Forecasting Equatorial Plasma Bubbles Based on K-Nearest Neighbor Approach at Chumphon, Thailand

Punyawi Jamjareegulgarn[1*], Lin Min Min Myint[2a], Phimmasone Thammavongsy[2b], Pornchai Supnithi[2c]

[1] King Mongkut's Institute of Technology Ladkrabang, Prince of Chumphon Campus, Chumphon 86160, Thailand, Email: kjpunyaw@gmail.com
[2] School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand, Email: linminmin.my@kmitl.ac.th[a], phim.thmvs@gmail.com[b], pornchai.supnithi@gmail.com[c]

Abstract: This paper proposes a forecasting model for range spread-F (RSF) events relied on the indices of solar and geomagnetic activities along with K-nearest neighbor approach. During the studied years 2013-2014, the indices of solar and geomagnetic activities were retrieved from the OMNIWEB, while the RSF values were scaled manually on the ionograms obtained from the SEALION's ionosonde at Chumphon station, Thailand. In this work, the mentioned indices consist of Kp-index, Ap-index, Dst index, 10.7 solar flux and sunspot number. Either RSF existence or RSF absence is employed to indicate whether the equatorial plasma bubble (EPB) occur or not. Hence, the scaled RSF values represented by two binary numbers (1 or 0) identifies the presence or absence of EPB phenomena, respectively. The RSF existence and absence in year 2013 and 2014 are used as the training and the testing datasets, respectively. Our results show that this KNN-based prediction model for EPB occurrence can be used an alternative practically due to its accuracy of greater than 85%.

Keywords—**Ap-index, Dst index, plasma bubble, F10.7, Kp-index, K-nearest neighbor, RSF, sunspot number.**

I. INTRODUCTION

PLASMA irregularities are recognized to be the magnetic field aligned density structures in the nighttime F-layer. They are frequently shown as the spreading in the F-layer traces of ionogram in range and frequency, namely "spread-F". Both the range spread-F (RSF) and frequency spread-F (FSF) have been the widespread studies at middle and low latitudes. Meanwhile, the RSF has been focused investigating at the equatorial latitude, so-called equatorial spread-F (ESF) [1]. The ESF phenomena is known to related to the equatorial plasma bubbles (EPB). The scale sizes of EPB structure range from meter to hundreds of kilometers. The ESF is stimulated to generate by the urgent recombination in F-layer after sunset resulting in the depletion of plasma bubble density where it rises up from the bottomside F-layer and penetrates toward the topside F-layer [2]. The ESFs are impacted from complex processes within the ionosphere where a main electrodynamic mechanism leading to the ESF onset is the generalized Rayleigh-Taylor instability (RTI) process. The rapid raising of equatorial F-layer is also driven by the pre-reversal

enhancement (PRE) during the evening and the post-sunset [3]. The formation of ESFs or EPBs can interrupt the high frequency (HF) communication and navigation by leading to the sudden amplitude and phase fluctuation (scintillation) on the HF signals. Therefore, the prediction of EPBs has been the issue of major endeavors and has drawn much attention among scientific communi-ties for better comprehension and prior knowledge [4].

Recently, the EPBs relating to the ESFs were studied by using an all-sky imager (ASI) during the 23rd-24th solar cycle over Brazilian sector. This investigation reported that EPB occurrences in every season correlate positively with the 10.7 cm solar radio flux (F10.7) [5]. In addition, Buhari et al. [6] investigated the EPB variabilities based on solar and geomagnetic activities over Malaysia during the years of 2008-2013. They disclosed that most of EPBs often generate and relate to the intense solar activities, but they don't depend on geomagnetic activities. Likewise, Abdu et al. [7] and Abdu [8] reported that the ESF events frequently occur at the same time periods as the EPB phenomena and are often generated when the electric fields are disturbed by geomagnetic storm. It shows that the geomagnetic indices (e.g., the F10.7 index, the sunspot number (R), etc.) are related directly to the EPB existence during the geomagnetic storm time which does not corresponds to the results of Buhari et al. [6]. Hence, the authors have an idea to investigate the impacts of solar and geomagnetic activities on RSF existence at Chumphon station during the years 2013-2014. The instantaneous RSF existence or nonexistence is a hypothesis that is possibly expected to

notify the instantaneous EPB occurrence or absence as well. Hence, the benefit of this paper is to employ this technique for predicting formerly the existence and the notification of EPBs. The structure of this paper are as follows. Data used in this work is stated in Section II. The indices of solar and geomagnetic activities as well as the k-nearest neighbor (KNN) used in this work are explained in Section III and IV, respectively. The results and discussion are described in Section V. Finally, the conclusion is given in Section VI.

II. DATA USED

Since 2003, the frequency modulation continuous waves (FMCW) ionosonde has operated to monitor the ionosphere at Chumphon station, Thailand (10.72°N, 99.37°E; dip latitude 3.3°N). This ionosonde station, one of the 8 SEALION (SouthEast Asia Low-Latitude Ionospheric Network) ionosondes [9], it is located at King Mongkut's Institute of Technology Ladkrabang, Prince of Chumphon Campus, Chumphon, Thailand, as depicted in Fig. 1. The operating parameters of the FMCW ionosonde are shown in Table 1. The recorded ionograms were retrieved and processed every 15 minutes from 18.00 - 06.00 LT for each day during 2013-2014. The RSF values related directly to EPBs are determined manually on each ionogram regarding to the scaling manual [10]. The scaled RSF values during 2013-2014 are represented by two binary numbers (1 or 0) that indicates the existence or absence of RSF phenomena, respectively [11]. The left panel of Fig. 2 show the normal trace in an ionogram on March 2, 2016, while the right panel of Fig. 2 show the RSF (trace spreading) at 04:00 UT on March 3, 2016 at Chumphon station [4]. Also, the indices of solar and geomagnetic activities during 2013-2014 used in this work can be downloaded from https://omniweb.gsfc.nasa.gov/.

III. THE INDICES OF SOLAR AND GEOMAGNETIC ACTIVITIES

The indices of solar and geomagnetic activities are employed in this work consist of 5 parameters, for example,
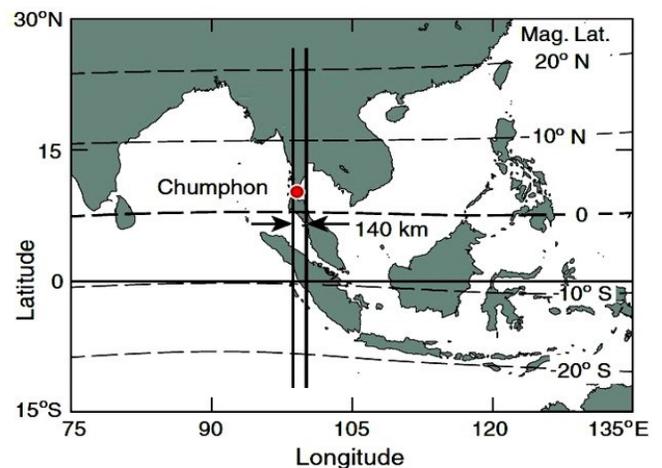

Fig. 1. Geographical Location of Chumphon, Thailand (geographical location 10.72°N, 99.37°E; dip latitude: 3.3°N) [9].

Kp-index, Ap-index, Dst index, 10.7 cm solar flux (F10.7) and sunspot number (R). Each parameter is explained briefly as follows.

A. Kp-index

Kp-index is a global parameter that is determined based on an algorithm running on all of the reported K values from every magnetometer station at the same time. The K value is a three-hour local index of the geomagnetic activity at the given time and location. The Kp-index spans from G0 to G5 as shown in Table II, for instance, G0 (0 <= Kp <= 4+) - quiet and G5 (Kp >= 9) - extreme etc.

B. Ap-index

Ap-index provides a daily average of geomagnetic activity. Due to the nonlinear relationship between the K value and magnetometer fluctuations, every three-hour K value is computed to be the linear scale a-value. Then, the average from eight daily a-values provides the Ap-index of a certain day [12]. The Ap-index covers from G0 to G5 like the Kp-index as shown in Table III such as G0 (Ap <= 39) and G5 (Ap > 300) etc.

TABLE I
THE OPERATIONAL PARAMETERS OF FMCW IONOSONDE

Ionosonde Parameters	Details/Values
Peak Tx power	20 W
Average Tx power	10 W
Frequency range	2-30 MHz
Sweep rate	100 kHz/s
Sweep repetition period	5 minutes
Antenna	Folded dipole on Tower 27 m

C. Disturbance Storm Time (Dst) Index

Dst index is a magnetic activity index that reports the level of the globally equatorial ring current obtained from lots of near-equatorial geomagnetic observatories . It is manipulated by NGDC and is available from 1957 to the present [13]. The Dst index spans from G0 to G5 as shown in Table IV, for instance, G0 (Dst >= -20 nT) as well as G5 (Dst < -320 nT) and so on.

D. F10.7 index

F10.7 index (or 10.7 cm solar flux) is a parameter of the noise level derived from the Sun that is measured at a wavelength of 10.7 cm at the Earth's orbit. It is one of the most widely used solar activity indices that is employed in several applications as follows: a) a simple activity level indicator, b) a proxy for other solar emissions or quantities that are more difficult to receive, and c) a commonly available datum for antenna calibration [14]. The F10.7 index ranges from G0 to G5 as shown in Table V such as G0 (0 < F10.7 <= 10) and G5 (F10.7 > 200) and so on.

E. The sunspot number (R)

Sunspots happen on the Sun's photosphere that appear as temporary spots that are darker than the surrounding areas. They are the regions of reduced surface temperature that are caused by concentrations of magnetic flux. Sunspots appear within active areas and the sunspot number varies about 11-year solar cycle [15]. The sunspot number (R) can be used to compute directly using the F10.7 index.

TABLE II
THE GEOMAGNETIC STORM LEVELS OF KP-INDEX

Geomagnetic Storm Levels		Kp-index Values
G0	quiet	0 <= Kp <= 4+
G1	minor	4+ < Kp <= 5+
G2	moderate	5+ < Kp <= 6
G3	strong	6 < Kp <= 7
G4	severe	7 < Kp < 9
G5	extreme	Kp >= 9

TABLE III
THE GEOMAGNETIC STORM LEVELS OF AP-INDEX

Geomagnetic Storm Levels		Ap-index Values
G0	quiet	Ap <= 39
G1	minor	39 < Ap <= 56
G2	moderate	56 < Ap <= 94
G3	strong	94 < Ap <= 154
G4	severe	154 < Ap <= 300
G5	extreme	Ap > 300

TABLE IV
THE GEOMAGNETIC STORM LEVELS OF DST-INDEX

Geomagnetic Storm Levels		Dst-index Values (nT)
G0	quiet	Dst >= -20
G1	minor	-50 <= Dst < -20
G2	moderate	-100 <= Dst < -50
G3	strong	-200 <= Dst < -100
G4	severe	-320 <= Dst < -200
G5	extreme	Dst < -320

TABLE V
THE GEOMAGNETIC STORM LEVELS OF F10.7 INDEX

Geomagnetic Storm Levels		F10.7 index Values
G0	quiet	0 < F10.7 <= 10
G1	minor	10 < F10.7 <= 50
G2	moderate	50 < F10.7 <= 100
G3	strong	100 < F10.7 <= 150
G4	severe	150 < F10.7 <= 200
G5	extreme	F10.7 > 200

IV. K-NEAREST NEIGHBOR (KNN)

Fig. 2. The normal trace in an ionogram without RSF on March 2, 2016 (left) and the RSF (trace spreading) occurred in an ionogram at 04:00 UT on March 3, 2016 (right) at Chumphon station [11].

The k-nearest neighbor algorithm (KNN) is a supervised learning method that was proposed firstly by Evelyn Fix and Joseph Hodges in 1951. The input consists of the k closest training samples in a data set in both cases, whereas the output depends on whether KNN is used for classification or regression [16]. In this work, the KNN is used for "EPB classification" to forecast whether the EPB event occur or not based on RSF consideration accordingly. If the RSF is predicted to occur, then the EPB also happen, and vice versa. Further KNN details can be read in [17]. Note that other classification methods (e.g. support vector machine, decision tree etc.) can be used and also compared to the performance of KNN methods proposed in this work.

V. Results and discussion

In this work, the RSF values are scaled manually on the ionograms obtained from ionosonde at Chumphon station. The amount of training and testing data during years 2013 - 2014 are considered as 70:30. Note that the RSF presence and absence in years 2013 - 2014 are considered, since they are in solar maximum of the recent 24th solar cycle. The results of testing dataset and EPB prediction are shown partially in Table VI with the correct predictions. We found that the proposed prediction model for RSF give the accuracy above 85%. The KNN-based predicted results in Table VI can be inferred that:

a) The R > 120 and the Kp-index > 2 are the main constraints that result in the RSF or EPB existence. Both the Ap-index and the Dst index do not impact to the RSF or EPB existence.

b) The R >= 140, the Kp-index >= 2.5, and the Ap-index > 10 are the main constraints that guarantee definitely the RSF or EPB occurrence. Here, it is show that the Ap-index can be used to identify the EPB when it is particularly higher than 10.

c) The F10.7 > 200 (extreme geomagnetic storm level or G5 level) is the main constraint that also guarantee definitely the RSF or EPB occurrence.

Table VII show partially the results of EPB occurrences with the incorrect predictions based on the proposed KNN method. The main causes of these errors are discussed as follows. Although the sunspot numbers are higher than 120 and the Ap values are larger than 10, both the positive Dst values and the low Kp values lead to very little geomagnetic storm. Therefore, the EPB doesn't exist in real

TABLE VI
THE PARTIAL RESULTS OF EPB OCCURRENCES WITH ACCURATE PREDICTIONS BASED ON KNN METHOD

No.	Kp-index	R	Dst-index	Ap-index	F10.7	Scaled RSF values	KNN-based EPB prediction
1	2.7	159	4	12	218.3	1 (RSF presence)	Yes (EPB exists)
2	2.7	159	-8	12	218.3	1 (RSF presence)	Yes
3	3.7	159	-17	22	218.3	1 (RSF presence)	Yes
4	3.7	159	-18	22	218.3	1 (RSF presence)	Yes
5	2.3	136	-4	9	186.7	1 (RSF presence)	Yes
6	2.3	136	-7	9	186.7	1 (RSF presence)	Yes
7	2.3	136	-10	7	186.7	1 (RSF presence)	Yes
8	3	107	-1	15	146.9	0 (RSF absence)	No (EPB does not exist)
9	3	107	-8	15	146.9	0 (RSF absence)	No
10	3	107	-12	15	146.9	0 (RSF absence)	No
11	3	178	-1	15	244.4	1 (RSF presence)	Yes
12	3	178	3	15	244.4	1 (RSF presence)	Yes
13	3	178	2	15	244.4	1 (RSF presence)	Yes
14	0.7	85	-11	3	116.7	0 (RSF absence)	No
	0.7	85	-10	3	116.7	0 (RSF absence)	No
16	5	125	-1	12	165.3	1 (RSF presence)	Yes
17	5	125	3	12	165.3	1 (RSF presence)	Yes
18	5	125	2	12	165.3	1 (RSF presence)	Yes
19	0.7	105	-5	3	125.4	0 (RSF absence)	No
20	0.7	105	-10	3	125.4	0 (RSF absence)	No

TABLE VII
THE PARTIAL RESULTS OF EPB OCCURRENCES WITH ACCURATE PREDICTIONS BASED ON KNN METHOD

No.	Kp-index	R	Dst-index	Ap-index	F10.7	Scaled RSF values	KNN-based EPB prediction
21	2.1	125	12	7	170.2	0 (RSF absence)	Yes (EPB exists)
22	2.1	125	10	7	170.2	0	Yes
23	2.5	133	-2	12	182.6	0	Yes
24	2.5	133	-5	12	182.6	0	Yes
25	1.5	100	-20	10	121.5	0	Yes

situation. In contrast, the proposed KNN method predicts in case of "Yes (EPB exists)". Note that although the R can be used to compute the F10.7 directly, the F10.7 computations are changed with respect to the solar activity level. Hence, the R equations should have at least two different equations to compute the F10.7 regarding to low and high solar activities, respectively [18].

The results of this proposed RSF prediction model are good agreement to the results of Thammavongsy et al. [4] such that the levels of F10.7 solar flux values (or the sunspot number) lead to RSF existence and also influence the frequency of RSF durations. It can be seen that the geomagnetic indices (particularly, the F10.7 index, the sunspot number (R), and the Kp-index) are related directly to the RSF (or EPB) existence during the geomagnetic storm time. Our results are good agreement to the results of Abdu et al. [7] and Abdu [8].

VI.   CONCLUSION

A prediction model for range spread-F (RSF) event using the indices of solar and geomagnetic activities as well as K-nearest neighbor method is proposed in this paper to forecast whether the EPS exist or not during the selected years 2013-2014. The indices of solar and geomagnetic activities were retrieved from the OMNIWEB including Kp-index, Ap-index, Dst index, 10.7 cm solar flux and sunspot number. Meanwhile, the RSF values were scaled manually on the ionograms at Chumphon station, Thailand. Since the RSF existence or absence is used to indicate accordingly whether the equatorial plasma bubble (EPB) occur or not. Hence, two binary numbers (0 or 1) are also implemented to identify achievably the nonexistence or existence of EPB phenomena, respectively. The studied results show that the KNN method can be used to forecast the EPB achievably with the accuracy above 85%. In the near future, we intend to study more years and compare the KNN method with the neural network (NN) approach as proposed in [11].

REFERENCES

[1] M. A. Abdu, *et al.*, "Equatorial spread F echo and irregularity growth processes from conjugate point digital ionograms," in *URSI GASS*, pp. 1-4, 2011.

[2] D. L. Hysell, "An overview and synthesis of plasma irregularities in equatorial spread F," *J. Atmos. Sol. Terr. Phys.*, vol. 62, pp. 1037-1056, 2000.

[3] M. A. Abdu, "Outstanding problems in the equatorial ionosphere-thermosphere system relevant to spread F," J. *Atmos. Terr. Phys.*, vol. 63, pp. 869-884, 2001

[4] P. Thammavongsy, *et al.*, "The Statistics of Equatorial Spread-F and Effects on Critical Frequency at Chumphon, Thailand," *Adv. Space Res.*, vol. 172, 2019.

[5] E. Agyei-Yeboah, *et al*., "Seasonal variation of plasma bubbles during cycle 23-24 over the Brazilian equatorial region.," *Adv. Space Res.*, vol. 19, 2019.

[6] S.M. Buhari, *et al*., "The variations of equatorial plasma bubble with solar and geomagnetic activities in Malaysia (2008-2013)," in *United Nations/Nepal Workshop on the App. of GNSS*, Dec. 12-16, 2016.

[7] M. A. Abdu, R. T. de Meiros, J. H. A. Sobrel, J. A. Bittencourt, "Spread F plasma bubble vertical rise velocities determined from spaced ionosonde observations," *J. Geophys. Res*., pp. 9197–9204, 1983.

[8] M.A. Abdu, "Equatorial spread F/plasma bubble irregularities under storm time disturbance electric fields," J. Atmos. Sol. Terr. Phys., vol. 75–76, 2011.

[9] T. Maruyama, "Low latitude ionosphere-thermosphere dynamics studies with ionosonde chain in Southeast Asia," Ann. Geophys., vol. 25, 2007.

[10] N. Wakai, H. Ohyama, and T. Koizumi, "Manual of Ionogram Scaling Third Version.," in *Radio Research Laboratory*, 1987.

[11] P. Thammavongsy, *et al.*, "Spread-F prediction model for the equatorial Chumphon station, Thailand," *Adv. Space Res*., vol. 65, pp. 152-162, 2020.

[12] https://www.spaceweatherlive.com/en/help/.

[13] https://www.ngdc.noaa.gov/stp/geomag/dst.html.

[14] https://spawx.nwra.com/spawx/f10.html.

[15] https://en.wikipedia.org/wiki/Sunspot.

[16] https://en.m.wikipedia.org/wiki/K-nearest_neighbors_algorithm.

[17] https://towardsdatascience.com/a-simple-introduction-to-k-nearest-neighbors-algorithm-b3519ed98e.

[18] D. Okoh and E. Okoro, "On the Relationships between Sunspot Number and Solar Radio Flux at 10.7 Centimeters," *Sol. Phys.*, vol. 295, no. 1, pp. 1-13, 2020.

Punyawi Jamjareegulgarn received Ph.D. in Department of Electrical Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand, in 2017. Since 2012, he has worked as an Associate Professor in Department of Engineering, King Mongkut's Institute of Technology Ladkrabang, Prince of Chumphon Campus, Chumphon, Thailand. His research interests include ionospheric irregularities, TEC modeling, RTK-based positioning and multi-GNSS signals, plasma bubbles, and earthquake investigations based on GPS TEC and GIM TEC.

Lin Min Min Myint received Ph.D. in Department of Electrical Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand. He has worked as an Assistant Professor at School of International Interdisciplinary Engineering (SIIE), School of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand. His research interests include signal processing on magnetic storage, GNSS positioning and space weather analysis.

Phimmasone Thammavongsy received Ph.D. in Department of Electrical Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2023. He has worked at National University of Laos, Vientiane, Laos.

Pornchai Supnithi received Ph.D. in electrical Engineering from Georgia Institute of Technology, USA in 2002. He is currently a full professor at Telecommunication Engineering Department, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok Thailand. His research interests are in the area of Telecommunications, Ionospheric and GNSS, Data storage and Engineering Education.