# CHAPTER 5

# CONCLUSIONS

## 5.1 INTRODUCTION

This chapter provides an overall summary of the research presented in this thesis; (i) identified the space weather parameters that contribute to TEC modelling, (ii) development of a local TEC estimation model based neural network and evaluates the capability of NN to interpolate and extrapolate the missing TEC data, (iii) investigated the predictability of TEC using NN to estimate seasonal TEC and disturbed days TEC and (iv) development of a new hybrid model to forecast the ionospheric TEC ahead. The major results of the study are presented. Several suggestions that can be further explored to improve this study in future are pointed out at the end of the chapter.

### 5.1.1 Identification on the space weather parameters that contribute to TEC modelling

Solar and magnetic activities are the primary factors that influence the ionospheric TEC variations. In this study, various solar and magnetic indices are examined to obtain the most pertaining proxies for TEC modelling over Parit Raja, Malaysia using a short term data. In terms of solar proxies; sunspot number (SSN), solar radio flux ($F_{10.7}$) and solar extreme ultraviolet flux ($S_{10.7}$) with their means over several period such as daily values, one solar rotation (27 days) and three solar rotations (81 days), are analysed to determine the most influenced parameters. Two types of smoothing schemes; namely centered and backward means are used to smooth the one and three solar rotation periods. In terms of magnetic indices, the three-hourly index (ap) and Disturbance Storm Time (Dst) are considered in this study. The neural network model is used to determine the most appropriate proxies for ionospheric TEC estimation. These proxies are included in the input space of the NN model along with the hour (HR) and day number (DN) parameters represent the diurnal variation and seasonal variation, respectively. Both the HR and DN are considered as the constant input parameters (Cp) in the TEC modelling. In order to determine the optimum input parameters, the combination of proxies which yielded the least RMSE is considered as the most influenced parameters on TEC variability. The outcomes of this study as follows:

i. In terms of solar proxies, the proxies with combination periods produced smaller RMSE than proxies with single period. The combination of daily and 27 day period concurrently gives an improvement in TEC modelling for all the solar proxies. The 27 day means tend to represent as a better index when a short term of dataset (less than 2 years) is used in ionospheric TEC estimation. The combination of different solar proxies improved the TEC modelling further. The best combination is the daily and 27 day backward means of SSN and $S_{10.7}$.

ii. The combination of solar and magnetic proxies along with the constant parameters yielded further improvement in ionospheric TEC modelling. The daily and 27 backward means of $S_{10.7}$ and SSN (solar proxies), ap (magnetic index), HR and DN (constant parameters) are the best combination to enhance the estimation accuracy. The optimum input parameters for TEC estimation model based NN are as below:

$$NN\,TEC \; = \; f(SSN(d+27b), S_{10.7}(d+27b), ap, DNs, DNc\; HRs, HRc)$$

### 5.1.2 Development of TEC estimation model based on neural network technique

In efforts to construct an optimum TEC model using neural network (NN), the number of hidden node(s) in the hidden layer is identified based on the RMSE results. The optimal number of hidden nodes that yielded the least RMSE is summarized as follows:

$$Number\ of\ hidden\ neurons = n + 2$$

where n is the number of input parameters in the NN model. Finally, to optimize the performance of the TEC modelling, the most effective training algorithm in the NN model that can maximize the performance of the NN to estimate the ionospheric TEC is determined. The Levenberg-Marquardt (LM) training algorithm achieved the fastest convergence rate, the least RMSE error value with the correlation coefficient more than 0.9 during the training process selection. With all the aforementioned criteria, a feed forward multi-layer network associated with Levenberg - Marquardt (LM) back-propagation algorithm is implemented to enhance the estimation accuracy. The optimum NN configuration used in this thesis; 9 numbers of nodes in the input layer, 11 nodes in a single hidden layer and 1 node represent the ionospheric TEC in the output layer designated as 9:11:1. The hyperbolic tangent sigmoid function is used as an activation function for all layers except in the input layer.

The developed NN is employed to investigate the NN's interpolation and extrapolation capabilities to estimate the missing GPS TEC. The NN1 and NN2 are developed to estimate the hourly ionospheric TEC. The NN1 model is used for interpolation technique. In this network, the March 2006 dataset is employed within the training set period (February 2005 - April 2006), however the data is set aside for testing purposes. The NN2 model is used for extrapolation technique, where in this network, the March 2006 dataset is outside the training sample (February 2005 - February 2006).

i. In general, the model's interpolation capability could be seen more evidently than the extrapolation, especially over longer periods of missing data. The NN2 model experienced more difficulty in extrapolating the TEC values outside the input space during the night time than the daytime. Overall, in the extrapolation analysis, there is a degradation of performance in NN2 model with increasing of missing value rates. The NN2 model has Crel below than 85% when the missing data are above 60%.

### 5.1.3 Investigation on the predictability of TEC using NN to estimate seasonal TEC and disturbed days TEC

In order to the access the predictability of TEC using NN more extensively, the NN2 model is further used to estimate the seasonal and disturbed days TEC over Parit Raja station.. To validate the performance of the developed model, the NN TEC values are compared with the global model IRI-2007 TEC values with respect to the corresponding GPS TEC. The overall results show that the developed NN model has high competence in estimating the TEC variability at this region. Following are the findings of this study:

i. The extrapolation capability of the NN2 model is further assessed by estimating the unseen TEC values for all seasons in 2006; namely March equinox, June solstice, October equinox and December solstice. Generally, the NN2 model tends to estimate the TEC values fairly well than the IRI-2007 model during both the equinoctial and solstitial months. The average RMSEs values of the NN and IRI models for all the four seasons are equal to 1.666 and 2.772 TECU, respectively. The NN model gives an improvement by ~39.9% over the IRI-2007 model. The performance of the NN model is more significant during solstice seasons. The NN model gives the best TEC approximation during December solstice with the smallest normalized RMSE of ~0.085 and largest Crel of ~90% while the model fails to estimate the TEC values accurately during October equinox with normalized RMSE of ~0.111 and Crel of ~82%.

Finally, the possibility of the NN2 model to estimate the unseen TEC values during the impulsive events is examined. The results show that the NN model fails to capture the TEC variability during negative ionospheric storm. The IRI-2007 model generally tends to provide more accurate estimation results than the NN model during this period. The Crel of the IRI model is about ~25% higher than NN model. In contrast, the NN model able to generalize the TEC trend more favourably than the IRI model during positive ionospheric storm effects. The Crel of the NN model is about ~30 to 35% higher than the global model during this period. Compare to negative storms, more positive storms experiences in the learning phase helps the NN model to generalize unseen data well during positive ionospheric storm.

### 5.1.4 Development of TEC forecasting model based on hybrid SARIMA-NN model

In this section, the study explored and developed a hybrid model which combines the linear and non-linear methods to forecast the GPS TEC values 3 days or 72 hours ahead. A seasonal autoregressive integrated moving average integrated (SARIMA) is used to model the linear component in the GPS TEC while the residuals obtained from the SARIMA model which assumed to be purely non-linear are modelled using neural network (NN). The ionospheric TEC data over a period of 20 months (February 2005 to September 2006) are used for model development to obtain the optimal model while three months TEC data (October 2006 to December 2006) are used for the model verification, forecasting and comparison purposes.

Using the MATLAB the best fitted SARIMA $(3,1,1)(0,2,2)_{24}$ model is attained to forecast linear components of the TEC time series. Then, a three layer feed forward network associated with Levenberg - Marquardt (LM) back-propagation algorithm is used to forecast the residuals (non-linear component) ahead. Nine consecutive lagged inputs as shown below are considered as the best combination to improve the forecasting accuracy and provide the optimum result:

$$NN\ \ residual(e_{t+71}) = f(e_{t-1}, e_{t-2}, e_{t-3},\ \dots, e_{t-8}, e_{t-9})$$

where $f$ is the non-linear function determined by the neural network structure and $e_t$ is the hourly lagged residuals. The optimal architecture used to forecast the non-linear component is 9:7:1, where nine input nodes, one output node depicted the forecast residuals values and seven hidden nodes. Finally, the forecast ionospheric TEC values are generated by integrating the forecast values yielded by both the models; SARIMA and NN, respectively. For the validation and comparison purposes, the forecast values of developed hybrid SARIMA-NN model are compared with the forecast values produced by the individual models, SARIMA and NN separately against the corresponding GPS TEC. The overall results show that the combined method is able to forecast the GPS TEC values three days ahead more favourably than either of the individual models, FCAST-SARIMA and FCAST-NN used separately. The outcomes of this study are as follows:

i. The hybrid SARIMA-NN, FCAST-SARIMA and FCAST-NN models tend to forecast the GPS TEC fairly well during the magnetic quiet condition (10 - 12 October 2006) since there is no drastic fluctuation in the TEC variations during this period. Among the forecast models, the hybrid model able to forecast the GPS TEC 72 hours ahead better than either of the individual models used separately. In terms of average RMSE, the combined model yielded improvements by ~11.7% and ~24.3% over both the single models SARIMA and NN, respectively.

ii. The applicability of the models is further assessed by forecasting the ionospheric TEC during the moderate condition from 9 - 11 November 2006. Generally the TEC values produced by the hybrid model reasonably match the GPS TEC than the individual models. The performance of the individual models degraded and the forecasting errors increased for both the single models as the time horizon became larger.

iii. The effectiveness of the forecast models are also investigated during the disturbed conditions which is from 14 - 16 December 2006. The hybrid SARIMA-NN and FCAST-SARIMA models tend to forecast the GPS TEC fairly well on the storm day. In contrast, the FCAST-NN able to forecast the GPS TEC accurately on the pre- and post-storm days. During disturbed condition, the percentage improvements of the hybrid model over FCAST-SARIMA and FCAST-NN in term of average RMSE are ~13.4% and ~26.1%, respectively. This shows the hybrid model tend to forecast the TEC dynamics ahead more accurately during the disturbed conditions than the other individual models. However, the overall Crel results indicate that the performance of all the three forecast models deteriorated during disturbed condition compared during the quiet and moderate conditions.

## 5.2 FUTURE WORKS

In this study, a data driven NN model is developed to estimate the GPS TEC and a hybrid SARIMA-NN model is designed to forecast the GPS TEC ahead over the Parit Raja station. Both the methods seem to be adequate and able to capture the TEC variations correctly. This may be due to the study period (2005 - 2006), which only covered the medium solar activity in the training and testing phases for both the developed models. However, it must be acknowledged that the usage of short duration data in the models may limit the validation of the models during low and high solar activity since the NNs in the developed models are unable to extrapolate outside the input space for a longer period of time. Therefore, to develop a truly representative ionospheric TEC estimation and forecasting models over this station in future, a longer data period at least for one solar cycle (~11 years) is required.

Furthermore, in this thesis both the developed models were restricted for a single station. Since fewer works have been done concerning on the TEC estimation and forecasting in the Malaysia region, future work should include more stations within Malaysia to develop the comprehensive representative TEC model that is able to provide insight into the TEC variations at any point within Malaysia.

In addition, the prominent input-output pairing would ensure the success of the designing process in NN. The inability of the developed NN model to extrapolate outside the input space especially during night time may attribute to the lack of input parameter(s) to represent the TEC variations during these hours. The existing input parameters; the solar and magnetic proxies were unable to determine the TEC variations during the recombination process in the night time. The number of electron contents during night hours is not associated with the ionisation due to the solar radiations. It may relate to other sources which highly influence the night time of the TEC variations at this region. Besides, the developed model did not yield accurate results during equinox months. The TEC values varied significantly during the equinox months and there are a few mechanisms that affect the enhancements or depletions of electrons during equinoxes in this region. Those mechanism parameters were not considered in the NN modelling at this level. Therefore, in future, the inclusion of other significant geophysical parameters in the NN modelling which determine the TEC variations is important to enhance the predictability on the developed NN model.

In the hybrid SARIMA-NN model, other than the longer period dataset and more stations, the inclusion of other input parameters in the hybrid NN model should be a key priority in future to reduce the forecasting errors, especially during the disturbed conditions. In this regard, other input parameters such as the first, second and relative differences of the residual values should be considered in the hybrid NN forecast model. Besides, the inclusion of solar and magnetic indices in the hybrid NN forecast model is also another key point to investigate in the future.

Lastly, taken into account the potential space weather threats such as the ionospheric and geomagnetic storms as well as the natural hazards e.g. pre-seismic and pre tsunami, in future this forecast work can be expanded further to detect the ionospheric anomalies in advance to reduce the impacts on the technical and ground based infrastructures which rely on space-based communication and technologies.