Version of Record: https://www.sciencedirect.com/science/article/pii/S0273117719307185
Manuscript_71acf4d8145c38cce2220a86d110ecb8

Spread-F prediction model for the equatorial Chumphon station, Thailand

P. Thammavongsy[a], P. Supnithi[a,*], W. Phakphisut[a], K. Hozumi[b], T. Tsugawa[b]

[a] *Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand*
[b] *Space Environment Laboratory, National Institute of Information and Communication Technology, Nukuitita, Koganei, Tokyo 184-8795, Japan.*

Abstract

This work proposes a range spread-F (RSF) prediction model using the neural network (NN) over the equatorial Chumphon (CPN) region in Thailand. The RSF model is constructed by using five input spaces including the diurnal variation, seasonal variation, geographic latitude, solar flux index ($F10.7$), and magnetic index (A_p). The RSF NN model is trained with three years of RSF data during 2013 to 2015 from Chumphon (CPN) station (Latitude = $10.7°N$, Longitude = $99.4°E$) and the performance of the proposed RSF NN model is validated using the dataset of 2016. As a result, the RSF NN model achieves 98.3% accuracy of all correct predictions even with the limited available data. The results show that the proposed NN model yields a lower RSF probability than the actual observation by about 7.3%, but the overestimation of the proposed NN model is 2.5% in both the equinoxes and solstices. In addition, we discover that the IRI-2016 model mostly overestimates the RSF probability when compared with the actual observation for all seasons in 2016, particularly, in equinoctial months over Chumphon station.

Keywords: Ionospheric Irregularity; Equatorial Spread-F; Spread-F Prediction Model; Neural Network; IRI-2016 model

1. Introduction

The nighttime irregularity of the equatorial ionosphere is a major issue that degrades the high frequency radio wave propagation. Such irregularity at the bottomside ionosphere is often displayed as the equatorial spread-F (ESF). The spread out of the equatorial F layer was first studied by Booker & Wells (1938) who discovered the scattering of F layer trace in the ionograms at nighttime near the magnetic equator. The spreading of the F layer along the frequency band and height range is due to the reflection of scattered signals from the irregularity of ionospheric electron densities and the spread sizes of the plasma irregularities vary from a few centimeters to hundreds of kilometers Basu et al. (1978). Rungraengwajiake et al. (2013) found out that the percentage of RSF occurrences is higher in low latitudinal region (Chumphon station: CPN) than in the mid latitudinal region (Chiangmai station: CMU) over Thailand sector and the onset of RSF is observed at CPN station before the station at conjugate points. Hence, it confirms that the onset of RSF/plasma bubble initiates from equatorial region and afterward extend to the higher latitudes (equatorial to northern) in Thailand sector. The study of Klinngam et al. (2015) analyzes ESF occurrence statistics during equinoctial months from 2006 to 2013 over the three conjugate stations in Southeast Asia; the RSF occurrence at Chumphon station is higher than those at Chiangmai station and Kototabang station, specifically, the

*Corresponding author. Tel.: +66 2 329 8000; fax: +66 2 329 4554.
Email addresses: 61601189@kmitl.ac.th (P. Thammavongsy), pornchai.su@kmitl.ac.th (P. Supnithi), watid.ph@kmitl.ac.th (W. Phakphisut), kukkai@nict.go.jp (K. Hozumi), tsugawa@nict.go.jp (T. Tsugawa)

most observed RSF occurrences are noticed within 5% to 85% in the equinoctial months. The previous works analyze the characteristics of RSF occurrence in low latitudinal areas are more predominant than the high latitudinal areas. Its effects are a major issue to degrade the quality level of communication systems such as the reduction in the performance of high frequency (HF) communication systems and the Global Navigation Positioning System (GNSS).

The plasma instability is caused by the uplifting of bottom side F-region plasma gradient and the action of the Rayleigh-Taylor (R-T) gravitation instability processes. The ESF/plasma bubble is driven by the evening pre-reversal enhancement (PRE) that causes the sudden uplift of equatorial F layer and the ESF occurrences lead to instability growth at the bottomside of the F-region so called the plasma bubble event Abdu (2001); Abdu et al. (2009). Hoang et al. (2010) proposes the statistical results of RSF occurrence over the two equatorial sites at Ho Chi Minh City (HCM) in Vietnam and Sao Luis (SL) in Brazil; one interesting result is the ESF occurrence rate at SL station is higher than HCM station during post-midnight in December and July months, in addition, the height of F layer over HCM station is lower than those over SL station. Pezzopane et al. (2013) presents the study of ESF phenomenon during equinoctial months in 2009 over American longitude sector including Palmas (PAL) and São José Dos Campos (SJC) in Brazil, Tucumán (TUC), Argentina and in Asian longitude sector composes of Chiangmai (CGM), Thailand, and Tanjungsari (TNJ), Indonesia. The observed ESF results in American sites are higher than those observed ESF occurrence rate at CGM and TNJ sites in the Asian longitude sector, moreover, this study confirms the fact of RSF occurrence at TUC, anticipated by satellite trace (ST) appearance, it is most probably related to gravity waves propagating through the F-region rather than the equatorial process. The previous studies indicate that the occurrence and development of ESF event depend on the several factors such as LSWS, PRE and PSSR. The study of Abdu et al. (2015) shows that the post-sunset development of spread F/plasma bubble is due to the amplification of the large-scale wave structure (LSWS) at the bottomside of equatorial ionospheric F-layer and also, the vertical/zonal perturbation winds is a common feature in the equatorial region for producing the form of F layer height oscillations. These two causes lead to development of spread-F/plasma bubble during the post-sunset. Huang (2018) proposes the pre-reversal enhancement (PRE) of the post-sunset vertical plasma drift is a significant factor in controlling the generation of ESF and the 100% of ESF occurrence probability can be actually observed when the PRE is large enough. The post-sunset rise (PSSR) of equatorial F layer is replaced by the upwelling growth for controlling the development of equatorial plasma bubble/spread-F during the low solar activity Tsunoda et al. (2018). However, the mechanism of ESF occurrences are not well understood, especially on the day-to-day variability.

Since the equatorial spread-F degrades quality of communication systems, it is reasonable in attempting to prevent this problem and develop a prediction model. One of well-known empirical models is the International Reference Ionosphere (IRI) model which has been continually updated; the current model is the IRI-2016 model `https://ccmc.gsfc.nasa.gov/modelweb/models/iri2016_vitmo.php`. Abdu et al. (2003) first develops the prediction model for the spread-F occurrence probability using the 13 years of spread-F data (1978 - 1990) over the equatorial site Fortaleza and low latitude site Cochoeira in Brazilian sectors; the model uses the method of Cubic-B spline interpolation. This model is being used as an optional parameter of the IRI model. Neural networks have been recently utilized to forecast bottomside ionosphere and spread-F events. Oyeyemi et al. (2005) proposes a prediction model for $foF2$ value using the neural network, which is trained with a large dataset from 59 ionospheric stations from 1964 to 1986. They indicate that the results of proposed NN2 model yield the better results than the Comité Consultatif International

des Rediocommunications (CCIR) model on average by a margin of 15 - 16%. Oyeyemi et al. (2006) suggests the improvement in the $foF2$ prediction model in that geophysical information and the recent-past observations of $foF2$ from several control stations which could be utilized as inputs of the NN for predicting the $foF2$ in a near-real time fashion. The NN is utilized to develop the prediction model of $M(3000)F2$, the designed network model consists of two hidden layers that contain 20 neurons and the response is compared with IRI model (CCIR $M(3000)F2$ model) Oyeyemi et al. (2007). McKinnell & Oyeyemi (2010) proposes the new NN based global model for the $foF2$ parameter concentrating on the ability of predicting the $foF2$ in the equatorial sector, this model uses the data from 135 global ionospheric stations including a number of equatorial stations. Equatorial stations are identified as problematic within the current IRI prediction, hence, this new model is considered as a suitable replacement for the International Union of Radio Science (URSI) and International Radio Consultative Committee (CCIR) model which are currently used within the IRI model with the objective of predicting the F2 peak electron density. Watthanasangmechai et al. (2012) proposes the prediction of Total Electron Content (TEC) utilizing the NN over equatorial station so called CPN station in Thailand. It is trained with the global positioning system (GPS) available data during the low solar activity from 2005 to 2009 and the training algorithm of network model is a back propagation algorithm. The three input spaces such as the day number, hour number, and sunspot number are used. The proposed model can predict the GPS total electron content (TEC) quite well over equatorial latitude station, even with the limited amount of available data. In 2019, Zhao et al. (2019) optimizes the NN based model of predicting the $foF2$ value using the Genetic Algorithm (GA) during disturbance time in China. GA is used to improve the initial weights to escape the local minimum during the training. The training data covers 13 years from 1995 to 2004. The prediction of the proposed model is more accurate in 1 hour ahead of prediction and furthermore, the prediction accuracy is better at middle latitude than low latitude. Importantly, the proposed NN model works well in several cases where the observed values are far away from the monthly median value. The development of ESF prediction model is proposed by McKinnel et al. (2010) utilizing the artificial neural network (ANN) which is trained using 12 years of range spread-F data (1978 - 1989) from Fortaleza and Cochaeira Paulista stations in Brazil and the design of NN model consists of seven input parameters as the geographic latitude, magnetic inclination, magnetic declination, day number, hour number, solar activity (R3), and magnetic activity (A8). This work indicates that the diurnal, seasonal, and solar cycle variability in spread-F occurrence is fairly well characterised statistically. Based on previous studies, the several solutions have been proposed to solve the ionospheric issues posed by the ionosphere through prediction.

Therefore, this study aims to develop spread-F model using the NN algorithm for predicting the percentage of spread-F occurrences at Chumphon, Thailand. We use the spread-F data from 2013 to 2015 during high solar activity phase of the 24^{th} solar cycle over the equatorial CPN station (Lat $= 10.7°N$, Lon $= 99.4°E$) in Thailand. Furthermore, this study demonstrates the comparison of spread-F NN model with the actual observed spread-F and the evaluation of IRI-2016 model performance is discussed in terms of servicing the forecast result of spread-F probability during descending solar activity in 2016 at CPN station.

There are four main sections in this paper. Section 1 discusses the motivation, previous works and objectives in this work. Section 2 describes the methodology of the designed NN model. Section 3 shows the results and the last section. Section 4 gives the conclusion.

2. Methodology

2.1. Data preparation

The spread-F data or target data is scaled from the ionograms of frequency modulation/continuous waves (FM/CW) ionosonde at CPN station as demonstrated in Fig.1. The measured RSF data covers a four-year period from 2013 to 2016 during the 24^{th} solar cycle with data availability as exhibited in Fig. 2. There are some obtained input data from the National Oceanic and Atmospheric Administration (NOAA) data center `ftp://ftp.swpc.noaa.gov/pub/warehouse/` including the solar activity ($F10.7index$) and the magnetic activity (A_p). The parameters of the diurnal variation and the seasonal variation are generated using the cyclic components of the sine and cosine functions of the hour and day numbers. In addition, the entire input parameters are scaled into range between -1 and 1resulting in a better convergence of training the neural network Zhao et al. (2019). The entire prepared inputs are grouped and arranged appropriately following the target/ESF data format. We choose the three years of data set from 2013 to 2015 for training the NN model, specifically, the training data set is partitioned randomly into the ratio of 70% for the training set and 30% for the testing set. To validate the network model, the data set of 2016 is assigned as an unknown data set for testing the designed NN model.

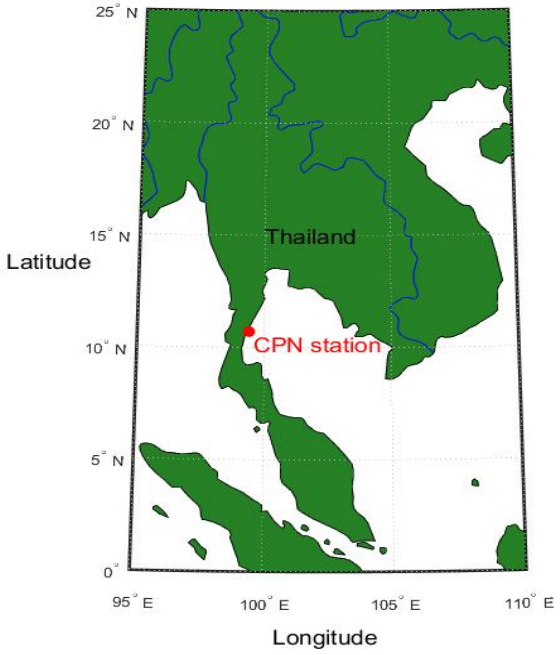

Figure 1: Geogrphical location of FM/CW ionosonde at CPN station.

2.2. Designing the structure of the neural network

The determination of the NN structure influences the desired performance of the designed NN structure and, furthermore, to overcome the major issues of training the NN model such as overfitting and underfitting. Each designed NN structure is trained with the primary input parameters consisting of the seasonal variation, diurnal variation, solar activity ($F10.7$) index and magnetic activity (A_p) index. The designed network structures are implemented on the three different networks which consist of a single hidden layer, two hidden layers and three hidden layers. Also, the

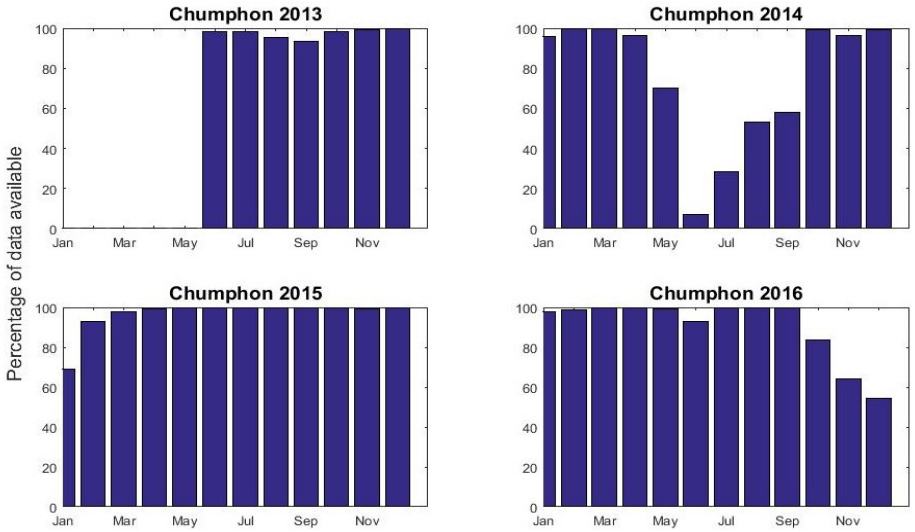

Figure 2: Availability of the spread-F data during 2013 to 2016 at CPN station.

neurons of each designed network are orderly adjusted by 5 interval neurons increasing from 10 neurons to 60 neurons, respectively. The initial weight values are randomly defined and the training times are repeated by 100 iterations, the network's error goal is set to 0.001 and the learning rate of network is 0.003. To evaluate the network model, the trained network performance is evaluated by the root mean square error ($RMSE$), i.e.,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{pred}(i) - P_{obs}(i)\right)^2}, \tag{1}$$

where $P_{obs}(i)$ represents the observed output or the actual output at time i, $P_{pred}(i)$ is the predicted output of NN model at time i and n is the total number of the data points.

2.3. Determination of the input parameters

We determine the input parameters depending on its relationships that influence and associate with the equatorial spread-F phenomenon. Following the numerous input parameters as successfully used in Abdu et al. (2003); McKinnell & Oyeyemi (2010), we then apply some input parameters such as the diurnal variation, seasonal variation, geographic latitude, solar activity ($F10.7$ radio flux) index and the magnetic activity (A_p) index and additionally, we adjust the $F10.7$ and A_p indices to the several levels to find the best input parameters. Therefore, in this work, we can optimize the input parameters as described below.

2.3.1. Diurnal variations

Usually, the occurrence of ESF event is observed only during the post-sunset to the morning. The development of ESF is caused by recombination of the ionospheric plasma density after sunset and its behaviors are varied following the diurnal variation. The diurnal variation is a basic factor of interpreting the variability of the ionospheric layer depending on time variation; it is successfully applied in variety of the space prediction models Watthanasangmechai et al. (2012); Zhao et al. (2019). Therefore, the diurnal variation is represented by the universal time (UT) from 1 to 24 over CPN station (LT+7). The diurnal variation is generated using the cyclic components of sine and cosine functions in universal time to give the continuous data as

$$HrS = \sin\left(\frac{2\pi * HrN}{24}\right), \tag{2}$$

and

$$HrC = \cos\left(\frac{2\pi * HrN}{24}\right), \tag{3}$$

where HrN is the hour number of universal time (1 - 24).

2.3.2. Seasonal variations

The RSF occurrences are generally found to be higher in the equinoctial months than the solstice months during minimum solar activity during 2006 to 2013 at CPN region Klinngam et al. (2015). The seasonal variation is thus used widely to describe the variability of RSF occurrences in each season Oyeyemi et al. (2005); Abdu et al. (2003). It can be generated by using cyclic components of sine and cosine functions as well as the diurnal variation, i.e.,

$$DNS = \sin\left(\frac{2\pi * DN}{365}\right), \tag{4}$$

and

$$DNC = \cos\left(\frac{2\pi * DN}{365}\right), \tag{5}$$

where DN is the day number (1 - 365).

2.3.3. Geographic latitude

The ESF occurrence is observed in low latitudinal region more often than the high latitudinal region, particularly, the area closes to the magnetic equator Rungraengwajiake et al. (2013); Pezzopane et al. (2013); Klinngam et al. (2015). Therefore, the geographic latitude is considered as an input of space parameters fed to the proposed NN model. The study area is referred to the geographic latitude of CPN station.

2.3.4. Solar activity ($F10.7$ index)

The solar activity is an indicator of the ionospheric parameters which can imply numerous irregularities within the ionospheric layer. Pietrella et al. (2017) study the occurrences of ESF at different longitude sectors under moderate and high solar activities. Here, we use the $F10.7$ radio flux index which has been used in the ionospheric prediction models McKinnell & Oyeyemi (2010); Hu & Zhang (2018). The dataset of $F10.7$ radio flux index is provided by world data center (WDC) from National Oceanic and Atmospheric Administration (NOAA) `ftp://ftp.swpc.noaa.gov/pub/warehouse/`. In addition, we optimize the $F10.7$ input in order to minimize the error by averaging the primary $F10.7$ index into a window of 15 previous days, 30 days (1 month), 2 months, 3 months, 4 months, 8 months and 12 months (1 year). Each case of the averaged $F10.7$ value is trained through the three different network models to find the best performance of $F10.7$ index.

2.3.5. Magnetic activity (A_p index)

The A_p index is an indicator of the geomagnetic activity based on the recorded Earth's geomagnetic activity. It is used to identify the major geomagnetic activity which influences the variations in the ionospheric layer. Stolle et al. (2006) proposes the relation of magnetic field to ESF using the CHAMP (Challenging Minisatellite Payload) satellite. The variations of magnetic field are most suitable for the detection of ESF occurrence. The A_p index is used in the

prediction model of Zhao et al. (2019) to forecast the critical frequency of F2 layer (foF2) in Beijing, China. It is computed by an 8-point running average of successive 3-hour *ap* indices during geomagnetic storm. Dataset of A_p index is derived from data center of NOAA. Therefore, to find the proper value of A_p index to minimize the error, we define the window size of computing the running average A_p value into different window sizes including the previous four, eight, sixteen and thirty two 3-hourly values, respectively.

2.4. The description of the proposed NN model

The designed RSF model is based on the defined input parameters with the overall network structure as shown in Fig. 3. The selection of input spaces is evaluated by the minimum error or the best performance of the forecaster model. The basic functionality of this proposed NN model can be written in a function form of the space parameters that are composed of the diurnal variation (HrS and HrC), seasonal variation (DNS and DNC), geographic latitude (Lat), solar index ($F10.7(4)$) and magnetic index ($A_p(8)$), i.e.,

$$SpreadFModel = f\left(HrS, HrC, DNS, DNC, Lat, F10.7(4), A_p(8)\right), \tag{6}$$

where $F10.7(4)$ is an averaged $F10.7$ value which is taken over the period of the previous four months, $A_p(8)$ is an averaged A_p value of the previous eight 3-hourly values. Therefore, these inputs are normalized and fed to the NN model. The proposed NN model applies a feed-forward network Haykin (2008) and its learning is trained by using the feedback error of the back propagation algorithm of Levenberg-Marchquardt which is referred to be the fast learning algorithm Kisi & Uncuoglu (2005). The responses after training are passed through the threshold to make a decision of RSF presence or RSF absence. Finally, the results are analyzed in terms of the percentage of RSF occurrence, i.e.,

$$\%ESF_t = \frac{\sum_{i=1}^{N} ESF_{t,i}}{N} \times 100, \tag{7}$$

where $\%ESF_t$ is the percentage of spread-F occurrence at time t, N represents the number of observed days and $ESF_{t,i}$ is the ESF occurrences at time t and day i.

2.5. The ESF probability of IRI-2016 model

The ESF probability has been included in the IRI model in 2003 under a regional model of Brazil longitude sector Abdu et al. (2003). The dataset covers 13-years period from 1978-1990 over the equatorial site Fortaleza (3.9° S, 38.45° W, dip angle: -9°) and coordinates of Cachoeira Paulista (22.6° S, 315° W, dip angle: -28°) is utilized to build this model. The model produces the prediction of spread-F occurrence in monthly percentage. Furthermore, the input parameters consist of the local time, latitude, season/month and the solar flux value, these inputs are fitted by the method of a cubic-B spline. The spread-F probability P can be shortly expressed as

$$P(t, m, F, \phi) = \sum_{i=1}^{23}\sum_{j=1}^{12}\sum_{k=1}^{3}\sum_{l=1}^{2} a_{i,j,k,l} N_{i,4}(m) N_{j,2}(m) N_{k,2}(F) N_{l,2}(\phi), \tag{8}$$

where t is the local time, m is the seasonal variations, F represents the solar flux and ϕ is the latitude dependence. In addition, $N_{i,4}(t)$ is a cubic-B spline of order four which is applied to local time dependence, $N_{i,2}(m)$, $N_{k,2}(F)$, $N_{i,2}(\phi)$ are cubic-B splines of order two which are assigned to season, solar flux and latitude, respectively, and $a_{i,j,k,l}$ are the monthly mean of the spread-F occurrence.

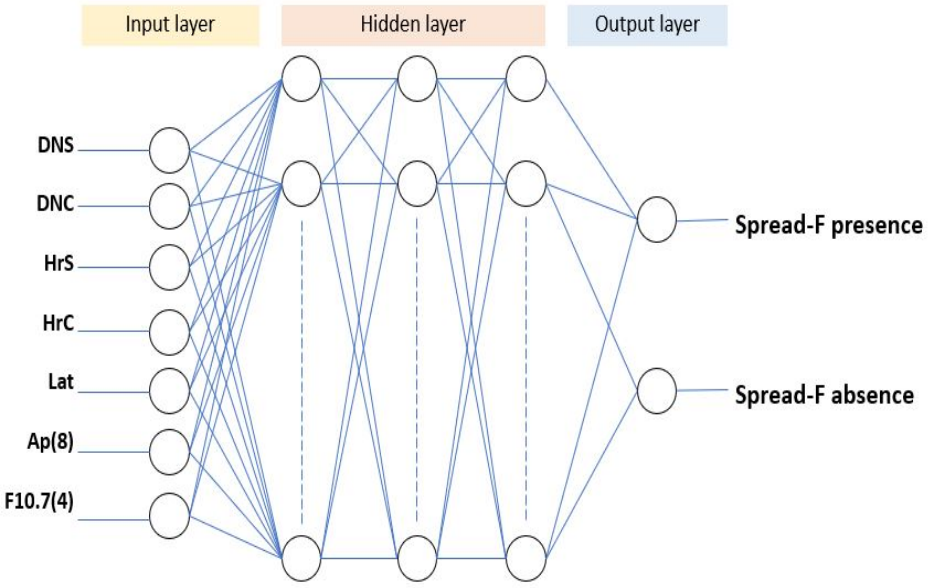

Figure 3: Schematic diagram of the proposed neural network model for spread-F prediction.

3. Results and discussions

3.1. The RMSE in relation to the network structure and input parameters

In this work, we utilize the RSF dataset during 2013 to 2015 for training and designing the network structure with the input parameters as illustrated in Fig. 3. The graphs in Fig. 4(a) exhibit the $RMSE$ of the designed network structure with single, two and three hidden layers, respectively. We obtain the NN structure with three hidden layers with 45 neurons which produces the minimum $RMSE$ at 0.185. In addition, the two increase values of $RMSE$ occur at three hidden layers with 45 and 50 neurons as depicted in Fig.4(a) However, these two rapid increase cases are due to the initial selected number of the neurons sizes but they do not affect the optimal number of neuron size as demonstrated in Fig. 4(a). Fig. 4(b) and (c) show the derived $RMSE$ of the designed NN models with different running mean values of A_p and $F10.7$ indices. The lowest $RMSE$ of the NN model (0.227) occurs when the averaged A_p value of the previous eight 3-hourly value $A_p(8)$ is used as shown in Fig. 4(b). On the other hand, the minimum $RMSE$ (0.206) occurs when the averaged $F10.7$ value of the four previous months $F10.7(4)$ is used as illustrated in Fig. 4(c). As a result, these two inputs are determined as network inputs together with the seasonal variation (DNS and DNC), diurnal variation (HrS and HrC), and the geographic latitude (Lat).

3.2. The performance of proposed NN model after training

Figure 5 shows the confusion matrix of the proposed NN model after training with the dataset from 2013 to 2015. The confusion matrix yields the informative details of how the proposed NN model learns to classify or predict the outputs with respect to the target classes. Based on the design of NN model, there are two target classes and also the outputs consist of two classes with respect to target classes, namely, the class 1 and 2 representing RSF presence and RSF absence, respectively. The proposed NN model is trained with 78,153 data samples which are randomly partitioned into ratio of 70% and 30% for the training. The output of class 1 represents the case of RSF presence; the correct classification of the proposed NN model is 7.6% against 0.5% of wrong classification regarding to all of the

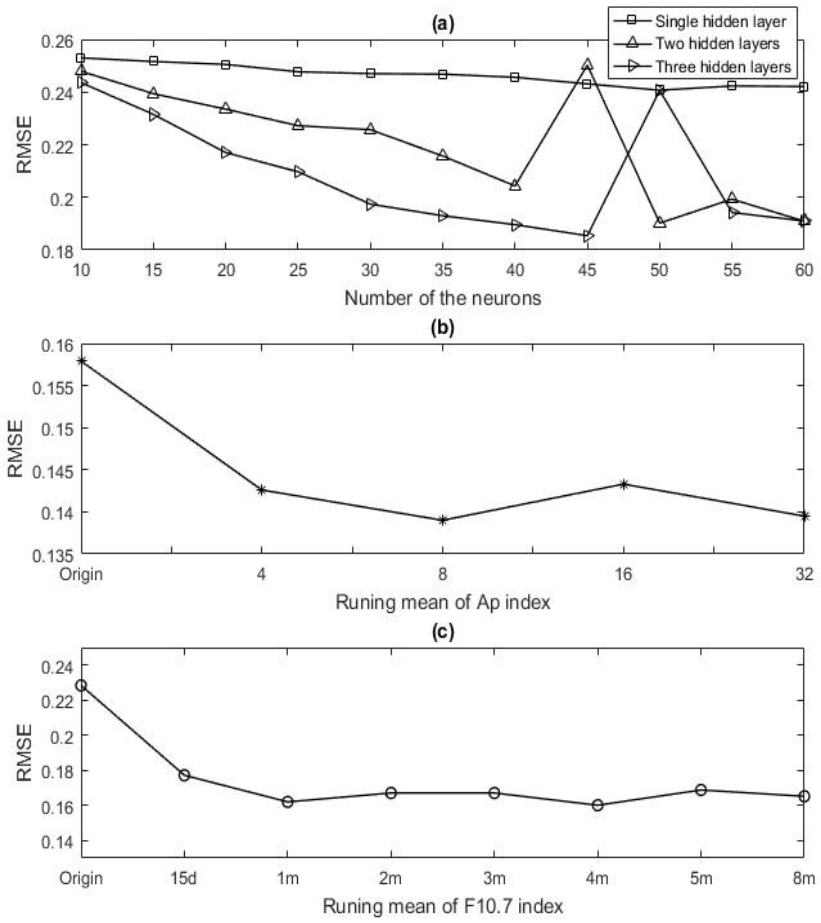

Figure 4: The RMSE performances of the NN models during training as a function of (a) the number of neurons (b) the A_p index and (c) the F10.7 index.

78,153 data samples. In addition, out of the 6,348 samples of prediction of RSF presence, 93.7% is correct and 6.3% is wrong. The case of RSF absence is represented by the output class 2; the 90.7% of the RSF absences is derived as correct prediction and 1.2% is for the incorrect classification. In total, the classification of the 71,805 samples of the RSF absences is 98.7% correct prediction and 1.3% incorrect prediction. On the other hand, the amount of the 6,877 samples of the target class 1(left column), the proposed NN model produces the 86.4% of the correct classification as the RSF occurrences and 13.6% belongs to the RSF absence. For the 71,267 samples in the target class 2(middle column), the proposed NN model predicts 99.4% spread-F absence and 0.6% RSF presence (or wrong classification). Overall, the learning of the proposed NN model achieves the 98.3% correct and 1.7% wrong classifications.

Figure 5: The confusion matrix of the proposed NN model (after training).

3.3. Comparison of the predicted results with the actual observed results in 2016

The comparison between the actual observed RSF and the predicted RSF using the designed NN model in 2016 are analyzed by the percentage of RSF occurrences following Fig. 6. The characteristic of RSF occurrences can be seen clearly in each month, particularly, the percentage of RSF occurrences mostly appears in the equinoctial months more than solstice months Klinngam et al. (2015). In addition, the prediction of proposed NN model produces both the overestimation and underestimation that do not differ much from the actual observed results in each month. The proposed NN model underestimates the RSF occurrences in March, April, June and July by about 4.3%, 10.2%, 10.6% and 4.4% of averaged values, respectively. Note that in November and December, there is no RSF occurrence due to the missing data and its particular characteristic in solstices are shown in Fig. 2, however, the proposed NN model could predict the RSF occurrences which are very close to the actual values. The largest overestimations of the proposed NN model are 0.3% to 2.5% in January, May, August, October, November, and December. Totally, the prediction of

the proposed NN model provides the lower overestimated results than the underestimated results. The percentage of predicted RSF probability from the proposed NN model is under 10.6% of the average deviation when compared with the actual observed RSF probability.

3.4. Evaluation of the IRI-2016 model performance through the comparison

Figure 7 shows the comparison of RSF probability of the IRI-2016 with the observation and the proposed NN model during 2016 over CPN station. As a result, the estimated results of IRI-2016 model has a particular trend in each month. We notice that the IRI model produces the predicted results which are higher than the observed results in all months during 18:00LT to 21:00LT and the IRI-2016 model overestimates the RSF probability in several months including January, February, March, September, October, November and December, by 12.7%, 5.8%, 8%, 6.8% and 6.9% of the averaged values, respectively. Therefore, those results imply that the most predictions of IRI model are higher than the results of the proposed NN model, particularly during the equinoctial months and in some solstice months, except in May, June, July where it has lower estimation than the proposed NN model and the observation is depicted in Fig. 7. The low percentages of RSF occurrence are observed to be about 0.4%, 4%, and 4.6% of the averaged values during June solstices, respectively. The IRI model underestimates the RSF probability which is lower than the actual observation even though the real observed value is very low during 22:00LT to 07:00LT. The low RSF probability of June solstices falls into the same case like December solstices including November and December, hence, this case is possibly due to the missing data and the particular feature of the solstice season. Therefore, depending on accuracy of IRI model is performed, we expect that the causes are from the dataset which uncovers CPN region and in addition, the differences in geographical location is also probably the difficulties of IRI model to forecast the RSF probability over CPN region.

4. Conclusions

In this work, the proposed NN model is designed and tested for spread-F at CPN station, Thailand. It achieves 98.3% accuracy after training with the available data which covers the four-year period from 2013 to 2015 during high solar activity of 24^{th} solar cycle. The proposed NN model provides the best performance with three hidden layers containing 45 neurons in each layer. The A_p and $F10.7$ indices can minimize the errors of the proposed NN model by utilizing the previous averaged values of the eight 3-hourly values of A_p ($A_p(8)$) and the four previous months of the averaged $F10.7$ values ($F10.7(4)$). The results demonstrate that the prediction of proposed NN model could forecast the RSF probability in solstice months more accurately than equinoctial months over CPN station. In addition, the validation of the IRI-2016 model shows clearly that the IRI model still overestimates the RSF probability in the equinoctial months and December solstices in 2016 over CPN station.

Development of the spread-F predictive model is still a challenging work in terms of the data limitation for training the model. In future works, we will increase the amount of the training dataset and add the new input parameters such as the virtual height of F layer ($h'F$) Wang et al. (2018), the velocity drift of F bottom side region, the rate of total electron content ($ROTI$) or gravity wave Manju et al. (2016) and the plasma vertical drift Huang (2018). Moreover, the another learning technique of the machine learning, deep learning and reinforcement learning can be used.

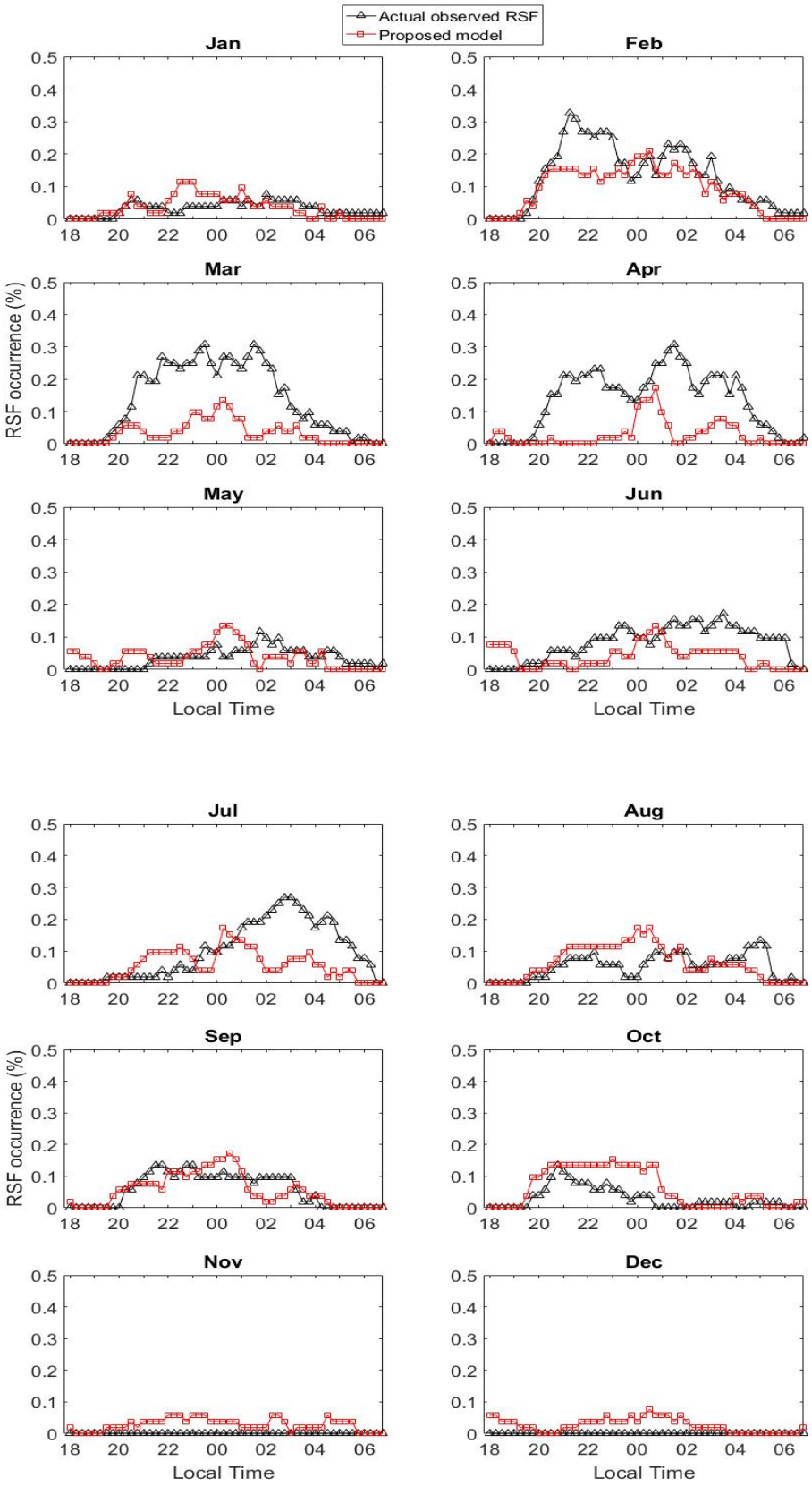

Figure 6: Comparison between of the actual observed spread-F with the predicted result of the proposed spread-F model in each month of 2016 at CPN station.

5. Acknowledgments

This work is supported by King Mongkut's Institute of Technology Ladkrabang under the grant #KDS2019/016. We thank NICT, Japan, and NOAA for supporting the research equipment and the data.

References

Abdu, M. (2001). Outstanding problems in the equatorial ionosphere-thermosphere system relevant to spread-F. *Journal of Atmospheric and Solar-Terrestrial Physics*, *63(9)*, 869–884. URL: https://doi.org/10.1016/S1364-6826(00)00201-7.

Abdu, M., Kherani, E., Batista, I., de Paula, E., Fritts, D., & Sobral, J. (2009). Gravity wave initiation of equatorial spread F/plasma bubble irregularities based on observational data from the spreadfex campaign. *Ann. Geophys*, *27*, 2607–2622. URL: https://doi.org/10.5194/angeo-27-2607-2009.

Abdu, M., Souza, J., Batista, I., & Sobral, J. (2003). Equatorial spread F statistics and empirical representation for IRI: A regional model for the brazilian longitude sector. *Advances in Space Research*, *31*, 703–716. URL: https://doi.org/10.1016/S0273-1177(03)00031-0.

Abdu, M., Souza, J. D., Kherani, E., Batista, I., MacDougall, J., & Sobral, J. (2015). Wave structure and polarization electric field development in the bottomside F layer leading to postsunset equtorial spread F. *J. Geophys. Res. Space Physics*, *120*, 6930–6940. URL: https://doi.org/10.1002/2015JA021235.

Basu, S., Aarons, J., McClure, J., & Cousins, M. (1978). On the coexistence of kilometer and meter-scale irregularities in the nighttime equatorial F region. *Journal of Geophysical Research*, *83*, 4219 – 4226. URL: https://doi.org/10.1029/JA083iA09p04219.

Booker, H., & Wells, H. (1938). Scattering of radio waves in the F-region of the ionosphere. *Journal of Geographical Research*, *43*, 249–256. URL: https://doi.org/10.1029/TE043i003p00249.

Haykin, S. (2008). *Neural Networks and Learning Machines, Third Edition*. PEARSON.

Hoang, T. L., Abdu, M., MacDougall, J., & Batista, I. (2010). Longitudinal differences in the equatorial spread F characteristics between vietnam and brazil. *Advances in Space Research*, *45*, 351 – 360. URL: https://doi.org/10.1016/j.asr.2009.08.019.

Hu, A., & Zhang, K. (2018). Using bidirectional long short-term memory method for the height of f2 peak forecasting from ionosonde measurements in the australian region. *Remote Sens*, *10*, 49–53. URL: https://doi.org/10.3390/rs10101658.

Huang, C. S. (2018). Effects of the postsunset vertical plasma drift on the generation of equatorial spread F. *Progress in Earth and Planetary Science*, *5*, 1–15. URL: https://doi.org/10.1186/s40645-017-0155-4.

Kisi, Ö., & Uncuoglu, E. (2005). Comparison of three back-propagation training algorithms for two case studies. *Indian Journal of Engineering and Materials Sciences*, *12*, 434–442. URL: https://pdfs.semanticscholar.org/0465/dca766776bdb51be0472a6654453673ae114.pdf.

Klinngam, S., Supnithi, P., Rungraengwajiake, S., Tsugawa, T., Ishii, M., & Maruyama, T. (2015). The occurrence of equatorial spread-F at conjugate stations in southeast asia. *Advances in Space Research*, *55*, 2139 – 2147. URL: https://doi.org/10.1016/j.asr.2014.10.003.

Manju, G., Madhav, Haridas, M., & Aswathy, R. (2016). Role of gravity wave seed perturbations in esf day-to-day variability: A quantitative approach. *Advances in Space Research*, *57*, 1021–1028. URL: https://doi.org/10.1016/j.asr.2015.12.019.

McKinnel, L., Paradza, M., Cilliers, P., Abdu, M., & de Souza, J. (2010). Predicting the probability of occurrence of spread-F over brazil using neural networks. *Advances in Space Research*, *46*, 1047–1054. URL: https://doi.org/10.1016/j.asr.2010.06.020.

McKinnell, L., & Oyeyemi, E. (2010). Equatorial predictions from a new neural network based global fof2 model. *Advances in Space Research*, *46*, 1016–1023. URL: https://doi.org/10.1016/j.asr.2010.06.003.

Oyeyemi, E., McKinnell, L., & Poole, A. (2006). Near-real time fof2 predictions using neural networks. *Atmospheric and Solar-Terrestrial Physics*, *68*, 1807–1818. URL: https://doi.org/10.1016/j.jastp.2006.07.002.

Oyeyemi, E., McKinnell, L., & Poole, A. (2007). Neural network-based prediction techniques for global modeling of m(3000)f2 ionospheric parameter. *Advances in Space Research*, *39*, 643–650. URL: https://doi.org/10.1016/j.asr.2006.09.038.

Oyeyemi, E., Poole, W., & Mckinnell, L. (2005). On the global model for fof2 using neural networks. *Radio Science*, *40*, 1–15. URL: https://doi.org/10.1029/2004RS003223.

Pezzopane, M., Zuccheretti, E., de Abreu, A., de Jesus, R., Fagundes, P., & Supnithi, P. (2013). Low-latitude equinoctial spread-F occurrence at different longitude sectors under low solar activity. *Ann. Geophys*, *31*, 153 – 162. URL: https://doi.org/10.5194/angeo-31-153-2013.

Pietrella, M., Pezzopane, M., Fagundes, P., de Jesus, R., Supnithi, P., Klinngam, S., Ezquer, R., & Cabrera, M. (2017). Equinoctial spread-F occurrence at low latitudes in different longitude sectors under moderate and high solar activity. *Atmospheric and Solar-Terrestrial Physics*, *164*, 149–162. URL: https://doi.org/10.1016/j.jastp.2017.07.007.

Rungraengwajiake, S., Supnithi, P., Tsugawa, T., Maruyama, T., & Nagatsuma, T. (2013). The variation of equatorial spread-F occurrences observed by ionosondes at thailand longitude sector. *Advances in Space Research*, *52*, 1809–1819. URL: https://doi.org/10.1016/j.asr.2013.07.041.

Stolle, C., Luhr, H., Rother, M., & Balasis, G. (2006). Magnetic signatures of equatorial spread F as observed by the champ satellite. *Journal of Geophysical Research*, *111*, 1 – 13. URL: https://doi.org/10.1029/2005ja011184.

Tsunoda, R., Saito, S., & Nguyen, T. T. (2018). Post-sunset rise of equatorial F layer — or upwelling growth? *Progress in Earth and Planetary Science*, *5*, 22 – 50. URL: https://doi.org/10.1186/s40645-018-0179-4.

Wang, N., Guo, L., Zhao, Z., Ding, Z., & Lin, L. (2018). Spread-F occurrences and relationships with fof2 and h'F at low and mid-latitudes in china. *Earth, Planets and Space*, *59*, 1–14. URL: https://doi.org/10.1186/s40623-018-0821-9.

Watthanasangmechai, K., Supnithi, P., Lerkvaranyu, S., Tsugawa, T., Nagatsuma, T., & Maruyama, T. (2012). TEC prediction with neural network for equatorial latitude station in thailand. *Earth Planets Space*, *64*, 473–483. URL: https://doi.org/10.5047/eps.2011.05.025.

Zhao, J., Li, X., Liu, Y., Wang, X., & Zhou, C. (2019). Ionospheric fof2 disturbance forecast using neural network improved by a genetic algorithm. *Advances in Space Research*, *63*, 4003–4014. URL: https://doi.org/10.1016/j.asr.2019.02.038.

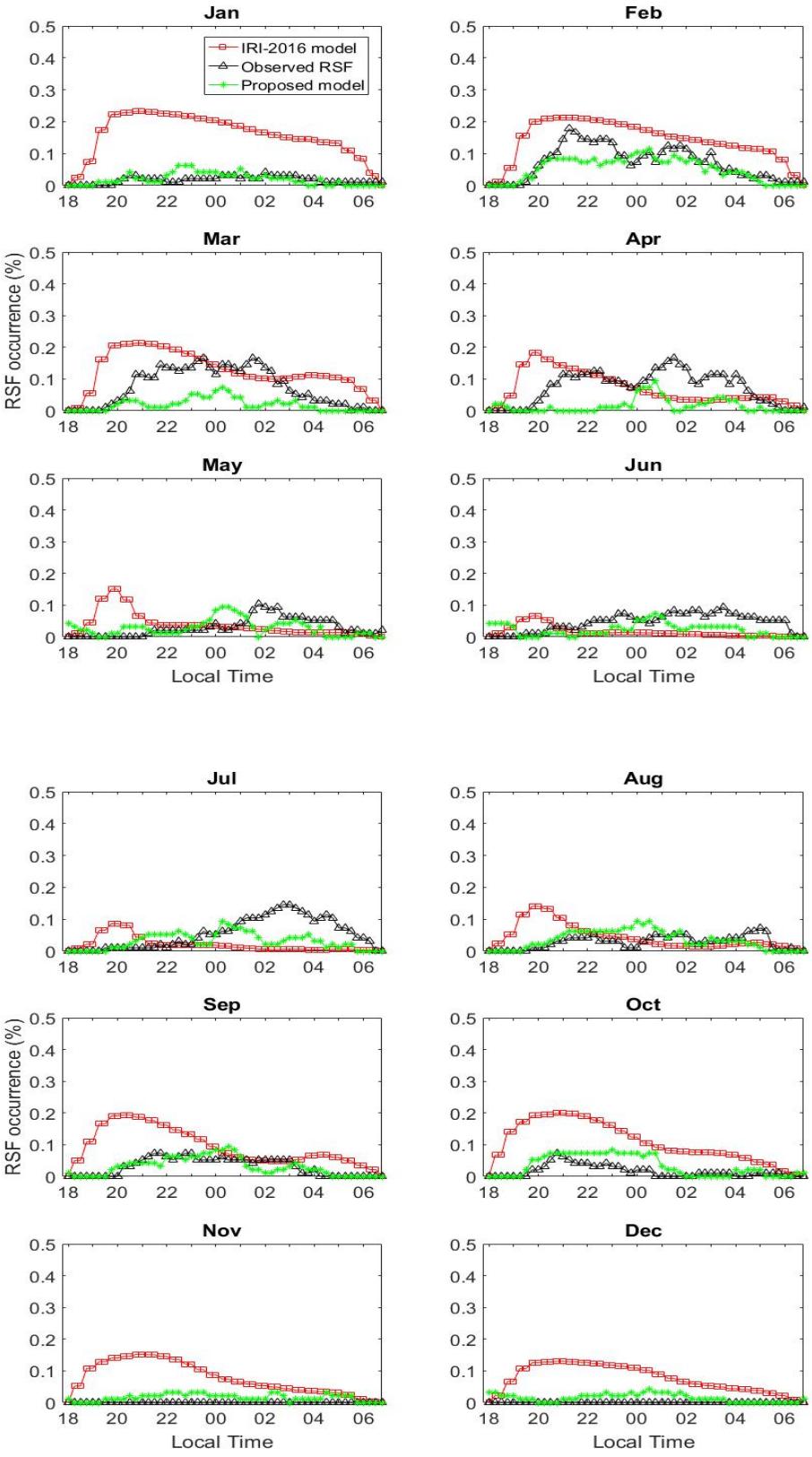

Figure 7: Comparison of the IRI-2016 model with the actual observed results and the results of the proposed NN model in each month of 2016 at CPN station.