3 是 2002 年 5 月 15 日 10 次体扫资料预处理结果的平均值,体扫资料包括 9 层。补测范围是 120km 以内, $k_2 = 8$, $k_{21} = 8$ 。

对于两种体扫模式,方法 A 填补的点数从低到高一般是逐渐减少的,所占比例一般不超过 5%。方法 B 填补的点数上下层差别不大,所占比例一般不超过 3%。两种方法相比较,仰角较低的 3 层,方法 A 比方法 B 填补的点多,其它的扫描层,方法 B 填补的点一般要多。两种方法共同填补的缺测点数

所占比例不超过 7%。

表2 2002 年 3 月 20 日 13 次体扫资料预处理结果的平均值

	方法 A		方法 B		(A+B)/%
	补测的点数	占总数的百分比	补测的点数	占总数的百分比	
一层	6268	4.55	3454	2.45	7.00
二	5231	3.83	2795	1.99	5.82
三	5012	3.66	3281	2.33	5.99
四	2015	1.49	3531	2.51	4.00
五	1647	1.37	2990	2.12	3.49
六	1200	0.98	2358	1.68	2.66

表3 2002 年 5 月 15 日 10 次体扫资料预处理结果的平均值

	方法 A		方法 B		(A+B)/%
	补测的点数	占总数的百分比	补测的点数	占总数的百分比	
一层	5209	3.91	1426	1.01	4.92
二	5483	4.02	2934	2.08	6.10
三	3350	2.48	2031	1.44	3.92
四	2208	1.65	3360	2.38	4.03
五	2061	1.54	3992	2.83	4.37
六	1661	1.36	3566	2.53	3.89
七	1240	0.99	3001	2.13	3.12
八	867	0.68	2539	1.80	2.48
九	633	0.63	2405	1.71	2.34

2.3 个例分析

图 3a 是第 200(50km)距离圈数据随方位的分布,图 3b、c 分别是方法 A 和方法 B 补测后的数据随方位的分布图。1、3 处的缺测数据方法 A 能完全填补,点 2 和 4 处的缺测数据方法 A 没有作用,方法 B 能完全填补点 2 处的缺测数据,点 4 处的缺测数据比较多,方法 B 也没有完全填补这一小片缺测数据。从图中可以看出,两种方法都对缺测的点进行了合理的填补,而方法 B 填补的点更多一些。

3 结论

(1) 利用 K-邻域频数法除孤立点时,统计窗口的大小和 k_1 值的选取对结果有重要影响。在统计窗口固定时,应根据原始数据的质量选择合适的 k_1 值,以达到最佳效果。

(2) 试验表明,补缺测值时,在统计窗口一定的情况下,存在最大可补测点数,并且这一数目占总数目的比例很低。另一方面,统计窗口的大小对最大可补测点数影响不大,因此,在补缺测点时, k_2 应取得小一些,使填

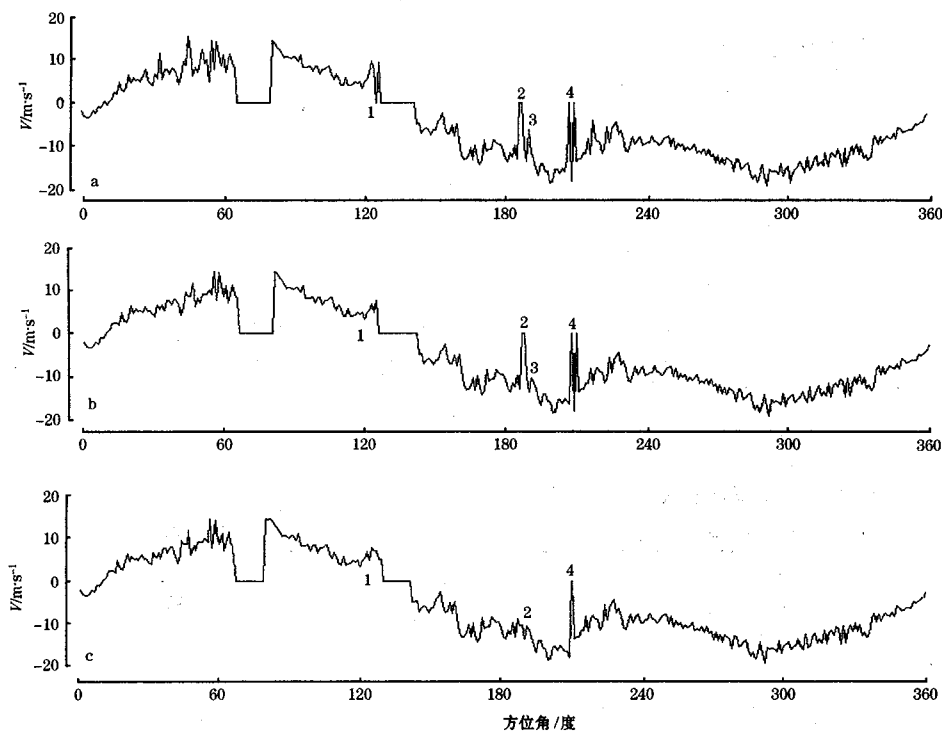


图3 径向速度随方位的分布

a 是第 200(50km)距离圈数据随方位的分布, b、c 分别是方法 A 和方法 B 补缺后的数据随方位的分布图

补的点数尽可能多,而不必担心影响整体数据结构。

(3) 根据 714SDN 多普勒雷达在体扫过程中,不同扫描仰角的 Nyquist 速度差别大的特点,采用大小变化的统计窗口,以保证数据分布在各个统计区间内的平均数差别不大。

(4) 在补缺测值时,仅利用 $\chi_{\max}$ 对应区间的中值来补缺测值有时效果不明显,进一步利用次最大频率 $\chi_{\max 1}$ 所对应区间的中值,可以更加有效地填补缺测点。

参考文献

- 1 陶祖钰,孙丽娟. 多普勒天气雷达 VAD 技术的扩展应用. 中尺度气象文集. 北京:气象出版社,1993:12~19.
- 2 陶祖钰. 单 Doppler 速度场反演风矢量场的 VAP 方法. 气象学报,1992,50(1):81~90.
- 3 周建芬,周文贤,曾西平. 多普勒雷达径向速度资料的预处理方法. 南京气象学院学报,1994,17(2):213~218.
- 4 白洁,陶祖钰. 多普勒雷达风场反演 VAD 方法的资料预处理. 应用气象学报,2000,11(1):21~26.
- 5 梁海河,张沛源,葛润生. 多普勒天气雷达风场退模糊方法的研究. 应用气象学报,2002,13(5):591~599.

Preprocess of 714SDN Doppler Weather Radar Data with K-Neighborhood Frequency Method

Wang Jun

(Shandong Meteorological Institute, Jinan 250031)

Abstract

A K-Neighborhood Frequency method is used to preprocess the Doppler weather radar data and its effect is analyzed, some significant conclusions are educed. In addition, the method is modified according to practical need, so its applied effect is better.

Key Words: K-neighborhood frequency method data preprocessing Doppler weather radar