0254-6124/2020/40(6)-1014–10 *Chin. J. Space Sci.* 空间科学学报

WANG Ling, YIN Fan. Distinguish small-scale traveling ionospheric disturbances and equatorial plasma bubbles by clustering algorithm (in Chinese). *Chin. J. Space Sci.*, 2020, **40**(6): 1014-1023. DOI:10.11728/cjss2020.06.1014

利用聚类算法区分小尺度电离层行扰事件与赤道等离子体泡事件*

汪 领 尹 凡

(武汉大学电子信息学院 武汉 430072)

摘 要 利用 Swarm 卫星 2015 年 1 月 1 日至 2019 年 12 月 31 日的 50 Hz 高频磁场数据, 根据阈值判断垂直于主磁场方向的扰动, 对磁纬 45°N－45°S 之间的小尺度电离层行扰事件进行探测. 为避免混淆而产生的干扰, 可以根据阈值判断平行于主磁场方向是否发生扰动, 从而排除典型的赤道等离子体泡事件. 但对于较弱的赤道等离子体泡事件, 扰动阈值判断无效. 为避免弱赤道等离子体泡事件的污染, 根据小尺度电离层行扰事件和赤道等离子体泡事件在不同参数空间中的密度分布差异, 利用基于密度的聚类算法将赤道等离子体泡事件进一步甄别提取. 结果表明, 聚类算法能够有效地将赤道等离子体泡事件从小尺度电离层行扰事件中甄选出来, 并使小尺度电离层行扰事件聚类与赤道等离子体泡事件聚类形成清晰的边界. 由聚类算法导出的弱赤道等离子体泡事件主要分布在磁纬 15°N－15°S, 地理经度 20°－60°W, 月份 10 至 3 月之间, 并且在 20:00 MLT－24:00 MLT 存在高发生率, 同时依赖于太阳活动, 这也验证了前人的相关研究结果.

关键词 聚类算法, 电离层, 小尺度电离层行扰, 赤道等离子体泡

中图分类号 P352

Distinguish Small-scale Traveling Ionospheric Disturbances and Equatorial Plasma Bubbles by Clustering Algorithm

WANG Ling YIN Fan

(*School of Electronic Information, Wuhan University, Wuhan* 430072)

Abstract By analyzing the 50 Hz high-frequency magnetic field data by Swarm from January 2015 to December 2019, according to the magnetic disturbances above the threshold in the perpendicular direction to the main field, Small-Scale Traveling Ionospheric Disturbances (SSTID) within ±45°N magnetic latitude have been detected. When detecting SSTID, Equatorial Plasma Bubbles (EPB) can be excluded by determining whether there is magnetic disturbance above the threshold in the

* 国家重点研发计划项目 (2018YFC1503501, 2018YFC1503501-01), 国家自然科学基金重点项目 (41431073) 和国家自然科学基金面上项目 (41474157) 共同资助

2020-01-13 收到原稿, 2020-04-17 收到修定稿

E-mail: wlydjqcg@whu.edu.cn. 通信作者 尹凡, E-mail: yinfan@whu.edu.cn

direction which is parallel to the mean ambient field. However, when the disturbance is weak, the algorithm cannot completely exclude it. According to the different density distribution of SSTID and EPB in parameter space, the EPB can be distinguished by a density-based clustering algorithm. The results demonstrate that the clustering algorithm is very effective on separating EPB and SSTID in different two-dimensional parameter spaces and brings out a clear boundary between the high density region and low density region. The EPB identified by the clustering algorithm shows the strong dependence on solar activity and their statistical features are consistent with the previous study.

Key words　Clustering algorithm, Ionosphere, Small-Scale Traveling Ionospheric Disturbance (SSTID), Equatorial Plasma Bubble (EPB)

0　引言

由于等离子体的不稳定性, 各种从上和从下的能量输入会导致电离层发生扰动现象, 而并不总是保持稳态和均匀分层结构. 波长小于 10 km 的小尺度电离层行扰 (Small-Scale Traveling Ionospheric Disturbance, SSTID) 是一种经常发生在中低纬地区的电离层扰动现象, 被认为是夜间波长几百 km 左右的中尺度电离层行扰 (Medium-Scale Traveling Ionospheric Disturbance, MSTID) 的一种演化, 二者的统计特征基本一致 [1]. 太阳活动平静期的夜间中纬度电离层经常发生 MSTID[2], 这是由电离层 F 层的 Perkins 不稳定和与其相连的偶发 E 层不稳定之间的耦合作用产生的 [3−7]. 与 MSTID 事件导致的中纬度中尺度磁场波动 (Mid-latitude Magnetic Field fluctuation, MMF) 类似, 中低纬度 SSTID 事件也会导致与场向电流相关的小尺度磁场波动 (Small-Scale Mid-Latitude Magnetic Fluctuation, SSMMF).

通常 MMF 以及 SSMMF 在电离层顶部都不伴随电子密度的扰动, 而且磁扰动只存在于主磁场的垂直方向 [1,8]. Yin 等 [1] 研究发现 SSMMF 的统计特征与夜间 MMF[8,9] 的统计特征存在一致性: SSMMF 与 MMF 均主要分布在南北半球磁纬 25°N−45°S 之间, 磁赤道低纬度附近 (磁纬 10°N−10°S 之间) 和南大西洋异常 (South Atlantic Anomaly, SAA) 区域分布极其稀疏; 在季节依赖性上均表现出在夏至季节分布最密集, 冬至其次, 春秋分是事件发生的低谷期. 在磁地方时上, SSMMF 主要分别分布在 08:00 MLT−12:00 MLT 以及 17:00 MLT −23:00 MLT. SSMMF 整体上对太阳活动没有依赖性.

另一种电离层扰动现象是发生在低纬赤道地区的由瑞利泰勒不稳定性导致的等离子体密度不规则结构, 通常被称为赤道扩展 F (Equatorial Spread F, ESF) 或赤道等离子体泡 (Equatorial Plasma Bubble, EPB)[10−12]. EPB 泡内电子密度远低于背景电离层电子密度, 其电子密度耗尽可以达到 90% 以上. Lühr 等 [13,17]、Park 等 [14]、Xiong 等 [15]、Stolle 等 [16]、Yokoyama 等 [18] 使用 CHAMP 卫星观测结果, 发现 EPB 产生的压强梯度驱动的电流和场向电流分别使平行于主磁场和垂直于主磁场方向的磁场强度发生扰动. EPB 主要出现在地理经度 20°−60°W (SAA 区域), 月份 10 至 3 月之间分布最密集, 夏至季节分布非常稀疏, 主要出现在 20:00 MLT−24:00 MLT, 并且依赖于太阳活动强度 [14−18].

中低纬度的 SSTID 事件和 EPB 事件都会导致磁场扰动. 利用星载高精度磁力计捕捉到的扰动信号对两种现象已进行了充分研究. 但在研究这两种不同机制的事件时, 必须对二者加以区分, 否则会产生混淆. 其区分的关键在于: SSTID 事件导致的磁扰动存在于主磁场的垂直方向, 在主磁场的平行方向上则表现平静; 赤道等离子体泡产生的磁场扰动也明显存在于主磁场平行方向. Park 等 [14]、Xiong 等 [15]、Stolle 等 [16]、Lühr 等 [17] 通过分析 CHAMP 卫星磁场数据中平行于主磁场方向上的扰动来确定 EPB 事件. Yin 等 [1] 和 Park 等 [8,9] 分析 CHAMP 以及 Swarm 卫星磁场数据时, 通过垂直于主磁场方向上的扰动来确定 MSTID 和 SSTID 事件, 并且通过排除平行于主磁场方向的磁扰来避免 EPB 事件的

污染. 一般地, 事件性质可以通过不同方向上的时序信号是否存在超过设定阈值的扰动来确定. 但是, 由于 EPB 事件磁扰动的复杂性, Stolle 等 [16] 利用人工判读的方法来验证 EPB 事件. 但对于事件数量巨大的 SSTID 事件, 人工判读显然不合适. Yin 等 [1] 在探测 SSTID 事件时, 虽然通过阈值判定的方法排除了 EPB 事件, 但观察最终得到的 SSTID 事件集合在地理经度、地方时与磁纬等二维参数空间中分布, 仍然可以发现明显的 EPB 事件的分布特征, 即 SSTID 事件集是不纯净的, 受到了未被完全排除的 EPB 事件的污染. 通过分析发现, SSTID 探测算法中仅通过判定平行于主磁场方向上的磁扰来排除 EPB 事件的污染是存在缺陷的. 一个非典型 EPB 事件, 当其平行于主磁场方向上的磁扰持续时间比较短且扰动幅度相对较小 (低于阈值但又不是噪声) 时, 会导致其不被判定为 EPB 事件, 而被误判为 SSTID 事件. 如果降低 SSTID 探测算法中排除 EPB 事件的判断阈值, 又会因为噪声信号产生事件漏报, 而排除了本该探测到的 SSTID 事件; 而升高判断阈值则会将更多的 EPB 污染事件误报为 SSTID 事件, 从而导致 SSTID 数据集污染更严重. 另外, 由于使用的等离子体密度数据的采样频率为 2 Hz, 远低于 50 Hz 磁场数据, 并且非典型 EPB 事件磁扰动及电子密度扰动均较弱, 因此通过观测等离子体密度数据的扰动来排除非典型 EPB 事件也不合适. 但非典型 EPB 事件和典型 EPB 事件的统计属性是一致的. 在 Yin 等 [1] 的统计结果中, 污染性的 EPB 事件地理上明显分布在磁赤道附近的 SAA 区域, 在地方时上则集中分布在 22:00 MLT 附近, 这与已有观测一致. 无论 EPB 污染事件是典型的还是非典型的, 其在参数空间中的统计特征明显与 SSTID 事件不同. 值得研究的是, 是否存在一种能利用这两种事件在不同参数空间中的密度分布差异来对 SSTID 事件集合中未排除的 EPB 事件进行进一步提取、剔除的方法, 利用该方法进行数据处理, 不仅能保证 SSTID 事件集的纯净性, 而且能对未排除的 EPB 事件特征进行深入研究. 基于密度的 DBSCAN (Density-Based Spatial Clustering of Applications with Noise)[19] 聚类算法能够弥补时序信号阈值判定事件算法存在的缺陷, 对 SSTID 事件集合中存在的非典型 EPB 污染事件进行甄别.

本文利用三颗 Swarm 卫星 2015 年 1 月 1 日至 2019 年 12 月 31 日的高精度磁场数据, 探测磁纬 45°N−45°S 之间的 SSTID 事件. 使用了通过判定平行主磁场方向上是否存在超过设定阈值的磁扰动来排除 EPB 事件的算法, 针对使用这种算法的处理结果, 使用基于密度的 DBSCAN 聚类算法对非典型 EPB 污染事件进行进一步甄选. 最后分析验证由聚类算法导出的 EPB 事件的地理经度、磁地方时、季节和对太阳活动依赖性特征.

1 数据来源及处理方法

1.1 数据来源

2013 年 11 月 22 日欧空局 (ESA) 启动了 Swarm 地球探测任务, 将三颗卫星 (Swarm-Alpha, Swarm-Bravo 和 Swarm-Charlie) 成功发射到近地极轨轨道. Swarm 卫星的主要目标是测量迄今为止最完整、最高精度的地球磁场数据. 经过卫星发射和轨道调试, 2014 年 4 月 17 日三颗卫星的飞行轨道最终配置完成. 其中, 相对较低的两颗卫星 (Swarm-Alpha 和 Swarm-Charlie) 以 462 km 的初始高度和 87.35° 的倾角轨道并行飞行. 二者南北方向的距离大约只有 75 km, 东西方向地理经度相差约为 1.5°. Swarm-Bravo 卫星单独飞行在 520 km 的高度上, 其轨道倾角为 87.75°. 本文使用了三颗 Swarm 卫星搭载的矢量场磁强计测量的 50 Hz 高采样率磁场数据和朗缪尔探针测量的 2 Hz 等离子体密度数据 *.

1.2 SSTID 事件探测方法

Swarm 卫星测量的 50 Hz 磁场数据是基于 North-East-Center (NEC) 地理学坐标系, 而 SSTID 事件导致的磁扰动更容易出现在强背景磁场的垂直方向上. 因此, 需要将磁场数据从 NEC 坐标系转换到平均场 (Mean-Field Aligned, MFA) 坐标系. 在 MFA 坐标系中, 磁场分量 B_x, B_z, B_y 构成 MFA 中的三个磁场正交分量, B_z 沿着平均磁场的方向, B_y 垂直于磁子午线并且指向东向 [20].

使用截止周期 1 s 的高通滤波器分别对 B_y 和 B_z 进行滤波处理, 滤波后得到磁场分量中大于 1 Hz 的高频磁场信号 σ_y, σ_z. 当 σ_y 至少存在连续 4 个磁场扰动点幅值超过 ±0.2 nT 时, 一个子事件被探测到. 将时间间隔小于 15 s 的相邻子事件 (Swarm 卫星 15 s 约飞行经纬度 1° 的距离) 进行组合, 直到无法

*HTTP://earth.esa.int/web/guest/swarm/data-access

组合相邻子事件为止, 即一个子事件与相邻子事件之间的时间间隔大于 15 s 时, 则一个待确定的 SSTID 事件被探测到. 判断该事件发生期间 σ_z 分量是否存在幅值超过 ±0.15 nT 的扰动, 如果存在, 则该待确定的 SSTID 事件被判定为 EPB 事件而被排除, 不存在则该事件最终被确定为 SSTID 事件. 该 SSTID 事件的参数如地理经纬度、磁经纬度、地方时、季节等由该事件的中心点确定或标记.

2　SSTID 探测算法探测的事件

2.1　SSTID 事件

图 1 为 Swarm-Charlie 卫星 2016 年 5 月 27 日观测到的典型 SSTID 事件. 探测到的 SSTID 事件使用绿框标记. 上下边界之间的宽度表示事件对应的最大峰峰值, 左右边界表示事件的起始和结束时间, 垂直于时间轴的品红色直线表示事件中心. 结合图 1(a) 和图 1(b) 可知, 在 2 个 SSTID 事件期间, σ_y 方向上发生了磁场波动, 而 σ_z 方向上没有发生磁场波动, 不伴随等离子体密度值的扰动, 且这两个 SSTID 事件关于磁赤道共轭对称. 图 1(c) 给出了 B_y, B_x, B'_z, 这里 B'_z 是被截止周期为 1 min 的高通滤波器处理后得到的. 虽然 B_x 方向上磁场分量发生了波动, 但是在事件发生期间, B_y 方向上磁场波动更加显著, 其波动的最大幅值达到了 8 nT, 且磁场幅值主要向东偏扰. 图 1(d) 给出的是对图 1(c) 中 E2 事件的 B_x, B_y 磁场分量的放大以及对应的场向电流密度曲线. 在此次 SSTID 事件发生期间, B_x 表现出与 B_y 反向变化的特征.

2.2　EPB 污染性事件

探测 SSTID 事件时也探测到了其他污染性事件. 为确保 SSTID 事件研究的可靠性, 应将这些污

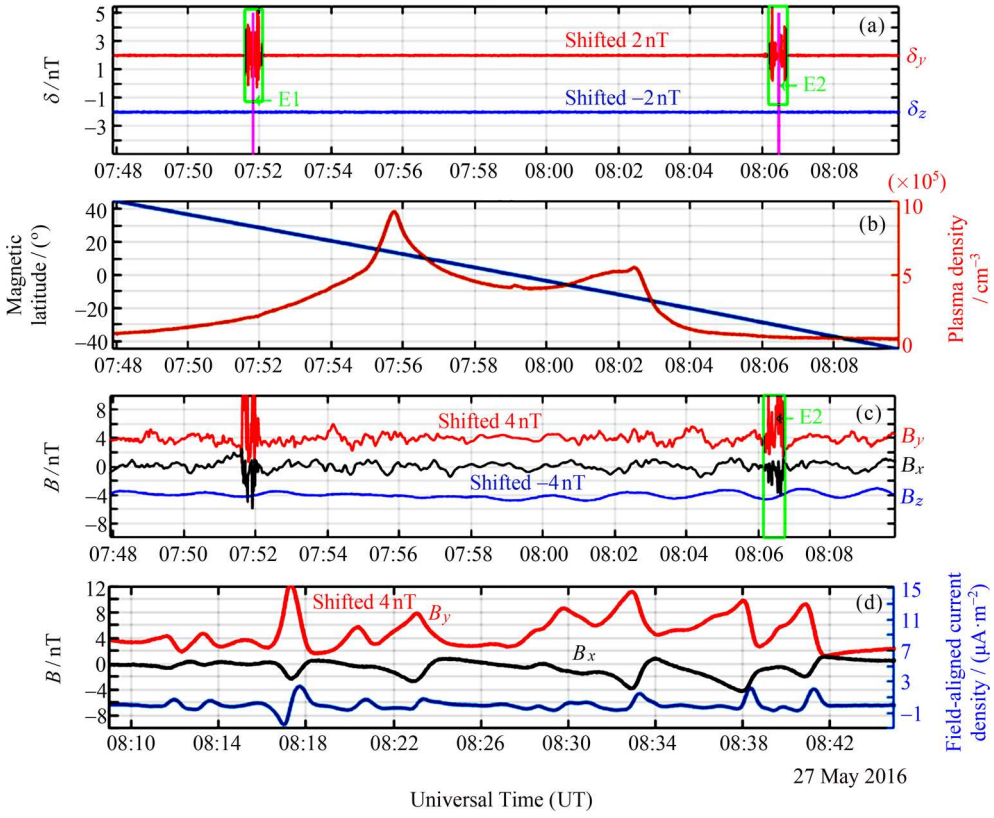


图 1　2016 年 5 月 27 日 Swarm-Charlie 卫星探测到的 SSTID 事件. (a) 使用 1 Hz 高通滤波器处理 B_y, B_z 磁场分量得到的高频磁场分量信号 σ_y, σ_z. (b) 等离子体密度和磁纬曲线. (c) B_y, B_x, B'_z, (被截止周期为 1 min 的高通滤波器处理后得到的) 三个磁场分量. (d) 对 (c) 中 E2 事件的磁场分量信号进行放大

Fig. 1　SSTID events detected by the Swarm-Charlie satellite on 27 May 2016. (a) 1 Hz high-pass filtered B_y (σ_y) and B_z (σ_z) components of magnetic field in MFA coordinates. (b) Plasma density and magnetic latitude. (c) B_y, B_x, B'_z (1 min high-pass filtered B_z) components. (d) Zoomed details of the event marked by the green box in (c)

染性事件排除. 图 2(a) 给出了 2015 年 3 月 25 日 Swarm-Bravo 探测到的 EPB 污染性事件, 图中用黑框标记. 其中, EB2, EB3, EB6 是探测到的三个典型 EPB 事件. 在 EPB 事件发生期间, σ_y 方向及 σ_z 方向上的磁场分量也发生了较强扰动. 由图 2(b) 可知, 等离子体密度也发生了较为强烈的耗空, 赤道双峰结构严重破坏, 电子密度下降到 $1\times10^5\,\mathrm{cm}^{-3}$ 量级, 表明出现了泡状结构. 图 2(a) 中较典型的 EPB 污染事件已经完全被排除, 但是对于在磁赤道附近的两个短事件 E4 和 E5, 由于 σ_z 方向上的磁场扰动并不强烈 (低于 0.15 nT 阈值), 并且如图 2(c) 和 2(d) 所示, 等离子体密度采样率为 2 Hz, 远低于磁场数据 50 Hz 的采样频率, 即也无法通过事件发生期间等离子体密度的扰动来判定 EPB 事件, 最终导致其不被判定为 EPB 事件. 但是从 E4, E5 前后关联的事件背景分析, 可以人工将 E4, E5 判断为 EPB 事件. 因此, 在探测 SSTID 事件时, 仅根据主磁场平行方向上的磁扰超过阈值来排除 EPB 事件的算法存在缺陷, 无法将磁赤道附近少量短、弱的非典型 EPB 事件完全排除.

2.3 SSTID 探测算法探测的短事件集分布

由于 SSTID 事件探测算法的缺陷, 探测到的 2015 年 1 月 1 日至 2019 年 12 月 31 日期间的 SSTID 事件集合中存在没有完全被排除的非典型 EPB 污染事件. 为了分析未被排除的短、弱 EPB 污染事件和 SSTID 事件在不同二维参数空间中的分布特征, 针对基于阈值判断的 SSTID 探测算法探测的事件集合, 选择持续时间小于 0.5 s 的短事件构造一个短事件集, 分析短事件集在地理经度–磁纬、磁地方–磁纬的二维参数空间中的分布特征. 图 3(a) 为短事件集在地理经度–磁纬二维参数空间中的散点分布, 其中每一个点代表一个事件. SSTID 探测算法无法排除短、弱 EPB 污染事件. 由图 3(a) 可以看出, 磁赤道附近明显存在一个比较密集的 EPB 污染事件区域. SSTID 事件主要分布在南北半球的磁纬 15°N−45°S 之间, 并且 EPB 污染事件与 SSTID 事件之间存在明显的密度分布差异. 图 3(b) 给出了短事件集在磁地方时–磁纬二维参数空间中的散点分布. 磁赤道附近的短、弱 EPB 污染事件的密集区域大致分布在 19:00 MLT 至午夜后的凌晨 01:00 MLT.

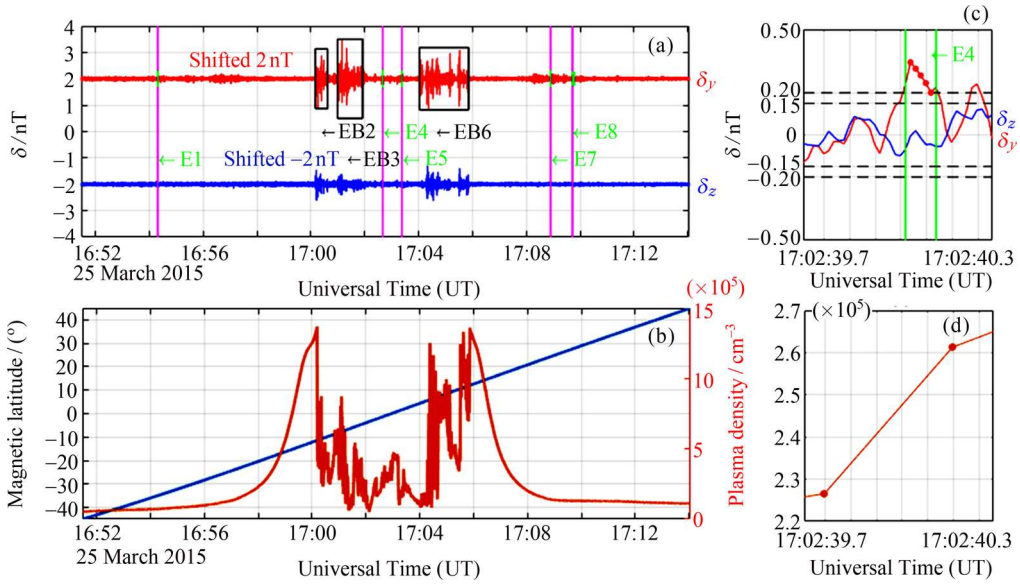


图 2 2015 年 3 月 25 日 Swarm-Bravo 卫星探测到的赤道等离子体泡事件. (a) (b) 与图 1(a) (b) 相同, 黑框标记的是利用平行于主磁场方向上的磁扰动探测的 EPB 事件, 绿框标记的是 SSTID 事件. E4 和 E5 是 EPB 判定算法未排除的 EPB 污染事件. (c) 和 (d) 分别为 (a) 中 E4 事件的放大和对应的等离子体密度曲线

Fig. 2 Equatorial Plasma Bubble (EPB) detected by the Swarm-Bravo satellite on 25 March 2015. (a) (b) are the same to Fig. 1 (a) (b). The black boxes mark the EPB events detected by disturbance in σ_z, and the green boxes mark the SSTID events. E4 and E5 are EPB events but not eliminated by SSTID algorithm. (c) and (d) are the amplification of E4 events in (a) and the corresponding plasma density curves

SSTID 事件呈现南北半球对称分布, 并且各半球分布特征也非常相似, 在 16:00 MLT−24:00 MLT 分布较多, 中心点在 20:00 MLT 附近. 白天主要分布在 08:00 MLT−12:00 MLT 之间, 中心点在 10:00 MLT 附近. 同时, 在此二维参数空间中两种事件也存在密度分布差异.

3　短事件集的聚类分析

探测 SSTID 事件时, 使用传统时域信号的阈值判断方法无法排除短、弱的非典型 EPB 污染事件. 这种污染事件保留了 EPB 事件的统计属性, 在参数空间与真实的 SSTID 存在明显的密度分布差异. 可利用两种电离层扰动事件在不同二维参数空间中的密度及区域分布差异, 使用基于密度的 DBSCAN 聚类算法甄选 SSTID 事件中的 EPB 污染事件, 并对提取出的 EPB 污染事件进行深入验证与分析.

3.1　聚类算法原理

DBSCAN 聚类算法是一种基于密度的分类算法. 其通过限定邻域 (ε_0) 和密度阈值 ($p_{\min}$) 这两个参数, 以某一数据点为中心, 反映该数据点周围 ε_0 范围内的数据稠密程度. DBSCAN 算法的基本原理是遍历数据集合中每个数据点, 并以当前遍历数据点为中心, 统计其周围 ε_0 范围内的数据点个数是否不小于 $p_{\min}$, 如果存在点 p 满足 ε_0 范围内的数据点个数不小于 $p_{\min}$, 则将点 p 标记为核心对象点, 而在点 p 周围 ε_0 范围内的数据点则由点 p 进行密度直达. 当多个点通过密度直达的特点传递和扩展时, 扩展的终止数据点是由扩展的起始数据点密度可达的, DBSCAN 算法将通过密度直达或密度可达关联起来的数据点标记为密度相连的点. 当无法再通过密度直达扩展时, 算法会将密度相连的点划分为一族即一类. 如果遍历到的数据点在其周围 ε_0 范围内的密度没有达到阈值且不由任何核心对象点密度直达, 则不划分为任何一类而被判定为噪声点.

图 4 给出了 DBSCAN 聚类算法原理. 由图 4 可以看出, 聚类算法中限定邻域 $\varepsilon_0 = \{(x,y)|x^2 + y^2 \leqslant 8$, 密度阈值 $p_{\min} = 4$, 箭头方向表示密度直达. 以 p_1∼p_6 分别为中心的 ε_0 范围内的数据点个数均不小于 $p_{\min}$, 将 p_1∼p_6 都判定为核心对象点. 通过由

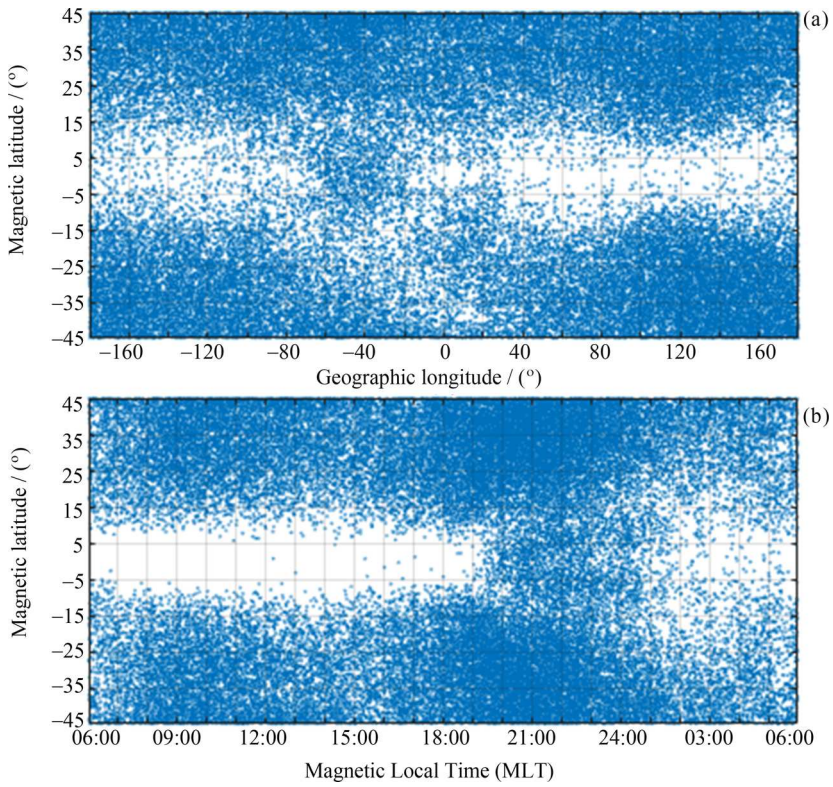


图 3　Swarm 卫星 5 年探测到的短事件分别在地理经度与磁纬 (a)、磁地方时与磁纬二维坐标系 (b) 中的分布

Fig. 3　Distribution of the short events detected by Swarm satellite in 5 years in 2D coordinates of geographic longitude versus magnetic latitude (a) and magnetic local time versus magnetic latitude (b)

核心对象点密度直达的特点扩展, 将 $p_1-p_2-p_3-p_4-B$, $p_4-p_3-p_2-p_1-C$ 判定为密度可达, 因此 p_1~p_6, B, C 这些数据点是密度相连的, 算法最终将其划分为一类. B, C 是由核心对象点密度直达的边界点, 而在以 A 点为中心的 ε_0 范围内的数据点个数小于 $p_{\min}$, 且 A 点不由任何核心对象点密度直达, 将 A 点判定为噪声点.

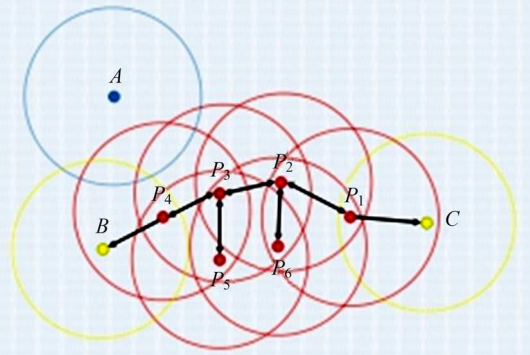


图 4 DBSCAN 聚类算法原理

Fig. 4 DBSCAN clustering algorithm principle

探测到的事件在参数坐标系空间中相对于不同坐标轴的分布差别较大, 因此在 DBSCAN 算法中应选择合适的椭圆邻域参数而不是圆形邻域参数, 以可靠区分参数空间中有密度分布差异的磁扰动事件.

3.2 短事件集聚类分析

利用 DBSCAN 聚类算法对图 3(a) 进行分析, 结果如图 5(a) 所示. 图 5(a) 中聚类参数

$$\varepsilon_0=\left\{(x,y)\middle|\frac{x^2}{6.5^2}+\frac{y^2}{2^2}\leqslant 1\right\},$$

椭圆邻域长轴为地理经度 6.5°, 短轴为磁纬 2°, $p_{\min}=48$. 聚类算法可以将待排除的 EPB 污染事件成功导出, 导出的黄色类 EPB 污染事件主要分布在地理经度 20°−60°W、磁纬 10°N−10°S 之间, 且大致关于磁赤道南北半球对称. 利用 DBSCAN 聚类算法也得到了 SSTID 事件高发生率区和低发生率区的清晰边界. 利用 DBSCAN 聚类算法对图 3(b) 进行分析, 结果如图 5(b) 所示. 图 5(b) 中聚类参数

$$\varepsilon_0=\left\{(x,y)\middle|\frac{x^2}{1.6^2}+\frac{y^2}{2^2}\leqslant 1\right\},$$

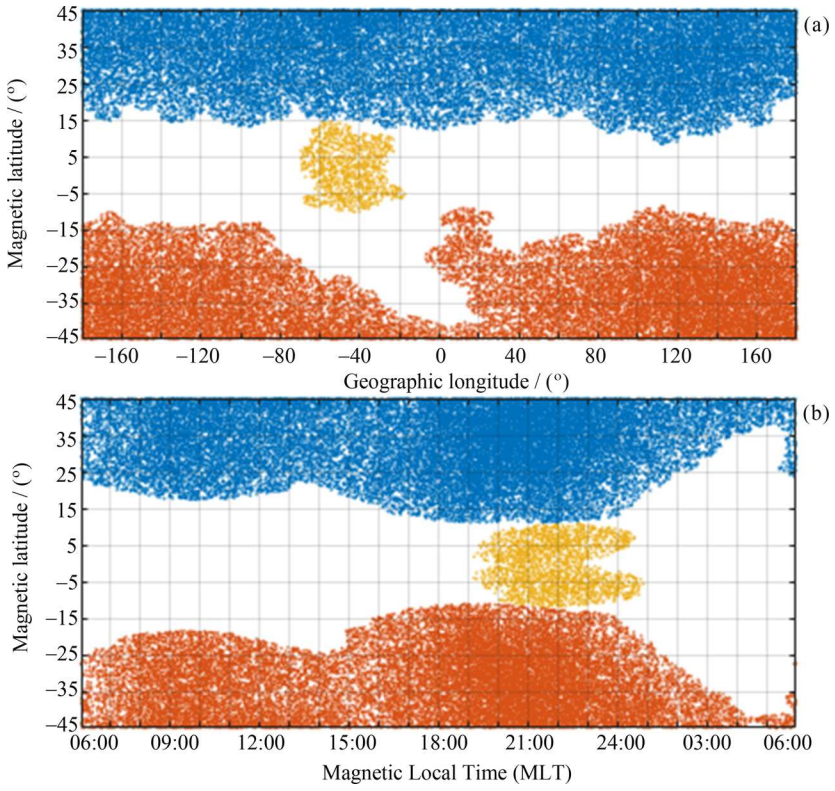


图 5 Swarm 卫星 5 年探测到的短事件分别在地理经度与磁纬 (a) 及磁地方时与磁纬 (b) 的二维坐标系中聚类后的分布

Fig. 5 Distribution of the short events detected by Swarm satellite in 5 years after cluster analysis in 2D coordinates of geographic longitude versus magnetic latitude (a) and magnetic time versus magnetic latitude (b)

椭圆邻域长轴为磁纬 2°, 短轴为磁地方时 1.6 h, $p_{\min}$ = 236. 由图 5(b) 可以看出, DBSCAN 聚类算法有效地将图 3(b) 中磁赤道附近夜间 19:00 MLT 时至凌晨 01:00 MLT 之间密集的短 EPB 污染事件从 SSTID 事件中提取出来 (图 5b 中黄色一类), 且这些 EPB 事件明显呈半球对称分布.

3.3　甄选出的 EPB 污染事件分析与验证

聚类算法能够将图 3(a)(b) 中磁赤道地区的短 EPB 污染事件有效甄选出来. 为进一步证明聚类结果的可靠性和合理性, 将图 5(a) (b) 中甄选出的黄色短 EPB 污染事件合并, 构造一个 EPB 事件聚类结果集, 分析其季节分布特征及对太阳活动的依赖性. 图 6 为 EPB 聚类结果集在地理经度与月份的二维坐标系中的散点分布. 由图 6 可以看出, EPB 事件主要分布在地理经度 20°−60°W 且月份在 10 至 3 月之间, 而在 4 至 9 月 (夏至季节) 相对稀疏很多. 图 7 给出了短 EPB 污染事件聚类结果集合中每年的污染事件数量和太阳活动强度年平均值的逐年变化曲线. 由图 7 可以看出, 2015 至 2018 年未排除的短 EPB 污染事件数逐年递减, 太阳活动强度逐年减弱, 二者变化趋势呈显著正相关, 这说明通过聚类方法得到的 EPB 事件依赖于太阳活动强度. 2019 年太阳活动强度相对于 2018 年几乎没有波动, 但是 2019 年的 EPB 事件个数略多于 2018 年. 这是由于 EPB 事件的发生率同时受卫星轨道地方时和季节的联合调控.

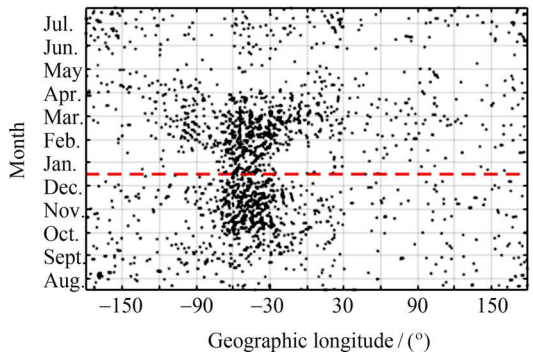


图 6　聚类结果中 EPB 事件并集在地理经度与月份的二维坐标系中的分布

Fig. 6　Distribution of the clustered EPB events in two-dimensional coordinate of geographical longitude versus month

3.4　分类质量

为衡量聚类算法的质量, 根据 Wang 等 [21] 研究评价分类质量的方法, 需要引入各种评价参数, 如正确接受数、错误拒绝数、错误接受数、正确拒绝数等. 要得到这些参数, 必须确定各种事件的属性. 实际上, 如果缺乏人工干预, 很难确定各种事件的属性. 这是因为意外事件或者噪声事件总是存在的, 例如一个 EPB 事件在较高纬度或者在非 SAA 地区发生, 本文聚类算法会拒绝该类 EPB 事件. 由于本文的事件数量太多, 人工判读显然缺乏可行性, 但是仍然可以通过已有的一些统计结果来部分验证该聚类算法的分类质量.

根据 Xiong 等 [15]、Stolle 等 [16]、Luhr 等 [17] 关于 EPB 事件对经度和月份的依赖性特征的研究结果可知, EPB 事件在月份和地理经度分布上都存在两个发生率几乎为 0 的区域. EPB 事件在 12 月初至 2 月底期间, 地理经度 150°−240°E 的发生率几乎为 0, 并且在 5 月初至 8 月中旬期间, 地理经度 20°−80°W 的发生率也基本为 0. 当 DBSCAN 聚类算法甄选的 EPB 事件出现在这两个发生率为 0 的区域时, 将其标记为非 EPB 类污染事件的错误接受事件, 而在这个区域之外的事件则标记为 EPB 类污染事件的正确接受事件.

对正确接受的 EPB 污染事件和错误接受的非 EPB 类污染事件分别使用黑点和红叉进行标记, 如图 8 所示. 最终统计出被正确接受的 EPB 事件数为 2330 个, 错误接受的非 EPB 事件为 101 个. 利用

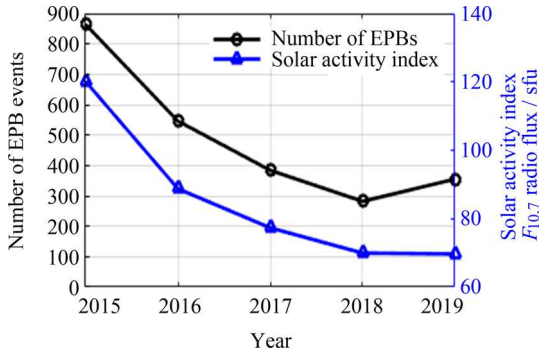


图 7　EPB 污染事件集事件个数及太阳活动年平均强度的逐年变化

Fig. 7　Annual changes of the number of the clustered EPB events and the average intensity of solar activity

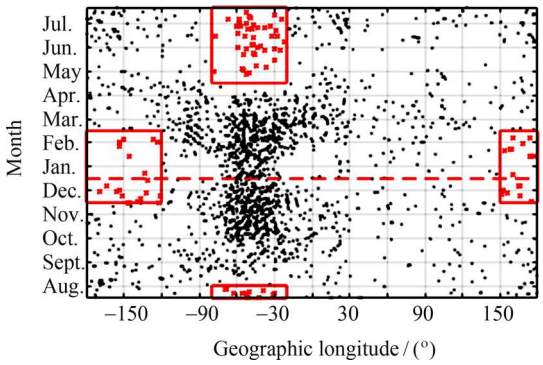


图 8 聚类算法甄选的 EPB 污染事件在地理经度与月份的二维参数空间中的分布 (正确接受的 EPB 污染事件和错误接受的非 EPB 类污染事件分别使用黑点和红叉进行标记)

Fig. 8 Distribution of the clustered EPB events in two-dimensional coordinate of geographical longitude versus month (EPB pollution events correctly accepted and non-EPB pollution events incorrectly accepted are marked with black dots and red crosses respectively)

Wang 等 [21] 评价分类质量方法中查准率和误检率的计算方法, 计算出聚类算法的查准率和误检率分别为 95.85% 和 4.15%, 这说明 DBSCAN 聚类算法甄选的 EPB 事件基本可靠. 需要指出的是, 这种聚类算法的分类质量是根据文献 [15−17] 研究 EPB 事件的经验计算的, 即本文统计的分类算法质量参数是理论值, 可能与真实值稍有偏差.

4 讨论

Yin 等 [1] 研究分析 SSTID 事件时, 通过判定平行于主磁场方向是否发生超过设定阈值的磁扰来排除 EPB 污染事件. 这种判定方法存在不足, 会导致 SSTID 事件集合中存在非典型 EPB 污染事件. 但这种非典型 EPB 污染事件与 SSTID 事件在参数空间存在明显密度分布差异. 本文利用这种差异, 使用聚类算法将 SSTID 事件中无法排除的磁赤道区域非典型 EPB 事件进一步甄别、选取. 全部短事件分别在地理经度–磁纬、磁地方时–磁纬二维空间的聚类结果表明, 由聚类算法甄选的非典型 EPB 事件在地理经度 20°−60°W 之间分布密集, 在磁地方时上主要分布在 20:00 MLT−24:00 MLT. 聚类结果中 EPB 事件的地理经度和磁地方时依赖性与文献 [15−18] 的统计结果基本一致. 对于通过聚类算法选取的 EPB 事件, 使用与文献 [15, 17] 相同的方法, 给出这些事件发生率在地理经度与月份二维参数空间中的分布特征. 结果表明, 图 6 中 EPB 事件在 10 月至 3 月冬季分布最为密集. 这与 Xiong 和 Stolle 研究结果非常相似. 图 7 中 EPB 聚类结果集事件数的逐年分布与太阳活动强度逐年变化趋势一致, 说明 EPB 依赖于太阳活动, 这也与以往研究结果相符 [16,17].

综上可知, 本文提出的基于密度的聚类算法能够弥补基于时序阈值判断的 SSTID 探测算法无法排除短、弱 EPB 事件的缺陷. 因此, 数据处理时可以使用聚类算法进一步将 SSTID 事件集合中无法排除的短、弱 EPB 事件剔除. 传统时序阈值分析方法不能甄选出 SSTID 事件中存在的全部非典型 EPB 事件, 而利用聚类算法不仅可以对非典型 EPB 事件进行提取, 而且能够明确给出 EPB 污染事件的具体分布区域, 同时使不同聚类事件在参数空间分布上产生明确边界.

但是, DBSCAN 聚类算法也存在一些不足. 首先, 该算法的密度阈值参数和限定邻域参数选取较麻烦, 如果参数选取不合适会影响聚类结果. 其次, 该算法不适合密度分布极其不均匀的数据集, 会出现本该划分为一类的数据点被分成多个不同类的现象. 最后, 该算法会将低于密度阈值的数据点判定为噪声点, 即对于分布极其稀疏的 EPB 污染事件数据仍旧无法提取.

5 结论

利用三颗 Swarm 卫星测量的高频磁场数据探测小尺度电离层行扰事件, 使用基于密度的 DBSCAN 聚类算法排除磁赤道附近的非典型 EPB 污染事件, 得到如下研究结果.

(1) 典型 EPB 事件已被基于主磁场扰动的阈值排除算法排除. 未被排除的 EPB 污染事件主要是持续时间短且扰动幅度弱的事件, 主要分布在磁赤道附近磁纬 15°N−15°S 之间.

(2) 使用基于密度的 DBSCAN 聚类算法, 能够将磁赤道附近磁纬 15°N−15°S 之间密集的 EPB 事件甄选出来并聚成一类. 由聚类算法提取出的 EPB 事件对地理经度、月份、磁地方时和太阳活动强度的依赖性特征与已有研究一致. 这说明利用聚类算法能

够甄别 EPB 污染事件并将其从 SSTID 数据集合中剔除, 保证 SSTID 数据集的纯净性.

(3) SSTID 事件聚类主要分布在南北半球磁纬 15°N−45°S, 而在磁赤道附近磁纬 10°N−10°S 和 SAA 区域则极其稀疏. 在地方时上, SSTID 事件聚类分别在 08:00 MLT−12:00 MLT 和 17:00 MLT−23:00 MLT 比较密集. 利用聚类算法进行数据处理, 可使 SSTID 事件在高发生率区和低发生率区之间形成清晰边界, 并与污染性的短 EPB 事件完全区分开. 研究结果有助于对 SSTID 事件的后继研究.

参考文献

[1] YIN F, LUHR H, PARK J, *et al.* Comprehensive analysis of the magnetic signatures of small-scale traveling ionospheric disturbances, as observed by Swarm [J]. *J. Geophys. Res*: *Space Phys.*, 2019, **124**(12): 10794-10815. DOI: 10.1029/2019JA027523

[2] MARTINIS C, BAUMGARDNER J, WROTEN J, *et al.* Seasonal dependence of MSTIDs obtained from 630.0 nm airglow imaging at Arecibo [J]. *Geophys. Res. Lett.*, 2010, **37**(11): 1-5. DOI: 10.1029/2010GL043569

[3] YOKOYAMA T, HYSELL D L, OTSUKA Y, *et al.* Three-Dimensional simulation of the coupled Perkins and Es-Layer instabilities in the nighttime midlatitude ionosphere [J]. *J. Geophys. Res.*: *Space Phys.*, 2009, **114**(A3): 1-16. DOI: 10.1029/2008JA013789

[4] BEHNKE R. F layer height bands in the nocturnal ionosphere over Arecibo [J]. *J. Geophys. Res.*: *Space Phys.*, 1979, **84**(A3): 974-978

[5] COSGROVE R B, TSUNODA R T. Instability of the E-F coupled nighttime midlatitude ionosphere [J]. *J. Geophys. Res.*: *Space Phys.*, 2004, **109**(A4): 1-7. DOI: 10.1029/2003-JA010243

[6] TSUNODA, ROLAND T. On the coupling of layer instabilities in the nighttime midlatitude ionosphere [J]. *J. Geophys. Res.*: *Space Phys.*, 2006, **111**(A11): 1-14. DOI: 10.1029/2006JA011630

[7] PARK J, LUHR H, STOLLE C, *et al.* Statistical survey of nighttime midlatitude magnetic fluctuations: their source location and Poynting flux as derived from the Swarm constellation [J]. *J. Geophys. Res.*: *Space Phys.*, 2016, **121**(11): 11235-11248. DOI: 10.1002/2016JA023408

[8] PARK J, LUHR H, STOLLE C, *et al.* Magnetic signatures of medium-scale traveling ionospheric disturbances as observed by CHAMP [J]. *J. Geophys. Res.*: *Space Phys.*, 2009, **114**(A3): 1-13. DOI: 10.1029/2008JA013792

[9] PARK J, LUHR H, KERVALISHVILI G, *et al.* Nighttime magnetic field fluctuations in the topside ionosphere at midlatitudes and their relation to medium-scale traveling ionospheric disturbances: the spatial structure and scale sizes [J]. *J. Geophys. Res.*: *Space Phys.*, 2015, **120**(8): 6818-6830. DOI: 10.1002/2015JA021315

[10] KELLY M. The Earth's Ionosphere: Plasma Physics And Electrodynamics [M]. 2nd. Boston: Academic Press, 2009

[11] WOODMAN R F. Spread F-an old equatorial aeronomy problem finally resolved [J]. *Ann. Geophys.*, 2009, **27**(5): 1915-1934

[12] KELLEY M C, MAKELA J J, LA BEAUJARDIERE O D, *et al.* Convective ionospheric storms: a review [J]. *Rev. Geophys.*, 2011, **49**(2): 1-26. DOI: 10.1029/2010RG000340

[13] LUHR H, ROTHER M, MAUS S, *et al.* The diamagnetic effect of the equatorial Appleton anomaly: its characteristics and impact on geomagnetic field modeling [J]. *Geophys. Res. Lett.*, 2003, **30**(17): 1-4. DOI: 10.1029/2003-GL017407

[14] PARK J, LUHR H, STOLLE C, *et al.* The characteristics of field-aligned currents associated with equatorial plasma bubbles as observed by the CHAMP satellite [J]. *Ann. Geophys.*, 2009, **27**: 2685-2697

[15] XIONG C, LUHR H, MA S Y, *et al.* Features of highly structured equatorial plasma irregularities deduced from CHAMP observations [J]. *Ann. Geophys.*, 2012, **30**(8): 1259-1269

[16] STOLLE C, LUHR H, ROTHER M, *et al.* Magnetic signatures of equatorial spread F as observed by the CHAMP satellite [J]. *J. Geophys. Res.*: *Space Phys.*, 2006, **111**(A2): 1-13. DOI: 10.1029/2005JA011184

[17] LUHR H, XIONG C, PARK J, *et al.* Systematic study of intermediate-scale structures of equatorial plasma irregularities in the ionosphere based on CHAMP observations [J]. *Front. Phys.*, 2014, **2**: 15

[18] YOKOYAMA T, STOLLE C. Low and midlatitude ionospheric plasma density irregularities and their effects on geomagnetic field [J]. *Space Sci. Rev.*, 2017, **206**(1): 495-519

[19] ESTER M, KRIEGEL H, SANDER J, *et al.* A density-based algorithm for discovering clusters in large spatial Databases with Noise [C]//Proceeding of 2nd International Conference on Knowledge Discovery and Data Mining. Portland: AIAA Press, 1996: 226-231

[20] PARK J, STOLLE C, LUHR H, *et al.* Magnetic signatures and conjugate features of low-latitude plasma blobs as observed by the CHAMP satellite [J]. *J. Geophys. Res.*: *Space Phys.*, 2008, **113**(A9): 1-10. DOI: 10.1029/2008JA-013211

[21] WANG Cheng, LIU Yafeng, WANG Xincheng, *et al.* Appraisal identification of classifier's performance [J]. *Electron. Design Eng.*, 2011, **19**(8): 13-15, 21(王成, 刘亚峰, 王新成, 等. 分类器的分类性能评价指标 [J]. 电子设计工程, 2011, **19**(8): 13-15, 21)