

王萍,王琮,王迪,2018. 低空急流识别及急流轴自动绘制方法研究[J]. 气象,44(7):952-960.

低空急流识别及急流轴自动绘制方法研究^{*}

王萍 王琮 王迪

天津大学电气自动化与信息工程学院,天津 300072

提 要: 低空急流对强对流天气的预报有重要意义。针对目前低空急流主要由手工绘制,存在效率低、易受主观因素影响的问题,本文基于 MICAPS 高空全要素填图数据中的探空站风场信息,提出了一种低空急流自动识别及急流轴自动绘制算法。算法从急流轴定义出发,依次从风速、风向、探空站分布、中轴线位置等多个角度对低空急流轴进行检测。经过传递闭包聚类、急流轴关键点提取、不同风向急流轴关键点归并、急流轴平滑等步骤,实现了低空急流自动识别及急流轴的自动绘制。测试表明,在识别的基础上自动绘制的急流轴具有位置准确、形态自然、能完整反映急流水汽输送路径、适应复杂环境低空急流等特点。在 291 组测试数据中未发现空报,准确击中率达到 94.96%。

关键词: 低空急流,自动识别,传递闭包聚类

中图分类号: P412

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2018.07.011

Research on Low-Level Jet Identification and Automatic Drawing Method

WANG Ping WANG Cong WANG Di

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072

Abstract: Low-level jet is important for predicting severe convective weather. At present, the identification of low-level jet is conducted mainly by handwork, which brings the problem of low efficiency, easily influenced by subjective factors. So, based on the wind field data of the sounding stations in MICAPS, we propose an automatic low-level jet identification and drawing algorithm in this paper. The algorithm is based on the definition of the low-level jet axis, and detects the low-level jet axis from several aspects in terms of wind speed, wind direction, sounding station distribution, and the central axis position. Then after the steps of transitive closure clustering, key points extracting, low-level jet axis merging and the axis smoothing, the automatic identification and drawing of low-level jet are achieved. The test result shows that the jet axis, which is automatically drawn, has the characteristics of accurate position and natural shape. Besides, it could reflect the transport path of water vapor in jet, and adapt to the complex environment of low-level jet. In the 291 test data, the identification rate reaches 94.96% and false alarm is not found.

Key words: low-level jet, automatic identification, transitive closure clustering

引 言

600 hPa 以下出现的窄而强的气流称为低空急

流(朱乾根,2007)。低空急流与强对流天气的形成有着密切的关系,它可以为强降水输送充足的水汽并提供动力条件(孙淑清,1979),是预报强对流天气的重要依据。因此,自20世纪30年代低空急流

^{*} 公益性行业(气象)科研专项(GYHY201406004)资助

2017年7月7日收稿; 2017年9月3日收修定稿

第一作者:王萍,主要从事计算机视觉、模式识别应用研究. Email: wangps@tju.edu.cn

这一概念被提出起便受到人们的广泛关注。王东阡等(2016)使用温度资料和再分析资料分析指出了低空急流对中国梅雨期雨量和雨期时长的影响。冯文等(2015)指出了低空急流对中国海南后汛期特大暴雨的促进作用。李进等(2015)利用ECMWF的全球再分析资料指出了高低空急流耦合对暴雨发生的重要作用。Higgins et al(1997)和 Weaver and Nigam(2008)分别利用高分辨率再分析资料研究了美国大平原地区低空急流对夏季降水和水汽输送的作用。Qian et al(2004)和 Zhao(2012)也分别利用中尺度模式研究了中国梅雨期强降水发生过程中潜热释放和低空急流的相互作用以及局地地形对低空急流的调制作用。

气象业务中,低空急流通常借助 MICAPS 平台人机交互绘制,效率低且易受主观因素影响。对此,金宏忆等(2008)从多普勒雷达径向速度图入手尝试进行低空急流的自动识别;王毅等(2013)基于 ECMWF 模式细网格风场资料,使用归并算法提取急流轴并进行多项式拟合,将低空急流自动识别出来。基于以上方法识别出的低空急流轴或者受雷达探测范围所限,仅能识别出低空急流的小局部,或者仅适于方向平缓的急流类型。

中国境内的低空急流主要以偏南向急流为主,存在方向缓变(或不变)及方向急变两种类型,通常会延绵上千甚至几千千米。基于探空站风场信息(850 hPa),本文提出一种能适应不同类型低空急流自动识别算法。该算法首先利用传递闭包聚类法形成满足风向条件的探空站聚类,每个子聚类与一个急流的全部或局部相关联。然后在每个子类中提取急流轴关键点,归纳出局部急流区域连接规则,给出全部关键点的平滑急流轴自动绘制方法。

1 低空急流自动识别算法

低空急流的分布特点一定会得到多个测站的数据支持,这些测站点虽非均匀布局,但它们覆盖的区域形同急流分布,同时,各站数据受到风速和风向的一致性 or 相似性制约。本文提出的从点(各测站)到面(急流域)、再到线(急流轴)的低空急流自动识别算法流程主要思想如图 1 所示。

1.1 探空站数据筛选

面对全部境内探空站,仅有一部分探空站能联

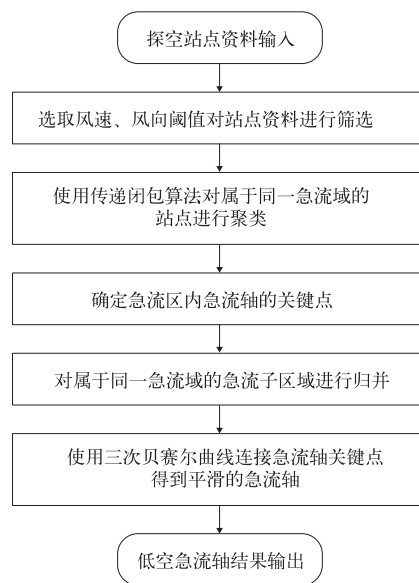


图 1 低空急流自动识别算法流程图

Fig. 1 Flow chart of low-level jet automatic identification algorithm

合给出低空急流的可用信息,它们之间满足以下约束条件:

(1) 在 850 hPa 等压面上,连续两个以上测站风速超过 $12 \text{ m} \cdot \text{s}^{-1}$ 。

(2) 中国境内的低空急流主要以偏南风急流为主。

用满足以上条件的探空站勾勒出大风区,并在大风区的几何中心分析低空急流轴。为此,设探空站筛选条件为:

$$(\text{风速} \geq \alpha_1) \cup (\text{风向} = \text{南风} \pm \alpha_2)$$

为了更加完整地体现水汽输送路径,对南部沿海地区将阈值 α_1 略微降低到 $10 \text{ m} \cdot \text{s}^{-1}$,其余设为 $12 \text{ m} \cdot \text{s}^{-1}$;另外设 $\alpha_2 \leq 90^\circ$ 。

1.2 急流区聚类

急流区是狭长的带状,我们称同属这样的带状区域的探空站为一个聚类。例如在图 2 中西南方向的急流轴周边,有 19 个探空站(图 2 中用椭圆标记的站点)位于由此急流所代表的带状区域内。不难分析出,在这 19 个探空站中,相邻探空站风向相近,而相距较远(例如区域两端)处探空站风向已相差甚远。因此,最常用的 K 均值聚类法没有能力将它们聚成一类;再者,在一个较大的范围内,是否存在低空急流,存在几条低空急流是未知的,这就使得很多以聚类数为先验的聚类算法失去了前提条件。

传递闭包聚类的主要思想是基于两两样本间的相似性来决定某样本是否具有进入当前子类的资格(Rosen, 2006; 陈显强, 2004)。通过选择合适的相似性度量及其使用规则, 一个样本属于哪一类, 仅需计算该样本与其相邻样本的相似度, 并被带入符合相似性度量阈值的相邻样本所属的一类。显然这种聚类思想符合急流区域内探空站间风向的渐变关系特点。

1.2.1 传递闭包聚类算法

经过筛选, 得到满足风向、风速要求的探空站 $P_1, P_2, P_3, \dots, P_n$ 。

定义 P_i 和 P_j 间的相似性度量 $\theta(i, j)$ 和 $D(i, j)$ 。其中 $\theta(i, j)$ 表示 P_i, P_j 坐标连线与这两探空站风向平均值的角度差。 $D(i, j)$ 表示 P_i, P_j 两探空站间的距离。

定义布尔关系矩阵 $\mathbf{R}_{n \times n} = [r_{ij}]_{n \times n}$, 其中 r_{ij} 非 0 即 1。

设布尔矩阵 $\mathbf{A}_{n \times n} = [a_{ij}]_{n \times n}$ 和布尔矩阵 $\mathbf{B}_{n \times n} = [b_{ij}]_{n \times n}$, 定义这两个布尔矩阵的积运算为 $\mathbf{C}_{n \times n} = \mathbf{A}_{n \times n} \times \mathbf{B}_{n \times n} = [c_{ij}]_{n \times n}$, 其中 $c_{ij} = \bigvee_{k=1}^n (a_{ik} \wedge b_{kj})$ 。特别的, $\mathbf{A}_{n \times n}$ 与自身的布尔积称为二次布尔幂, 记作 $\mathbf{A}_{n \times n}^{[2]}$ 。

传递闭包聚类算法步骤如下:

步骤 1: 生成一个布尔矩阵 $\mathbf{R}_{n \times n} = [r_{ij}]_{n \times n}$ 。在 P_i, P_j 均为东南风或者均为西南风的前提下, 对给定阈值 δ 和 d , 若 $\theta(i, j) \leq \delta$ 且 $D(i, j) \leq d$, 则令 $r_{ij} = 1$; 否则, 令 $r_{ij} = 0$ 。

步骤 2: 对布尔矩阵 $\mathbf{R}_{n \times n} = [r_{ij}]_{n \times n}$ 执行 p (p 为整数, $p \geq 2$) 次布尔幂运算得到 $\mathbf{R}_{n \times n}^{[p]}$ 。若 $\mathbf{R}_{n \times n}^{[p]}$ 为全 1 矩阵或 $\mathbf{R}_{n \times n}^{[p]} = \mathbf{R}_{n \times n}^{[p-1]}$, 聚类结束。否则令 $p \rightarrow p+1$, 重复执行步骤 2。

步骤 3: 分析 $\mathbf{R}_{n \times n}^{[p]}$, 若某一行中标记为 1 的元素大于两个, 则认为该行中所有标记为 1 的探空站组成一个有效聚类, 提取所有有效聚类, 将其中风向为东南风的探空站聚类记作 Ω_{ek} ($k=1, 2, 3, \dots, m$), 风向为西南风的探空站聚类记作 Ω_{uk} ($k=1, 2, 3, \dots, n$)。

图 3 为对 2009 年 7 月 23 日 08 时数据执行传递闭包聚类算法后得到的结果。其中, 矩阵 \mathbf{R} 的维度 $n=8$, 表明数据筛选后满足风向风速条件的探空站有 8 个, 对这 8 个探空站生成的布尔矩阵进行 3 次布尔幂运算后变为全 1 矩阵, 因此 8 个探空站构成一个聚类, 即图 3b 中用椭圆标记的站点。图中还

给出了这个聚类区外包椭圆及其长轴。在此称聚类区外包椭圆的长轴为该区域的主方向, 并称长轴明显大于短轴的情形为主方向显著。可以看出, 经传递闭包得到的图 3 聚类区域的主方向显著且与探空站风向趋于一致, 这是相似性度量 $\theta(i, j)$ 和 $D(i, j)$ 对探空站的距离约束以及探空站相对位置与风向之间约束的联合反应。

在聚类过程中为了尽可能找到支持急流的探空站, 需要适当将阈值 δ, d 条件放宽, 这可能会导致聚到一起的探空站区域主方向不十分显著的情形。例如对 2009 年 7 月 20 日 08 时数据进行聚类的结果(图 4)中, 探空站风向均为西南方向, 但 9 个探空站的外包椭圆趋于圆形, 主方向不显著, 这样会给之后准确定位急流轴带来困难。为此, 对聚类结果再次执行传递闭包聚类算法。

1.2.2 二次传递闭包聚类算法

仅保留图 3b 和图 4 中需聚类处理的探空站得到图 5a 和 5b。将这些探空站分别在 $\pm 45^\circ$ 方向上投影, 不难发现, 图 5a 所示的聚类集较理想(探空站分布主方向显著且与风向趋于一致), 其 $l(pr_{45i}) = \gamma_1 l(pr_{-45i})$ 且 γ_1 往往大于 1.5; 图 5b 给出的聚类集不甚理想(探空站分布主方向不显著), 其 $l(pr_{45i}) \approx l(pr_{-45i})$ 。设这种不理想聚类集中, 实际存在着密集站点子集合(见图 5b 椭圆内的站点), 由它们形成的子区域, 其主方向与风向一致。从分布上看, 这些密集站点之间关系较为紧密, 而其他站点与这些站点之间的关系则较为松散, 导致这一现象的原因在于进行首次传递闭包聚类时, 所选择的相似性度量阈值 δ 和 d 较为宽松。为此, 可以当 $l(pr_{-45i})$ 约等于甚至小于 $l(pr_{45i})$ 的情形出现之后, 收紧阈值 δ, d , 对该局部区域进行再次传递闭包聚类, 将首次聚类给出的大风区分解成若干个关系紧密的子域, 并保留其中站点数最多、最能反映急流趋势的那个子域。

算法步骤如下:

步骤 1: 假设 Ω_i 为第 i 个聚类, P_{ij} ($j=1, 2, 3, \dots, n$) 为 Ω 中的第 j 个探空站。 P_{ij} 在 45° 方向上投影长度记为 pr_{45ij} , P_{ij} 在 -45° 方向上投影长度记为 pr_{-45ij} , 则 $l(pr_{45i}) = \max\{pr_{45ij} - pr_{45ik}\}$, $l(pr_{-45i}) = \max\{pr_{-45ij} - pr_{-45ik}\}$ ($j, k=1, 2, 3, \dots, n$ 且 $j \neq k$)。并以此计算 $\gamma = \frac{l(pr_{45i})}{l(pr_{-45i})}$

步骤 2: 给定 $\gamma_1 = 1.5$ 。若 $\Omega_i \in \Omega_{ek}$, 当 $\gamma > \frac{1}{\gamma_1}$ 时, 执行步骤 3, 否则算法结束; 若 $\Omega_i \in \Omega_{uk}$, 当 $\gamma < \gamma_1$

时,执行步骤 3,否则算法结束。

步骤 3:对于 $P_{ij} (j=1,2,3,\cdots,n) \in \Omega$, 给定新的阈值 δ_{new} , 重新进行传递闭包聚类, 得到针对 Ω_i 的二次聚类结果 $\Omega_{i1}, \Omega_{i2}, \Omega_{i3}, \cdots, \Omega_{im}$ 。

步骤 4:选择 $\Omega_{i1}, \Omega_{i2}, \Omega_{i3}, \cdots, \Omega_{im}$ 中站点数最多的子类代替原来的第 Ω_i 类。

为确定二次传递闭包聚类时相似性度量 θ 的收紧程度, 选用 2009 年 3—8 月全部数据执行首次传递闭包聚类算法, 筛选出所有聚类区域主方向不显著的聚类结果, 将其中的所有探空站进行“保留”和“去除”的人工标注。并计算“保留”类站点的 θ 和

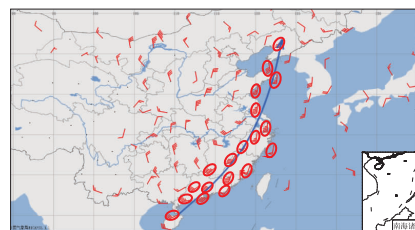


图 2 基于 2009 年 4 月 19 日 20 时 850 hPa 探空站风场数据的低空急流轴

Fig. 2 The low-level jet axis at 850 hPa wind field from sounding stations at 20:00 BT 19 April 2009

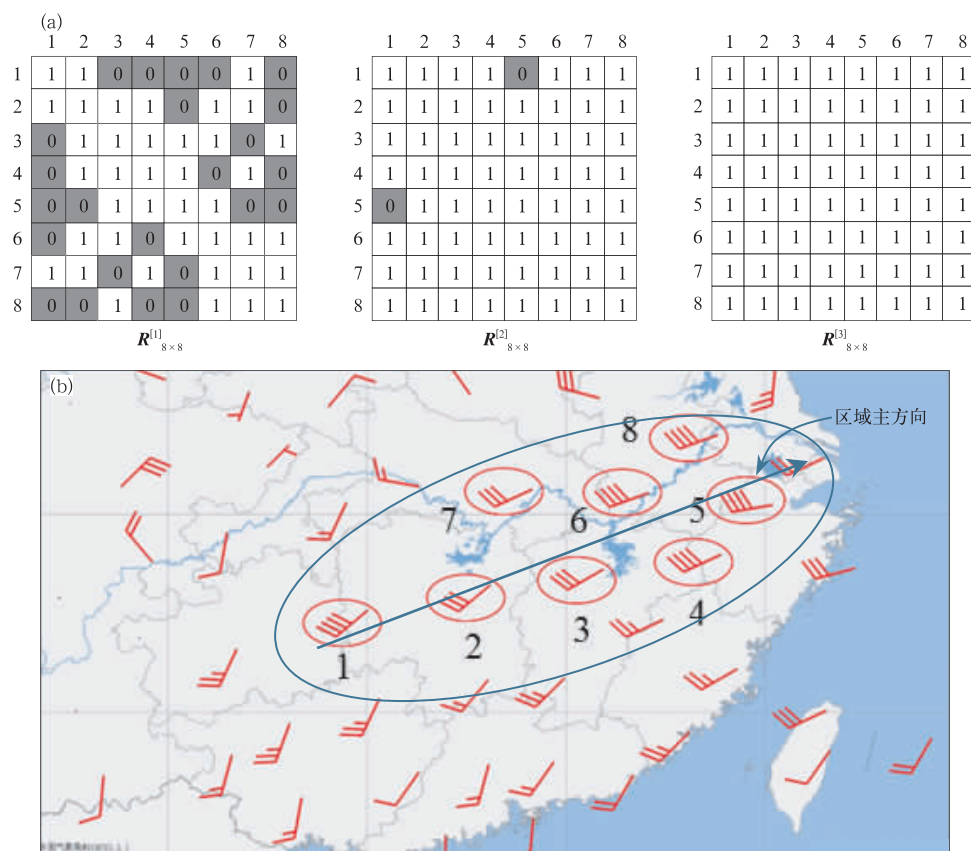


图 3 对 2009 年 7 月 23 日 08 时数据, 在 $\delta=60^\circ$ 、 $d=600$ km 条件下执行传递闭包聚类算法的结果

(a) 1~3 次布尔矩阵幂, (b) 探空站聚类结果示意

Fig. 3 Result of the transitive closure clustering algorithm performed under the condition of $\delta=60^\circ$ and $d=600$ km based on the data of 08:00 BT 23 July 2009

(a) the first three power operations of Boolean matrices,

(b) clustering results at the sounding stations

“去除”类站点的 θ , 形成 θ 的统计分布(图 6)。可以看出, “保留”类站点间的 θ 值主要分布在 $0^\circ \sim 30^\circ$, 且分布较为均匀。而“去除”类站点间的 θ 值集中分

布在 $40^\circ \sim 55^\circ$ 。所以用于二次聚类的 θ 的阈值 δ_{new} 可选在 $30^\circ \sim 40^\circ$ 。图 7a 为对图 4 的 9 个站点在 $\delta_{new}=30^\circ$ 、 $d=600$ km 条件下进行二次传递闭包聚类的



图 4 同图 3b, 但为 2009 年 7 月 20 日 08 时数据
(探空站分布的主方向不显著)
Fig. 4 Same as Fig. 3b, but for
08:00 BT 20 July 2009
(Main direction of the distribution
is not significant)

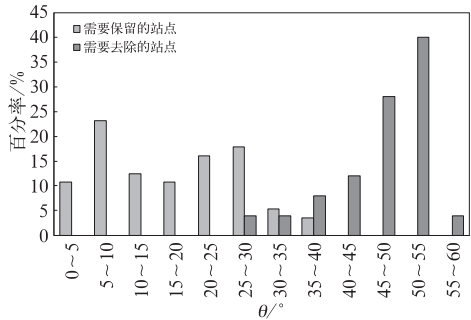


图 6 二次聚类中需要“保留”类站点
和需要“去除”类站点间的分布
Fig. 6 Distribution of stations in the
secondary clustering to be
“reserved” or “removed”

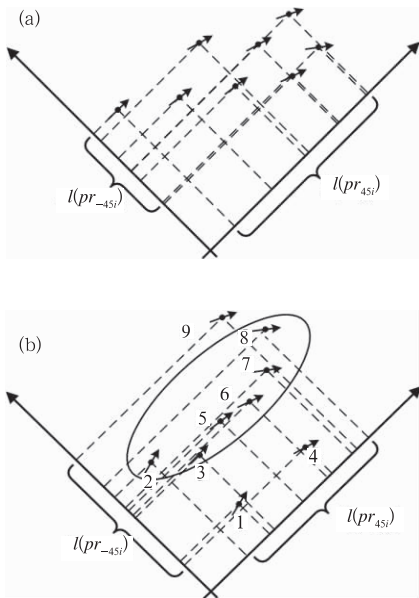


图 5 2009 年 7 月 23 日 08 时(a), 20 日 08 时(b)
聚类结果在 $\pm 45^\circ$ 方向上进行的投影
Fig. 5 Clustering results projected in the
direction of $\pm 45^\circ$ at 08:00 BT 23 (a)
and 08:00 BT 20 (b) July 2009

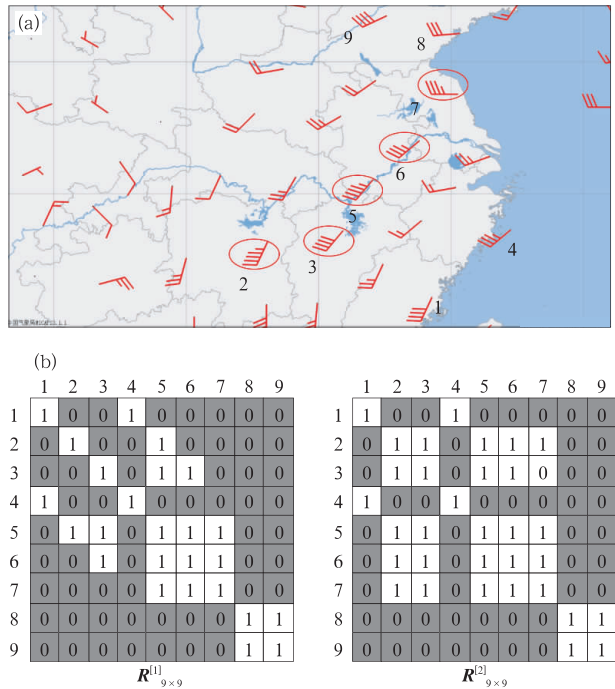


图 7 对图 4 站点的再次传递闭包聚类
($\delta_{\text{new}} = 30^\circ, d = 600 \text{ km}$) 的结果
(a) 最终保留的子聚类,
(b) 二次聚类中布尔矩阵幂变化

Fig. 7 Performing the secondary transitive closure
clustering algorithm ($\delta_{\text{new}} = 30^\circ, d = 600 \text{ km}$)
for the result of Fig. 4
(a) the final retention of sub-clusters,
(b) boolean matrix transformation
in secondary clustering

结果。对比图 4 可以看出,图 7a 中剔除了首次聚类中较为松散的 8、9、1、4 号站点,使二次聚类给出的探空站分布主方向与其风向趋于一致。从布尔矩阵 $R_{9 \times 9}$ 可以看出,因为阈值的收紧,切断了 6 号探空站与 8、1 号探空站的关系,继而剔除了 9 和 4 号探空站。

1.3 提取绘制急流轴的关键点

二次传递闭包聚类后,得到若干组沿风向呈带

状分布的探空站点集合,急流轴应从它们中间穿过。由于探空站地理位置分布的不均匀性,不宜采用简单的取中法。下面给出一种根据站点数量及站点之

间相对关系确定急流轴关键点的算法。算法步骤如下:

步骤1:假设 Ω_i 为第 i 个聚类, $P_{ij} (j=1, 2, 3, \dots, n) \in \Omega_i$ 。若 $\Omega_i \in \Omega_{ek}$, 将 P_{ij} 按照 $pr_{-45ij} (j=1, 2, 3, \dots, n)$ 由小到大进行排序; 若 $\Omega_i \in \Omega_{uk}$, 将 P_{ij} 按照 $pr_{45ij} (j=1, 2, 3, \dots, n)$ 由小到大排序。当 $n=2$ 时执行步骤2, $n=3$ 时执行步骤3, 否则执行步骤4。

步骤2:直接将 P_{i1} 、 P_{i2} 作为第 Ω_i 类中轴的关键点, 如图 8a 所示。

步骤3:若 $P_{i1}P_{i2}$ 与 $P_{i1}P_{i3}$ 的夹角大于 45° , 则将 P_{i1} 、 P_{i2} 的加权中点 P_{i0} 与 P_{i3} 一起作为第 Ω_i 类中轴的关键点, 如图 8b 所示; 若 $P_{i1}P_{i2}$ 与 $P_{i2}P_{i3}$ 的夹角大于 45° , 则将 P_{i2} 、 P_{i3} 的加权中点 P_{i4} 与 P_{i1} 一起作为第 Ω_i 类中轴的关键点; 除此之外, 则直接将 P_{i1} 、 P_{i3} 作为第 Ω_i 类中轴的关键点。

步骤4:

(1) 求取 $P_{i1}, P_{i2}, P_{i3}, \dots, P_{in} (n \geq 4)$ 的凸包 (Graham, 1972), 记作 A_i 。

(2) $P_{i1}, P_{i2}, P_{i3}, \dots, P_{in}$ 中纬度最大的站点记作 $P_{iy_{max}}$, 纬度最小的站点记作 $P_{iy_{min}}$ 。取 $P_{iy_{min}}$ 与 P_{i1} 的

加权中点 S 作为第 Ω_i 类中轴起始关键点; 取 $P_{iy_{max}}$ 与 P_{in} 的加权中点 E 作为第 Ω_i 类中轴的终止关键点。

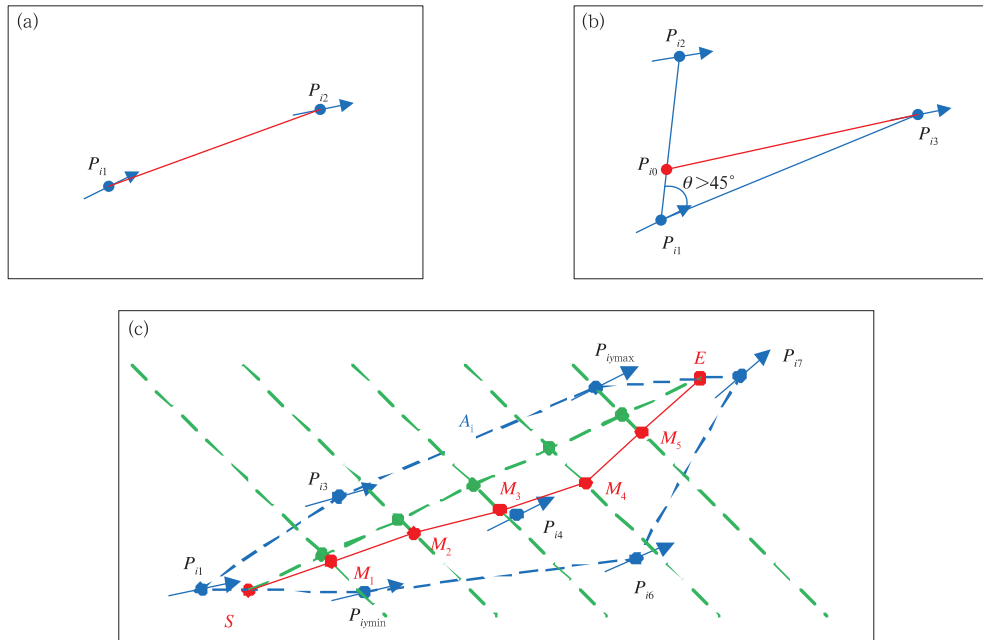
(3) 将线段 SE 五等分, 等分点分别记为 Q_1, Q_2, Q_3, Q_4 。若 $\Omega_i \in \Omega_{ek}$, 则分别过 $Q_i (i=1, 2, 3, 4)$ 做正北偏东 45° 的直线; 若 $\Omega_i \in \Omega_{uk}$, 则分别过 $Q_i (i=1, 2, 3, 4)$ 做正北偏西 45° 的直线。直线分别交 A_i 于 Q_{i1}, Q_{i2} , 记 Q_{i1}, Q_{i2} 的中点为 $M_i (i=1, 2, 3, 4)$ 。

(4) 将 S, M_1, M_2, M_3, M_4, E 作为第 Ω_i 类中轴的关键点, 如图 8c 所示。算法结束。

按照以上步骤, 分别找到 $\Omega_{ek} (k=1, 2, 3, \dots, m)$ 中每一子类的中轴 $L_{ek} (k=1, 2, 3, \dots, m)$ 的关键点, 及 $\Omega_{uk} (k=1, 2, 3, \dots, n)$ 中每一子类的中轴 $L_{uk} (k=1, 2, 3, \dots, n)$ 的关键点。

1.4 急流轴关键点归并

行进过程中方向发生较大改变的低空急流下的探空站, 一般会被传递闭包算法至少聚成两个子类 (参见图 9), 因此需要将原本属于一条急流轴但又不属于一个聚类子集的关键点进行合并。



首先定义不同聚类子集的连接规则:设两个聚类子集区域中轴为 L_1 和 L_2 ,记 L_1 中的最后一个关键点为 L_{1n} , L_2 中的第一个关键点为 L_{21} 。计算 L_{1n} 、 L_{21} 的距离 $d_{l_{1n},l_{21}}$,若 $d_{l_{1n},l_{21}} \leq \delta$,则认为 L_1 、 L_2 同属一个急流区,将 L_1 和 L_2 的关键点集合的并集作为新的关键点集合,并在连接处以 l_{1n} 、 l_{21} 的中点代替两点作平滑处理。若 $d_{l_{1n},l_{21}} > \delta$,则在沿 L_1 方向 $\pm 15^\circ$ 、半径 δ 的范围内搜索是否存在满足本急流轴风向条件,但风速略小的探空站点。若存在,则将该站点作为关键点加入 L_1 ,并按照上述连接规则判断并归并两关键点集,否则不进行关键点集的归并。

急流轴关键点归并算法步骤如下:

步骤 1:取 $L_{ei} \in L_{ek} (k=1,2,3,\cdots,m)$,依次判断 L_{ei} 与 $L_{uj} \in L_{uk} (k=1,2,3,\cdots,n)$ 是否满足上述归并规则。若满足,则将 L_{ei} 与 L_{uj} 的关键点归并,记作 L_k ,并从 L_{ek} 中删除 L_{ei} ,从 L_{uk} 中删除 L_{uj} 。否则将 L_{ei} 记作 L_k ,并从 L_{ek} 中删除 L_{ei} ,对 L_{uk} 不作处理。执行此步骤直到各 L_{ek} 均为空集。

步骤 2:若 $L_{uk} (k=1,2,3,\cdots,n)$ 为空集,则结束算法, $L_k (k=1,2,3,\cdots)$ 即为所求;否则取 $L_{uj} \in L_{uk} (k=1,2,3,\cdots,n)$,依次判断 L_{uj} 与 L_k 是否满足上述连接规则,若满足,则将 L_{uj} 与 L_k 归并以代替之前的 L_k ,并从 L_{uk} 中删除 L_{uj} 。否则将 L_{uj} 记作新的 L_k ,并从 L_{uk} 中删除 L_{uj} 。进行此步骤直到各 L_{uk} 均为空集。 $L_k (k=1,2,3,\cdots)$ 即为所求。

1.5 急流轴平滑

经过以上步骤,得到低空急流轴 $L_k (k=1,2,3,\cdots)$ 的一系列关键点。为使低空急流轴展现为一条光滑的曲线,用三次贝塞尔曲线将 $L_k (k=1,2,3,\cdots)$ 中的关键点平滑地连接起来(王家润等, 2010)。

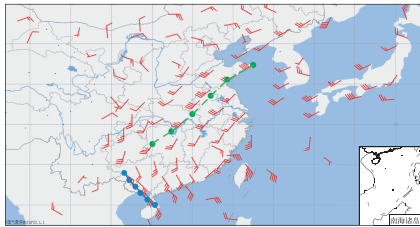


图 9 对 2009 年 7 月 17 日 08 时数据, 得到两个子类的区域中轴的关键点
Fig. 9 The key points of the axis in two subclasses obtained from the data at 08:00 BT 17 July 2009

图 10 给出了基于 2009 年 7 月 20 日 08 时数据的平滑急流轴。

2 算法应用

本算法基于 Visual Studio 2015 平台使用 C++ 编程语言实现,输入数据为 MICAPS 第 2 类数据中的 850 hPa 探空站风场信息,输出为低空急流轴的经纬度坐标。在 Intel i5-4590 硬件平台下整个算法运行平均耗时 0.0249 s。

使用 2009 年 3 月 12 日至 8 月 27 日及 2012 年 5 月 31 日至 12 月 18 日 08—20 时中的 291 组数据对算法的准确性和鲁棒性进行测试。并将测试结果交予气象专家核实,若识别结果与人工分析结果一致则定义为准确检出;若识别结果与人工分析结果不完全一致但自动识别结果能大致反映低空急流趋势,则定义为偏差;若自动识别存在低空急流但人工分析不存在低空急流,则认为空报;若人工分析存在低空急流但本算法并未识别出,则认为漏报。在上述 291 组数据中,有位于中国地区的低空急流 139 处,准确检测出 132 处,有偏差的为 4 处,漏报 3 处,空报 0 处,检出率达到 97.84%,准确击中率为 94.96%(表 1)。

图 11 给出了一组低空急流轴识别结果示例,其中图 11a 为识别并绘制出的单一风向的低空急流

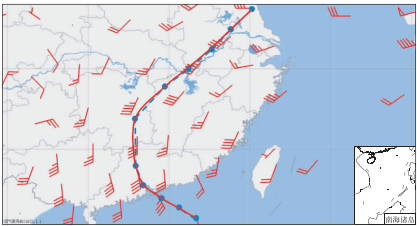


图 10 对 2009 年 7 月 20 日 08 时数据, 经过急流轴提取并平滑的结果

Fig. 10 Result from the data at 08:00 BT 20 July 2009 after the extraction and smoothing of jet axis

表 1 算法性能测试统计结果

Table 1 Result of algorithm performance test

样本数	击中率		漏报率 /%(样本数)	空报率 /%
	准确击中率 /%(样本数)	偏差击中率 /%(样本数)		
139	94.96%(132)	2.88%(4)	2.16%(3)	0

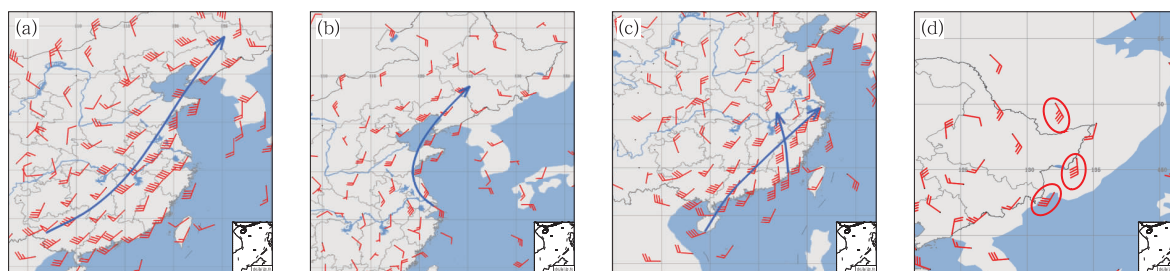


图 11 2009 年 3 月 21 日 08 时(a),2012 年 6 月 3 日 20 时(b),2009 年 4 月 19 日 08 时(c),
和 2009 年 6 月 10 日 08 时(d)低空急流轴检测结果

Fig. 11 Result of low-level jet identification and drawing at 08:00 BT 21 March 2009 (a),
20:00 BT 3 June 2012 (b), 08:00 BT 19 April 2009 (c), and 08:00 BT 10 June 2009 (d)

轴,位置准确,形状自然。图 11b 为风向发生改变的低空急流轴,位置准确,在连接处曲线平滑,没有明显连接痕迹。图 11c 为 4 例低空急流识别发生偏差的典型代表,其特征为:一块区域中,同时存在风向临界于东南风与西南风之间的站点。这样会对传递闭包聚类时的风向划分带来影响。从而影响急流轴的起止位。造成以距离为度量的急流轴关键点归并算法失效,导致在同一区域内识别出多条急流轴。但此种情况并不会使识别出的急流轴过于偏离急流区域的中轴,识别出的多条急流轴仍能大致反映急流的形态和位置。图 11d 为 3 例漏报的典型代表,其特征为:发生漏报的低空急流长度较短(由 3 个探空站组成),且发生风向改变。造成漏报的原因:对于风向发生改变的低空急流,本算法首先按不同风向进行聚类,再将其关键点进行归并。发生漏报的低空急流虽然总探空站数满足低空急流要求(大于 2 个探空站),但按风向聚类时却由于属于每个风向的站点数过少而不满足条件,最终导致漏报。

3 结论与讨论

本文根据低空急流定义及其特点,首先,以相邻探空站位置分布与其风向的一致性关系为依据,对探空站进行传递闭包聚类分析,得到多个聚类子集。针对同属一个聚类子集但分布不均的探空站,提出探空站凸包法,得到基于这些探空站的一片区域,并对该区域求取满足风向约束的关键点。针对急流行进过程中可能产生的风向改变的情形,提出一种距离约束下的急流轴关键点归并算法,最后选用贝塞尔曲线模型实现基于关键点的低空急流的自动绘制。

经对 291 组数据检验,结果经人工核实,得出以下结论:

(1) 算法中的两次传递闭包聚类可以在最大程度上识别出急流完整路径的基础上,圈定急流的准确区域范围,达到较为完整和准确地反映水汽输送路径的目的。

(2) 算法同时适用于单一风向的低空急流和方向发生改变的低空急流的自动识别,算法鲁棒性强。

(3) 绘制出的急流轴均处于急流区域的中轴位置且表现为光滑的曲线,与手工绘制的结果相符。

(4) 在 291 组数据中,准确检出 139 例低空急流中的 132 例,准确击中率达到 94.96%,未发现空报。

(5) 本算法使用 C++ 语言实现,便于编译成库文件以提供给气象部门进行业务应用。同时本算法适用性较强,稍作改动亦可用于格点数据中的低空急流的识别。

本文的算法还存在一些缺陷,(1)当支持方向急转型急流的探空站过少时,会出现漏检。(2)当在一个区域内满足风向条件的探空站过于密集但同时存在两类风向子集时,可能出现急流轴关键点归并算法失效或同一区域识别出多条急流轴的情况。这需要在更大规模样本的支持下,发现问题的规律,对本文方法进行改进和补充。

参考文献

- 陈显强,2004. 二元关系的传递性和传递闭包探讨[J]. 数学的实践与认识,34(9):135-137.
- 冯文,符式红,赵付竹,2015. 近 10 年海南岛后汛期特大暴雨环流配置及其异常特征[J]. 气象,41(2):143-152.
- 金宏忆,顾松山,王珊珊,等,2008. 多普勒雷达低空大风轴线自动识别方法的探讨[J]. 南京气象学院学报,31(5):702-710.

- 李进,丁婷,赵思楠,等,2015. 2013年6月7日浙江省中北部暴雨过程诊断分析[J]. 气象,41(10):1215-1221.
- 孙淑清,1979. 关于低空急流对暴雨的触发作用的一种机制[J]. 气象,5(4):8-10.
- 王东阡,王艳姣,崔童,等,2016. 2015年夏季气候异常特征及其成因简析[J]. 气象,42(1):115-121.
- 王家润,赵南松,华文元,等,2010. 分段连续三次Bezier曲线控制点的构造算法[J]. 计算机工程与应用,46(22):190-193.
- 王毅,代刊,黄小玉,2013. 高、低空急流的客观识别及其初步应用[J]. 天气预报技术总结专刊,5(5):46-50.
- 朱乾根,2007. 天气学原理和方法:第4版[M]. 北京:气象出版社:375-381.
- Graham R L,1972. An efficient algorithm for determining the convex hull of a finite planar set[J]. Inf Process Lett,1(4):132-133.
- Higgins R W,Yao Y,Yarosh E S,et al,1997. Influence of the Great Plains low-level jet on summertime precipitation and moisture transport over the central United States[J]. J Climate,10(3):481-507.
- Qian Jianhua,Tao Weikuo,Lau K M,2004. Mechanisms for torrential rain associated with the Mei-Yu development during SCS-MEX 1998[J]. Mon Wea Rev,132(1):3-27.
- Rosen K H,2006. Discrete Mathematics and Its Applications[M]. 6th ed. New York:McGraw-Hill:544-555.
- Weaver S J,Nigam S,2008. Variability of the Great Plains low-level jet:large-scale circulation context and hydroclimate impacts[J]. J Climate,21(7):1532-1551.
- Zhao Yuchun,2012. Numerical investigation of a localized extremely heavy rainfall event in complex topographic area during midsummer[J]. Atmos Res,113:22-39.