Using k-Means Cluster Analysis to find and classify Ion Density Irregularities consistent with Equatorial Plasma Bubbles from Multi-year Formosat-5 Advanced Ionospheric Probe observations

Cornelius Csar Jude Hisole Salinas[1], Loren C. Chang[1], Chi-Kuang Chao[1], Jann-Yenq Liu[1], and Charles C. H. Lin[2]

[1]National Central University
[2]National Cheng Kung University

November 22, 2022

Abstract

This work explores the results of applying Square Euclidean and Correlation k-means cluster analysis on ion density irregularity profiles observed by the Advanced Ionospheric Probe (AIP) onboard the Formosat-5 (F-5) satellite from November 2017 to November 2020. The Square Euclidean cluster analysis yielded separate clusters each for ion density irregularities consistent with Equatorial Plasma Bubbles (EPBs) occurring over the southern low-latitude, northern low-latitude and equator. The Correlation k-means cluster analysis only yielded one cluster characterized by ion density irregularities consistent with EPBs occurring over the equator. Thus, this work suggests that a cluster analysis preferably a Square Euclidean Cluster analysis can be used to find and classify ion density irregularities consistent with EPBs. This work also shows that the F-5/AIP can perform multi-year observations of ion density irregularities at ˜710 km altitude and at ˜2230 pm local-time that are consistent with irregularities due to EPBs.

Using k-Means Cluster Analysis to find and classify Ion Density Irregularities consistent with Equatorial Plasma Bubbles from Multi-year Formosat-5 Advanced Ionospheric Probe observations

Cornelius Csar Jude H. Salinas[1, 2], Loren C. Chang[1,2], Chi-Kuang Chao[1,2], Jann-Yenq Liu[1,2] and Charles C.H. Lin[3]

[1]*Department of Space Science and Engineering, National Central University, Zhongli, Taiwan*

[2]*Center for Astronautical Physics and Engineering, National Central University, Zhongli, Taiwan*

[3]*Department of Earth Science, National Cheng Kung University, Tainan, Taiwan*

Corresponding author: Loren C. Chang, Department of Space Science and Engineering, National Central University, Zhongli, Taiwan (loren@g.ncu.edu.tw)

Abstract

This work explores the results of applying Square Euclidean and Correlation k-means cluster analysis on ion density irregularity profiles observed by the Advanced Ionospheric Probe (AIP) onboard the Formosat-5 (F-5) satellite from November 2017 to November 2020. The Square Euclidean cluster analysis yielded separate clusters each for ion density irregularities consistent with Equatorial Plasma Bubbles (EPBs) occurring over the southern low-latitude, northern low-latitude and equator. The Correlation k-means cluster analysis only yielded one cluster characterized by ion density irregularities consistent with EPBs occurring over the equator. Thus, this work suggests that a cluster analysis preferably a Square Euclidean Cluster analysis can be used to find and classify ion density irregularities consistent with EPBs. This work also shows that

the F-5/AIP can perform multi-year observations of ion density irregularities at ~710 km altitude and at ~2230 pm local-time that are consistent with irregularities due to EPBs.

Plain Language Summary

Equatorial plasma bubbles (EPBs) predominantly cause the problems of space-based communication and navigation systems over the low-latitudes. These communication and navigation systems involve the use of signals that propagate through our ionosphere. EPBs cause problems in these systems by disturbing the ions along the path of these signals. A well-known way of characterizing the occurrence of these EPBs is to look at the fluctuations or irregularities of ionospheric ion density profiles and determine whether such irregularities are due to EPBs. However, satellite observations have already amassed millions of ion density profiles. Classifying these profiles is a major challenge. This work aims to help tackle this challenge by presenting a data-driven approach to classifying these profiles. This work shows the use of k-means cluster analysis on ion density profiles measured by the Advanced Ionospheric Probe onboard the Formosat-5 satellite. This work will show that the approach successfully finds and classifies ion density irregularities that have characteristics consistent with EPBs.

Index Terms/ Keywords: Ion Density, Equatorial Plasma Bubbles, Ionosphere, Data Science

Key Points:

- K-means cluster analysis is applied to multi-year ion density irregularity profiles.
- K-means cluster analysis finds clusters of ion density irregularities consistent with equatorial plasma bubbles.
- Formosat-5's Advanced Ionosphere Probe observes ion density irregularities consistent with equatorial plasma bubbles.

1. Introduction

Ionospheric irregularities cause significant disruptions in satellite communication and navigation systems (Basu et al, 2001). In the equatorial region, equatorial plasma bubbles (EPBs) cause most of these ionospheric irregularities. They form as a result of the nonlinear evolution of the generalized Rayleigh-Taylor instability in the night-time ionosphere (Dungey, 1956; Kelley, 1989). EPBs grow to be several hundred kilometers in the east-west direction, thousands of kilometers in the north-south direction and also hundreds of kilometers in the vertical direction (Kil, 2015; Yokoyama, 2017). Their overall shape also varies significantly. Thus, observations and identification of EPBs are a challenge.

While observing the complete three-dimensional image of an EPB is currently still impossible, we further our knowledge on EPB structure by analyzing when, where and under what conditions they frequently occur. In the past three decades, significant progress has been made in characterizing the occurrence rates of EPBs by analyzing the irregularities observed using satellite in-situ ion density measurements. When looking at ion or electron density profiles, EPBs manifest as sudden drops in density values. These analyses involve grouping the ion density profiles often in terms of EPB parameters. One of the most commonly used parameters is the depth of depletion $\Delta N/N_0$ where N_0 is a background ion density and ΔN is the perturbation. In the early 2000s, the first global climatology of EPB occurrence rates was formed by applying this method on more than a decade of measurements from 6 Defense Meteorological Satellite Program (DMSP) satellites (Huang et al, 2001; Burke et al, 2004; Gentile et al, 2006). DMSP satellites are in a circular sun-synchronous polar orbit at an altitude of ~840 km and they all cross the magnetic equator post-sunset. With this orbit, the Special Sensor-Ions, Electrons and Scintillation (SSIES) instruments onboard consistently measure night-time in-situ ion density latitude profiles with specific local-time coverages ranging between 1900 and 2100 pm local-time. These provided EPB occurrence rates as a function of longitude and month. It showed that EPBs inducing large plasma depletions occur most frequently during equinox and that in this season, most of the EPBs occur over the American, Atlantic and African sectors. They also showed that EPBs generally follow the magnetic declination angle. These are all consistent with theoretical studies (Tsunoda, 1985).

The same approach was utilized on measurements by the Republic of China Satellite – 1 (ROCSAT-1) as well as the Communication Navigation Outage Forecast System (C/NOFS) satellite. ROCSAT-1 is in a circular orbit at an altitude of 600 km and an inclination of 35 degrees. With this orbit, ROCSAT-1's ascending node regressed westward at a rate of ~7 degrees per day. With this nodal regression, the local-time as well as the magnetic latitude of measurements performed by the Ionospheric Plasma and Electrodynamics Instrument (IPEI) changed significantly each day (Yeh et al, 1999; Chen et al, 2001; Su et al, 2001; Le et al, 2003). Evening local-times were measured around 3 to 4 times a day. While this reduced the number of measurements at night compared to DMSP, this enabled the examination of the magnetic latitude dependence of EPBs. Burke et al (2004) analyzed ROCSAT-1 IPEI measurements in March and April of 2000 and 2002. They showed that EPBs occur more frequently around the magnetic equator.

C/NOFS used an Ion Velocity Meter (IVM) to measure ion density. It was in an elliptical orbit with a 13-degree inclination. At the time of launch in 2008, its perigee was around 400 km and its apogee was around 850 km. The precession of C/NOFS allowed it to attain full local-time coverage in 2 months. Heelis et al (2010) also grouped the profiles in terms of depth of plasma depletion to determine the seasonal and local-time-dependencies of EPB occurrence rates. They found that EPBs occur most frequently post-midnight and they suggested that this may be related to the seeding of EPBs in the bottom-side ionosphere.

More recent studies tried to actually isolate the individual EPBs along an ion density profile. Smith and Heelis (2017) employed edge detection on C/NOFS ion density profiles to isolate individual plasma bubbles along an ion density profile. This determined the occurrence rates of different scales of plasma bubbles. They found that most bubbles have a depletion length of around 200 km. Wan et al (2018) processed SWARM data to isolate the depletions and then the amplitude of these depletions. This determined the occurrence rate of different EPB intensities.

This work builds off of these previous studies in two ways. One, we present multi-year observations of ion density irregularities consistent with EPBs from a new space-based observational platform, the Advanced Ionospheric Probe (AIP) onboard the Formosat-5 (F-5). Two, we present a new method of finding and classifying ion density irregularity profiles consistent with EPBs. As shown by the aforementioned studies, most methods of EPB detection frequently involve grouping ion density profiles in terms of hyper-parameters such as ion density dip magnitude, depletion edge criteria or length of depletion. The chosen values of these parameters are based on looking at just a few ion density profiles. The need to group the profiles actually calls for classification algorithms. Classification algorithms offer a more data-driven approach compared to choosing hyper parameters as well as choosing the values for these parameters by only looking at a few ion density profiles. This work does exactly this by utilizing k-means cluster analysis to group ion density irregularity profiles. The clusters formed by the algorithm will then be characterized in terms of when (month) and where (in longitude and latitude) the irregularities occur. These are the first reported multi-year F-5/AIP observations on ion density irregularities consistent with EPBs. To the best of the authors knowledge, this is also the first use of cluster analysis for the purpose of ion density irregularity classification.

2. Methodology

This work utilizes F-5/AIP night-time ion density profiles from November 2017 to November 2020. F-5 was launched into a repeating sun-synchronous orbit (orbital inclination of 98.28 degrees) at an altitude of 720 km on August 25, 2017 (Lin et al, 2017; Chao et al, 2020). It orbits along the 1030-2230 LT sectors. This orbit allows AIP to globally sample the 2230 local-time in just 2 days. Figure 1A and 1B show this sampling. Figure 1A shows the sampling track of AIP for one day while figure 1B shows the sampling track for two days. The geographic latitudinal coverage of AIP, on average, is between -35S and 65N. Since this work is focused on EPBs, we only utilize data between 30S and 30N. AIP utilizes an ion trap sensor to measure ion concentration (Lin et al, 2017). Its sampling rate is up to 8192 Hz which can resolve ion density structures as small as 7.4 m. However, this work utilizes data that is the median of ion density measurements in 1 second. This yields orbital profiles with a spatial resolution of around 0.05 degrees or ~5 km. There are approximately 7 orbital profiles per day. To further minimize data-gaps, these profiles are interpolated into 0.5-degree latitudinal resolution.

To isolate the irregularities in a given ion density profile and to determine their amplitudes, we first filter the ion density profiles using a 5th-order Butterworth band-pass filter that cuts off fluctuations with horizontal wavelengths less than 200 km and greater than 500 km between 40S and 40N. These are well within the range of EPB dimensions (Kil, 2015; Yokoyama, 2017). Then, to get the amplitudes, we take the square-root of the square of this filtered form. Figure 1C shows a sample profile, its filtered form and the amplitude profile. The strongest irregularities in this sample profile are clearly between 10N and 20N. The filtered form of the profile effectively removes the background ion density profile but retains the strongest irregularities. Note though that this approach isn't meant to mimic the exact dimensions of the unfiltered irregularities. This approach is just meant to provide an amplitude profile whose regions of highest amplitudes coincide with the strongest irregularities. These amplitude profiles are then subject to the cluster analysis.

K-means cluster analysis is an unsupervised clustering method that groups data in terms of their similarities (Wu, 2012). The general algorithm of a k-means cluster analysis is as follows:

Step 1: Given a set of data points, randomly choose N number of initial cluster centers.

Step 2: Calculate the difference between each data point and the initial cluster centers and determine which cluster each data point is closer to.

Step 3: After all data points are classified, calculate the average of each cluster and set this average as the new cluster center.

Step 4: Repeat steps 2 and 3 until the cluster centers no longer change.

This work uses MATLAB's built-in k-means cluster analysis function to perform an 8-cluster k-means cluster analysis utilizing the square Euclidean distance and the correlation distance similarity metric. This means that the analysis will yield 8 numbers clusters and the square Euclidean distance and the correlation distance are our measures for the differences of the data points. For a dataset made of profiles (e.g. ion density profiles), a square Euclidean distance metric will group profiles whose element-by-element differences are minimal. If x and y are profiles, the square Euclidean metric is defined as:

$$\left||x-y|\right| = \sum_{i=1}^{d} x_i - y_i \tag{1}$$

where d is the total number of elements in a profile, x_i is element number i in profile x and y_i is element number i in profile y. On the other hand, a correlation distance similarity metric will group profiles with the highest element-by-element correlation. The standard statistical definition of correlation is used. With regards to the number of clusters, previous studies grouped the profiles into 4 groups depending on their depth of plasma depletion. To find the optimal number of groups for a cluster analysis, we look at the average of the sum of data point-to-cluster center distance (PtoD distance hereafter) for all groups (Wu, 2012). The lower the PtoD is in a group, the higher the similarity of the data-points in that group. The best number of clusters is the number such that decreasing this number significantly reduces the average PtoD while increasing this number minimally changes the average PtoD. When plotting the average PtoD as a function of number of clusters, the value for the best number of clusters is the 'knee' of this plot. We found that 8 clusters satisfied these conditions. Hence, we chose this as the number of clusters. Finally, for each group, we tabulate how many profiles are found on each month for all years from 2017 till 2020 as well as for a given longitude bin. This work utilizes a 30-degree longitude bin. This allows us to see the occurrence frequencies of each group as a function of longitudinal sector and month. Previous studies calculated occurrence rates as the ratio of the number of profiles for a given group and the total number of possible profiles. However, we found that for a grid of month and longitude, the total number of possible profiles is uneven. This is shown in figure 1D. Hence, we don't use this definition because the uneven distribution will cause biases.

3. Results

Figure 2 shows the results of the Square Euclidean Cluster Analysis. Figure 2 shows 8 panels with each panel containing two plots. Each panel corresponds to a cluster. The left plot of each panel shows the average of all profiles under the given cluster as a function of geographic latitude. The errorbars are the standard deviation of the profiles. Note that the y-axis for the left plots differ per panel. The right plot of each panel shows the number of profiles found in each month and longitude bin. This will be called the occurrence frequency plot. For example, the occurrence frequency plot of cluster 1 in figure 2 shows that of all profiles for all years, there are only 5 profiles of cluster 1 found in November and over geographic longitude -60°. Note that we don't separate the profiles in terms of year. Hence, this right plot actually shows the seasonality of the occurrence frequency for each cluster.

Apart from the geographic location, amplitude and seasonality of these irregularities, it is also important to characterize the magnetic inclination angles over these profiles because this also helps determine whether the irregularity within these profiles may be due to EPBs (Tsunoda, 1985). One way to do this is show a plot of the magnetic inclination angles as a function of latitude and longitude. Then, for each cluster, we first determine the latitudinal location of the peak amplitude and then also determine the most frequent longitudinal location of the cluster's profiles. We then check what the magnetic inclination angle is over this latitudinal and longitudinal location. This work takes a different approach. Instead of showing a separate plot for the magnetic inclination angles, we include the information on left plot of each panel. To do this, first, for each ion density profile, we use MATLAB's built-in IGRF model to determine the magnetic inclination angles at each data point in the profile. Thus, for each ion density profile, there is a corresponding magnetic inclination angle profile. Then, from this magnetic inclination angle profile, we note the average geographic latitude of magnetic inclination angles -15°, 0° and 15°. We only look at these magnetic inclination angles because we are only interested in knowing the magnetic low-latitude's geographic location. Finally, for each cluster, we average the geographic latitudes of magnetic inclination angles -15°, 0° and 15°. The red line in each panel's left plot is situated over the average geographic latitude of magnetic inclination angle 0° for the given cluster. The green line to the left (right) of the red line in each left plot is situated over the average geographic latitude of magnetic inclination angle -15° (15°) for the given cluster. Plotting these lines over the cluster's average profile quickly shows whether the ion irregularity profiles in the cluster have peaks found within the magnetic low latitude. If they are, this would be additional evidence that most of the ion density irregularities in the cluster may be attributed to EPBs.

For all clusters found using the Square Euclidean cluster analysis, it will be shown that the errorbars are small indicating that the number of profiles deviating from each cluster's average latitudinal variation are minimal. This suggests that most of the profiles in the clusters have magnitudes and latitudinal variations consistent with the average profiles. Figure 2's panel A shows the Square Euclidean cluster-1 irregularities (Sq-Cluster-1 irregularities hereafter and other Square Euclidean clusters will follow this labeling). The average profile plot shows that cluster-1 irregularities have peak amplitudes of 80 units (units = 10^3 #/cc) between geographic latitudes 10S and 10N. The occurrence frequency plot of Sq-Cluster-1 shows that Sq-Cluster-1 irregularities are mostly found in February and November over the American sector. The green and red lines indicate that the geographic latitude of the peak amplitudes in Sq-Cluster-1 irregularities are within the geomagnetic low-latitudes.

Figure 2's panel B shows Sq-Cluster-2 irregularities. Its average profile plot shows that Sq-Cluster-2 irregularities have peak amplitudes of 40 units between latitudes 5S and 5N. The occurrence frequency plot of Sq-Cluster-2 shows that Sq-Cluster-2 irregularities are mostly found in February and in November also over the American sector. The green and red lines indicate that the geographic latitude of the peak amplitudes are within the geomagnetic low-latitudes.

Figure 2's panel C shows Sq-Cluster-3 irregularities. Its average profile plot shows Sq-Cluster-3 irregularities have peak amplitudes of 50 units between latitudes 20S and 5S. The occurrence frequency plot of Sq-Cluster-3 shows that Sq-Cluster-3 irregularities are mostly found in February and November over the complete American sector. The green and red lines indicate that the geographic latitude of the peak amplitudes are partially within the geomagnetic low-latitudes.

Figure 2's panel D shows Sq-Cluster-4 irregularities. Its average profile plot shows Sq-Cluster-3 irregularities have peak amplitudes of 50 units between latitudes 5N and 20N. The occurrence frequency plot of Sq-Cluster-4 shows that Sq-Cluster-4 irregularities are mostly found in March and in November over the Atlantic sector. The green and red lines indicate that the geographic latitude of the peak amplitudes are within the geomagnetic low-latitudes.

Figure 2's panel E shows Sq-Cluster-5 irregularities. Its average profile plot shows Sq-Cluster-3 irregularities have peak amplitudes of 20 units between latitudes 20S and the equator. The location of the peak amplitudes is similar to that of Sq-Cluster-3 but the amplitudes are weaker. The occurrence frequency plot of Sq-Cluster-5 shows that Sq-Cluster-5 irregularities are mostly found from November to April over all longitudes except the Atlantic to Indian sector. The green and red lines indicate that the geographic latitude of the peak amplitudes are within the geomagnetic low-latitudes.

Figure 2's panel F shows Sq-Cluster-6 irregularities. Its average profile plot shows Sq-Cluster-6 irregularities have peak amplitudes of 10 units over latitudes 30S and 10S. The occurrence frequency plot of Sq-Cluster-6 shows that over the East Pacific and American sector, Sq-Cluster-6 profiles are mostly found between October and April while over the West Pacific, Sq-Cluster-6 profiles are found throughout the year with peak occurrence in June. The green and red lines indicate that the geographic latitude of the peak amplitudes are very far from the geomagnetic low-latitudes.

Figure 2's panel G shows Sq-Cluster-7 irregularities. Its average profile plot shows Sq-Cluster-7 irregularities have peak amplitudes of 10 units over latitudes 5S and 20N. The occurrence frequency plot of Sq-Cluster-7 shows that Sq-Cluster-7 irregularities are mostly found in March and September over all longitudes except the Pacific and American sectors. Over the American and Atlantic sectors, Sq-Cluster-7 irregularities are mostly found in April and June. Over the East Pacific and American sectors, Sq-Cluster-7 irregularities are mostly found in May and August. The green and red lines indicate that the geographic latitude of the peak amplitudes are partially over the geomagnetic low-latitudes.

Figure 2's panel H shows Sq-Cluster-8 irregularities. Its average profile plot shows Sq-Cluster-8 irregularities have minimal latitudinal variation with magnitudes of around 2 units throughout. The occurrence frequency plot of Sq-Cluster-8 shows that Sq-Cluster-8 irregularities are mostly found between March and September over the American and Indian sectors.

The occurrence frequency plots of Sq-Cluster-1, Sq-Cluster-2, Sq-Cluster-3 and Sq-Cluster-4 show that the ion density irregularities in these clusters frequently occur in months and longitudinal sectors when and where EPBs frequently happen (Huang et al, 2002; Gentile et al, 2006). These are months and sectors when the magnetic field is aligned with the solar terminator enhancing EPB rates (Tsunoda, 1985). In addition, the latitudinal location of these irregularities follow the location of the magnetic low-latitudes. This dependency with the magnetic equator is also consistent with how irregularities due to EPBs behave (Sultan, 1996). Thus, the ion density irregularities comprising Sq-Cluster-1, Sq-Cluster-2, Sq-Cluster-3 and Sq-Cluster-4 may mostly comprise of irregularities associated with EPBs.

Figure 3 shows the results of the Correlation Cluster Analysis. The figure is formatted in the same way as figure 2. Figure 3's panel A shows the Correlation cluster-1 irregularities (Corr-Cluster-1 irregularities hereafter and other Correlation clusters will follow this labeling). The average profile plot shows that Corr-Cluster-1 irregularities have peak amplitudes of 15 units between geographic latitudes 10S and 10N. The occurrence frequency plot of Corr-Cluster-1 shows that Corr-Cluster-1 irregularities are mostly found over the East Pacific, American and Asian sector between February and April as well as between August and November. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-1 irregularities are mostly within the geomagnetic low-latitudes.

Figure 3's panel B shows Corr-Cluster-2 irregularities. The average profile plot shows that Corr-Cluster-2 irregularities have peak amplitudes of 12 units between latitudes 10S and 5N. The occurrence frequency plot of Corr-Cluster-2 shows that Corr-Cluster-2 irregularities are mostly found over the Asian, Pacific and the American sectors during equinox seasons. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-2 irregularities are not within the geomagnetic low-latitudes.

Figure 3's panel C shows Corr-Cluster-3 irregularities. The average profile plot shows that Corr-Cluster-3 irregularities have peak amplitudes of 9 units between latitudes 20S and 5S. The occurrence frequency plot of Corr-Cluster-3 shows that over the Asian sector, Corr-Cluster-3 profiles are found throughout the year with the most occurrences in June. It also shows that over the American sector, Corr-Cluster-3 profiles are found between October and March. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-3 irregularities are not within the geomagnetic low-latitudes.

Figure 3's panel D shows Corr-Cluster-4 irregularities. The average profile plot shows that Corr-Cluster-4 irregularities have peak amplitudes of 12 units over latitudes 5N and 20N. The occurrence frequency plot of Corr-Cluster-4 shows that Corr-Cluster-4 profiles are mostly found over the American sector from January till June. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-4 irregularities are partially within the geomagnetic low-latitudes.

Figure 3's panel E shows Corr-Cluster-5 irregularities. The average profile plot shows that Corr-Cluster-5 irregularities have peak amplitudes of around 8 units over latitudes 30S and 20S. The occurrence frequency plot of Corr-Cluster-5 shows that Corr-Cluster-5 profiles are mostly found over the Asian sector from April to August as well as over the American sector from August to April. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-5 irregularities are very far from the geomagnetic low-latitudes.

Figure 3's panel F shows Corr-Cluster-6 irregularities. The average profile plot shows that Corr-Cluster-6 irregularities have peak amplitudes of 9 units over latitudes 30S and 25S. The occurrence frequency plot of Corr-Cluster-6 shows that Corr-Cluster-6 profiles are also mostly found over the Asian sector from April to August as well as over the American sector from August to April. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-6 irregularities are very far from the geomagnetic low-latitudes.

Figure 3's panel G shows Corr-Cluster-7 irregularities. The average profile plot shows that Corr-Cluster-7 irregularities have peak amplitudes of 6 units over latitudes 10N and 25N. The occurrence frequency plot of Corr-Cluster-7 shows that Corr-Cluster-7 profiles are mostly found over the American sector between April and August. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-7 irregularities are very far from the geomagnetic low-latitudes.

Figure 3's panel H shows Corr-Cluster-8 irregularities. The average profile plot shows that Corr-Cluster-8 irregularities have peak amplitudes of 4 units over latitudes 20N and 30N. The occurrence frequency plot of Corr-Cluster-8 shows that Corr-Cluster-8 profiles are mostly found over the American sector between April and August. The green and red lines indicate that the geographic latitude of the peak amplitudes in Corr-Cluster-8 irregularities are very far from the geomagnetic low-latitudes.

Unlike the results of the Square Euclidean cluster analysis, only one cluster, Corr-Cluster-1 has occurrence frequency plots showing ion density irregularities frequently occurring in months and longitudinal sectors when and where EPBs frequently happen. The latitudinal location of Corr-Cluster-1 irregularities' peak amplitude is also within the geomagnetic low-latitudes. The other clusters' occurrence frequency plots indicate that most of the irregularities in the clusters don't occur in months and longitudinal sectors when and where EPBs frequently happen.

4. Discussion and Conclusions

This work explores the results of applying Square Euclidean and Correlation k-means cluster analysis on ion density irregularity profiles observed by the Advanced Ionospheric Probe (AIP) onboard the Formosat-5 (F-5) satellite from November 2017 to November 2020. To isolate the ion density irregularities, the profiles were first filtered to give amplitude profiles whose regions of highest amplitudes coincide with the strongest irregularities. These were then subject to the cluster analyses. Of the 8 clusters, the Square Euclidean k-means cluster analysis was able to classify the irregularities into 4 clusters that exhibited characteristics consistent with irregularities due to EPBs. 2 clusters namely Sq-Cluster-1 and Sq-Cluster-2 mostly comprised of ion density irregularities occurring over the geographic equator and during equinox seasons. They also occurred over Eastern American sector where the geographic equator can be found within the geomagnetic low-latitudes. The only difference between the 2 clusters is in the amplitudes. Another cluster labelled Sq-Cluster-3 mostly comprised of ion density irregularities occurring over the southern geographic low-latitudes and during equinox seasons. They also occurred over the Western American sector where the southern geographic low-latitudes can be found within the geomagnetic low-latitudes. Finally, Sq-Cluster-4 mostly comprised of ion density irregularities occurring over the northern low-latitudes and during equinox seasons. They also occurred over Atlantic sectors where the northern geographic low-latitudes can be found within the geomagnetic low-latitudes.

Of the 8 clusters, the Correlation k-means cluster analysis was only able to yield one cluster that had characteristics consistent with EPBs. This was Corr-Cluster-1. Similar to both Sq-Cluster-1 and Sq-Cluster-2, Corr-Cluster-1 mostly comprised of ion density irregularities occurring over the equator and during equinox seasons. The Correlation k-means cluster analysis was also able to yield clusters that have, for example, peaks over the northern and southern low-latitudes. However, the longitudinal location of the peaks isn't within the geomagnetic low-latitudes.

Previous studies grouped ion density profiles in terms of hyper-parameters such as ion density dip magnitude, depletion edge criteria or length of depletion (Huang et al, 2001; Burke et al, 2004; Gentile et al, 2006). However, the chosen values of these parameters are based on looking at just a few ion irregularity profiles. With a cluster analysis, the algorithm looks at all of the ion irregularity profiles before grouping them. This is a more data-driven approach and is thus more advantageous.

These results suggest that a cluster analysis preferably a Square Euclidean cluster analysis can be used to find and classify ion density irregularities consistent with EPBs. This work also shows that the F5/AIP is able to make multi-year observations of ion density irregularities at ~710 km altitude and at ~2230 pm local-time that are consistent with irregularities due to EPBs.

Acknowledgements

This research was supported by grants 108-2636-M-008-002, 109-2636-M-008-004, and 110-2636-M-008 -002 from the Taiwan Ministry of Science and Technology to L.C.C. and C.C.J.H.S., as well as the Higher Education SPROUT grant to the Center for Astronautical Physics and Engineering from the Taiwan Ministry of Education. The Formosat-5 Advanced Ionospheric Probe ion density data used in this study may be downloaded free from http://sdc.ss.ncu.edu.tw/ after registering.

References

Basu, S., Basu, S., Valladares, C. E., Yeh, H. C., Su, S. Y., MacKenzie, E., ... & Bullett, T. W. (2001). Ionospheric effects of major magnetic storms during the International Space Weather Period of September and October 1999: GPS observations, VHF/UHF scintillations, and in situ density structures at middle and equatorial latitudes. Journal of Geophysical Research: Space Physics, 106(A12), 30389-30413.

Burke, W. J., Gentile, L. C., Huang, C. Y., Valladares, C. E., & Su, S. Y. (2004). Longitudinal variability of equatorial plasma bubbles observed by DMSP and ROCSAT-1. Journal of Geophysical Research: Space Physics, 109(A12).

Chao, C. K., Su, S. Y., & Liu, C. H. (2020). Initial nighttime ionospheric observations with advanced ionospheric probe onboard FORMOSAT-5. Advances in Space Research, 65(10), 2405-2411.

Chen, K. Y., Yeh, H. C., Su, S. Y., Liu, C. H., & Huang, N. E. (2001). Anatomy of plasma structures in an equatorial spread F event. Geophysical research letters, 28(16), 3107-3110.

Dungey, J. W. (1956). Convective diffusion in the equatorial F region. Journal of Atmospheric and Terrestrial Physics, 9(5-6), 304-310.

Gentile, L. C., Burke, W. J., & Rich, F. J. (2006). A climatology of equatorial plasma bubbles from DMSP 1989–2004. Radio Science, 41(5).

Heelis, R. A., Stoneback, R., Earle, G. D., Haaser, R. A., & Abdu, M. A. (2010). Medium-scale equatorial plasma irregularities observed by Coupled Ion-Neutral Dynamics Investigation sensors aboard the Communication Navigation Outage Forecast System in a prolonged solar minimum. Journal of Geophysical Research: Space Physics, 115(A10).

Huang, C. Y., Burke, W. J., Machuzak, J. S., Gentile, L. C., & Sultan, P. J. (2001). DMSP observations of equatorial plasma bubbles in the topside ionosphere near solar maximum. Journal of Geophysical Research: Space Physics, 106(A5), 8131-8142.

Kelley, M. C., The Earth's Ionosphere: Plasma Physics and Electrodynamics, Academic, San Diego, Calif., 1989.

Kil, H. (2015). The morphology of equatorial plasma bubbles-a review. Journal of Astronomy and Space Sciences, 32(1), 13-19.

Le, G., Huang, C. S., Pfaff, R. F., Su, S. Y., Yeh, H. C., Heelis, R. A., ... & Hairston, M. (2003). Plasma density enhancements associated with equatorial spread F: ROCSAT-1 and DMSP observations. Journal of Geophysical Research: Space Physics, 108(A8).

Lin, Z. W., Chao, C. K., Liu, J. Y., Huang, C. M., Chu, Y. H., Su, C. L., ... & Chang, Y. S. (2017). Advanced Ionospheric Probe scientific mission onboard FORMOSAT-5 satellite. Terrestrial, Atmospheric & Oceanic Sciences, 28(2).

Smith, J., & Heelis, R. A. (2017). Equatorial plasma bubbles: Variations of occurrence and spatial scale in local time, longitude, season, and solar activity. Journal of Geophysical Research: Space Physics, 122(5), 5743-5755.

Su, S. Y., Yeh, H. C., & Heelis, R. A. (2001). ROCSAT 1 ionospheric plasma and electrodynamics instrument observations of equatorial spread F: An early transitional scale result. Journal of Geophysical Research: Space Physics, 106(A12), 29153-29159.

Sultan, P. J. (1996). Linear theory and modeling of the Rayleigh-Taylor instability leading to the occurrence of equatorial spread F. Journal of Geophysical Research: Space Physics, 101(A12), 26875-26891.

Tsunoda, R. T. (1985). Control of the seasonal and longitudinal occurrence of equatorial scintillations by the longitudinal gradient in integrated E region Pedersen conductivity. Journal of Geophysical Research: Space Physics, 90(A1), 447-456.

Wan, X., Xiong, C., Rodriguez-Zuluaga, J., Kervalishvili, G. N., Stolle, C., & Wang, H. (2018). Climatology of the occurrence rate and amplitudes of local time distinguished equatorial plasma depletions observed by Swarm satellite. Journal of Geophysical Research: Space Physics, 123(4), 3014-3026.

Wu, J. (2012). Advances in K-means clustering: a data mining thinking. Springer Science & Business Media.

Yeh, H. C., Su, S. Y., Yeh, Y. C., Wu, J. M., Heelis, R. A., & Holt, B. J. (1999). Scientific mission of the IPEI payload onboard ROCSAT-1. Terrestrial, Atmospheric and Oceanic Sciences, (1), 19-42.

Yokoyama, T. (2017). A review on the numerical simulation of equatorial plasma bubbles toward scintillation evaluation and forecasting. Progress in Earth and Planetary Science, 4(1), 1-13.

Figure captions

Figure 1: A) Formosat-5 AIP orbital profiles for one day. B) Formosat-5 AIP orbital profiles for two days. C) Unfiltered and filtered sample AIP ion profile as well as the amplitude calculated from the filtered profile as a function of geographic latitude. D) Total number of profiles for each longitude-month bin accumulated between November 2017 and November 2020.

Figure 2: Each panel corresponds to a cluster (e.g. cluster-1 is in panel A, cluster-2 is in panel B, etc). The left plot of each panel shows the average of all profiles under the given cluster as a function of geographic latitude. The errorbars are the standard deviation of the profiles. Note that the y-axis for the left plots differ per panel. The right plot of each panel shows the number of profiles found in each month and longitude bin. The red line in each panel's left plot is situated over the average geographic latitude of magnetic inclination angle 0° for the given cluster. The green line to the left (right) of the red line in each left plot is situated over the average geographic latitude of magnetic inclination angle -15° (15°) for the given cluster. See text for more details.

Figure 3: Same as figure 2 but for the Correlation cluster analysis.