ESTIMATION AND FORECASTING OF IONOSPHERIC TOTAL ELECTRON CONTENT BASED ON NEURAL NETWORK AND HYBRID SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE–NEURAL NETWORK MODELS OVER EQUATORIAL REGION

VIKNESWARY JAYAPAL

UNIVERSITI SAINS ISLAM MALAYSIA

ESTIMATION AND FORECASTING OF IONOSPHERIC TOTAL ELECTRON CONTENT BASED ON NEURAL NETWORK AND HYBRID SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE–NEURAL NETWORK MODELS OVER EQUATORIAL REGION

Vikneswary Jayapal
(Matric No. 4130261)

Thesis submitted in fulfillment for the degree of
DOCTOR OF PHILOSOPHY
SCIENCE AND TECHNOLOGY

Faculty of Science and Technology
UNIVERSITI SAINS ISLAM MALAYSIA
Nilai

July 2016

DECLARATION OF THESIS AND COPYRIGHT
إقرار البحث العلمي وحق النشر

Student's Full Name / اسم الطالب بالكامل	VIKNESWARY JAYAPAL		
Academic Session / الفصل الدراسي	2013/2014	Matric No. / رقم التسجيل الجامعي	4130261
Research Title / عنوان البحث	ESTIMATION AND FORECASTING OF IONOSPHERIC TOTAL ELECTRON CONTENT BASED ON NEURAL NETWORK AND HYBRID SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE - NEURAL NETWORK MODELS OVER EQUATORIAL REGION		

I hereby declare that the work in this thesis/ project paper is my own except for quotations and summaries which have been duly acknowledged /

إني أقر وأعترف بأن هذا البحث من عملي وجهدي الشخصي، أما المقتطفات والاقتباسات، فقد أشرت إلى مصادرها في هامش البحث

I acknowledged that Universiti Sains Islam Malaysia reserves the right as follows /

أقر بأن تحتفظ جامعة العلوم الإسلامية الماليزية بالحقوق التالية:

1. The thesis/ postgraduate project paper is the property of Universiti Sains Islam Malaysia /
إن ملكية هذا البحث العلمي لجامعة العلوم الإسلامية الماليزية

2. The library of Universiti Sains Islam Malaysia has the right to publish my thesis/ postgraduate project paper as online open access (fulltext) and make copies for the purpose of research or teaching and learning only /
تحتفظ مكتبة جامعة العلوم الإسلامية الماليزية بحق النشر لهذا البحث العلمي أو المشروع العلمي على الإنترنت (البحث الكامل) وحق الطباعة بهدف إجراء البحث العلمي أو التعليم والتعلم فقط

.....	.....
(Student's Signature)	(Supervisor's Signature)
820523-08-5060	.....
(IC No./Passport No.)	(Name)
.....	.....
(Date)	(Date)

BIODATA OF AUTHOR

Vikneswary Jayapal (4130261) was born in Perak, Malaysia on 23rd May 1982. She is currently residing at Batu Gajah, Perak.

Vikneswary received her B. Eng. (Hons) degree in Electrical (Telecommunications) Engineering from Universiti Tun Hussein Onn Malaysia, in 2006. Upon completion her first degree, she was attached with Agilent Technologies, Penang Malaysia, in 2006 as a RF engineer. In 2007, she joined the Department of Communication Engineering, Universiti Tun Hussein Onn Malaysia as a tutor. She then took a leave of absence to pursue her studies and obtained her M.Eng degree in Electrical, Electronics and Telecommunications from Universiti Teknologi Malaysia, in 2009. Her master's project was on security in Wireless Sensor Network using IPv6.

At present, she is a Ph.D. candidate at Faculty of Science and Technology, Universiti Sains Islam Malaysia. Her current research interests are in modelling the Earth's ionosphere, radio propagation in the equatorial ionosphere, linear and non-linear estimation models.

ACKNOWLEDGEMENTS

First and foremost I would like to thank God, who has blessed me towards the completion of my research.

I would like to thank my supervisor Prof. Ir. Dr. Ahmad Faizal Bin Mohd Zain for his invaluable role and guidance in this research work. I'm very much grateful for his crucial ideas and valuable resources that he gives me throughout my studies. I would like to take this opportunity to acknowledge the following organization for providing a quality of data which have been used to produce this thesis; WARAS, Universiti Tun Hussein Onn Malaysia for the GPS data, WDC, Japan for Ap and Dst indices, NOAA, USA and CELIAS/SEM spectrometer aboard SOHO for the solar activity indices.

My greatest gratitude to my family especially to my parents Mr and Mrs. Jayapal Jayalachimy , beloved husband Shamuga Dass, dearest siblings Amutha and Thanaalan and never forgotten my family in law for their love, support and understanding that kept me going all the way through my PhD's study. I'm highly indebted to them.

Lastly, thanks to all my friends and people who have helped and supported me in different ways. Special thanks to Nurul Hafizah, Roshayati, Shamala, Siva Ranjini, and Shameela for lent me their precious time and support me from time to time. Their contributions are greatly appreciated.

ABSTRAK

Kebolehubahan jumlah kandungan elektron (TEC) yang tidak dapat diramal di kawasan khatulistiwa dan kekosongan dalam pengkalan data TEC yang disebabkan oleh kegagalan infrastruktur Bumi mewujudkan keperluan untuk membangunkan satu teknik penganggaran TEC. Pendekatan berasaskan NN didapati sangat sesuai di dalam pemodelan parameter ionosfera kerana keupayaannya yang fleksibel memeta fungsi tak linear, yang boleh menganggarkan sebarang fungsi boleh ukur tak linear dengan ketepatan yang dikehendaki. Kajian ini mengemukakan pembangunan model rangkaian neural (NN) untuk menganggar TEC ionosfera bagi sebuah stesen penerima GPS tunggal (Lat. 1˚52'N, Long. 103˚48'E, Junaman magnetik 14.3˚) di Malaysia dari Februari 2005 hingga Desember 2006. Di dalam NN, kebolehubahan TEC dimodelkan sebagai fungsi untuk variasi harian, variasi bermusim, solar, dan proksi magnet. Model NN1 dan NN2 masing-masing digunakan untuk menginterpolasi (tentu dalam) dan mengekstrapolasi (tentu luar) data TEC setiap jam yang hilang. Keupayaan NN menginterpolasi boleh dilihat dengan lebih jelas berbanding dengan mengekstrapolasi, terutamanya bagi tempoh yang lebih panjang untuk data yang hilang. Model NN2 mengalami lebih banyak kesukaran dalam mengekstrapolasi (tentu luar) nilai TEC pada waktu malam berbanding waktu siang. Model NN2 mempunyai pembetulan relatif (Crel) kurang daripada 85% apabila data TEC yang hilang adalah lebih daripada 60%. Bagi pengesahsahihan model, nilai TEC bagi NN2 dibandingkan dengan model Ionosfera Rujukan Antarabangsa (IRI) berdasarkan kepada GPS TEC. Keputusan anggaran bagi empat musim pada tahun 2006 menunjukkan bahawa, model NN2 sepadan dengan GPS TEC ketika musim solstis. Dari segi purata RMSE, model NN2 menunjukkan peningkatan sebanyak 39.9% berbanding dengan model IRI bagi empat musim tersebut. Kebolehramalan TEC semasa peristiwa spontan menunjukkan bahawa model IRI-2007 cenderung untuk memberikan hasil anggaran yang lebih tepat berbanding dengan model NN2 semasa ribut ionosfera negatif dengan Crel kira-kira ~25% lebih tinggi daripada model NN2. Sebaliknya, model NN2 mampu menjanakan trend TEC yang lebih baik berbanding dengan model IRI semasa kesan ribut ionosfera positif dengan Crel kira-kira ~30 hingga 35% lebih tinggi daripada model IRI. Selain penganggaran, model ramalan TEC ionosfera juga sangat bermanfaat sebagai sistem amaran awal bagi mengurangkan kesan cuaca buruk di angkasa dan bencana alam terhadap kehidupan manusia dan teknologi. Berdasarkan data pembaikan GPS TEC, model ramalan siri masa dibangunkan dengan menggunakan model hybrid yang mengintegrasikan purata gerakan bersepadu autoregresif bermusim (SARIMA) dan rangkaian neural untuk meramal nilai TEC sehingga tiga hari lebih awal. Nilai ramalan TEC daripada model hibrid dibandingkan dengan model individu SARIMA (FCAST-SARIMA) dan NN (FCAST-NN) secara berasingan berdasarkan kepada GPS TEC. Keputusan menunjukkan bahawa nilai TEC bagi ketiga-tiga model ramalan agak baik semasa keadaan tenang. Sementara itu, pencapaian model individu turun semasa keadaan sederhana, dan ralat ramalan meningkat pada kedua-dua model tunggal apabila ufuk masa menjadi lebih besar. Selain itu, dari segi purata RMSE, peningkatan peratusan model hibrid terhadap SARIMA dan NN semasa keadaan terganggu ialah masing-masing mencatatkan ~13.4% dan ~26.1%. Model penganggaran (*estimation*) dan peramalan (*forecasting*) digunakan secara bersama-sama untuk menyediakan satu kajian yang menyeluruh mengenai pemodelan ionosfera TEC di Parit Raja, Malaysia.

ABSTRACT

Unpredictable variability of total electron content (TEC) in the equatorial region and gaps in the TEC database due to Earth infrastructure failures creates a need to develop a TEC estimation model. NN-based approaches are found promising in modelling the ionospheric parameters because they have flexible non-linear function mapping capability, which can estimate any non-linear measurable function with arbitrarily desired accuracy. This work presents, the development of neural network (NN) based model to estimate the ionospheric TEC for a single GPS receiver station (Lat. 1˚52'N, Long. 103˚48'E, Magnetic dip 14.3˚) over Malaysia from February 2005 to December 2006. In NN, the TEC variability is modelled as a function of diurnal variation, seasonal variation, solar and magnetic proxies. The NN1 and NN2 models are used to interpolate and extrapolate the missing hourly TEC data, respectively. The NN's interpolation capability could be seen more evidently than the extrapolation, especially over longer periods of missing data. The NN2 model experienced difficulty in extrapolating the TEC values during the night time than the daytime. The NN2 has relative correction (Crel) below than 85% when the missing TEC data are above 60%. For model validation, TEC values from NN2 are compared with the International Reference Ionosphere (IRI) model with respect to GPS TEC. The estimation results for the four seasons in 2006 show that, the NN2 model agrees well with GPS TEC during solstice seasons. In terms of average root mean square error (RMSE), NN2 model shows an improvement about 39.9% compared to the IRI model over the four seasons. The predictability of TEC during negative ionospheric storm revealed that the IRI-2007 model tends to yield more accurate estimation results than the NN2 model with Crel is about ~25% higher than NN2 model. Conversely, the NN2 model able to generalize the TEC trend more favourably than the IRI model during positive ionospheric storm effects with Crel is about ~30 to 35% higher than the IRI model. Besides estimation, an ionospheric TEC forecasting model can be highly beneficial as an early warning system in order to lessen the adverse space weather and natural hazard impacts on human life and technologies. Based on the recovery GPS TEC data, a time series forecasting model is developed using a hybrid model that integrates the seasonal autoregressive integrated moving average (SARIMA) and neural networks to forecast the TEC values up to three days ahead. The forecast TEC values from the hybrid model are compared with the individual models SARIMA (FCAST-SARIMA) and NN (FCAST-NN) separately with respect to GPS TEC. Results show that all the three models forecast TEC values fairly well during quiet condition. Meanwhile, the performance of the individual models degraded during moderate condition and the forecasting errors increased for both the single models as the time horizon became larger. Besides, in term of average RMSE the percentage improvements of the hybrid model over SARIMA and NN during disturbed condition are ~13.4% and ~26.1%, respectively. The estimation and forecasting models are used in tandem to provide a complete investigation on ionospheric TEC modelling at Parit Raja, Malaysia.

ملخص الأطروحة

نموذج التنبؤ بالمحتوى الكامل للألكترونات في غلاف الأيونوسفير يمكن أن يكون ذا جدوى عالية, حيث يستخدم لأنظمة التحذير المبكر والذي يساعد في التقليل من تأثير الفضاء العكسي على حياة الإنسان والتقنية. عدم صحة البيانات و فقدها يمكن أن يوثر سلباً على أداء نموذج التنبؤ. هذه الأطروحة تتحقق من قدرة تقنية الشبكات العصبية في التنبؤ بالمحتوى الكامل للألكترونات لمحطة استقبال اشارة نظام تحديد المواقع GPS والتي تقع على خط عرض ٥٢°١ شرقا وبخط طول ٤٨°١٠٣ غربا وثنائي قطبية مغناطيسية ١٤,٣° لتوليد بيانات ملائمة المحتوى الكامل للألكترونات في ماليزيا. في تقنية الشبكات العصبية, تغير المحتوى الكامل للألكترونات تمت نمذجته كدالة في التغير اليومي والتغير الفصلي و أيضا كدالة في المفوض الشمسي و المغناطيسي. ولزيادة التحقق, الشبكة العصبية لقيم المحتوى الكامل للألكترونات قد تمت مقارنتها مع المحتوى الكامل للألكترونات لأشارة نظام تحديد المواقع وكذلك مع النموذج العالمي IRI TEC. نموذج الشبكات العصبية يستخدم لأستيفاء و استقراء البيانات الضائعة للمحتوى الكامل للالكترونات للنظام العالمي لتحديد المواقع. قدرة استيفاء نوذج الشبكات العصبية يمكن رؤيته بوضوح بالمقارنة مع قدرة الإستقراء خصوصاً للفترات الطويلة من البيانات الضائعة.

نموذج الشبكات العصبية يواجه صعوبة اكبر في استقراء قيم المحتوى الكامل للألكترونات خلال الأوقات الليلية أكثر من أي وقت أخر, حيث لديه دقة أقل من ٨٥٪ عندما تكون قين معدل ضياع البيانات اكبر من ٦٠٪. علاوة على ذلك, نتائج التنبؤ لفصول السنة الأربعة لسنة ٢٠٠٦ تبين أن نموذج الشبكات العصبية يميل لتنبؤ قيم المحتوى الكامل للألكترونات بشكل أكثر إعتدالاً مقارنة مع النموذج العالمي IRI-2007. نموذج الشبكات العصبية متلائم جداً مع المجتوى الكامل للألكترونات الخاص بنظام تحديد المواقع العالمي خلال فصول انقلاب الشمس أكثر من ما هو خلال فصول تساوي الليل مع النهار. من ناحية متوسط RMSE نموذج الخلايا البصرية تبين تحسن بمقدار ٣٩,٩٪ مقارنة بالنموذج العالمي IRI-2007. امكانية تطبيق نموذج الشبكات البصرية للتنبؤ بمجتوى الألكترونات الكامل الغير مرئي لغلاف الأيونوسفير خلال الأحداث المتهورة قد تمت دراستها. نموذج RII-2007 في العموم يميل لأعطاء تنبؤات للنتائج بأكثر دقة مقارنة بنموذج الشبكات العصبية خلال العواصف السالبة للأيونوسفير مع دقة تساوي تقريباً ٢٥٪ اكثر بالمقارنة بنموذج الشبكات البصرية. لكن نموذج الشبكات البصرية قادر على استقراء اتجاه المحتوى الكامل للألكترونات بشكل أفضل من نموذج IRI خلال ظواهر عواصف الايونوسفير الموجبة أعلى بحوالي ٣٠ الى ٣٥٪. هذا العمل قد تم توسعته بإستخدام استرداد المحتوى الكامل للإلكترونات لأنظمة تحديد المواقع, حيث تم تطوير نموذج التنبؤ بإستخدام نموذج مهجن والذي يجمع كلاً من: الشبكات العصبية و معدل الإنحدار المتحرك المتكامل الفصلي (SARIMA). النموذج يستخدم لتوليد المحتوى الكامل للإلكترونات لطبقة الأيونوسفير لمدة تصل الى ثلاثة أيام مستقبلية و مقارنتها مع القيم الفعلية وكذلك قيم التنبؤ لنماذج SARIMA و نموذج الشبكات البصرية بشكل منفصل. النتائج توضح أن كلا النموذجين تعطي أداء يمكن مقارنته مع النموذج الهجين خلال وضع السكون. في الوقت ذاته, أداء كلا النموذجين يتضعضع خلال الوضع المعتدل و كذلك أخطاء التنبؤ تزداد لكلا النموذجين كلما زاد وقت الأفق. إلى جانب ذلك, خلال الحالة المضطربة, تحسن النسبة المئوية للنموذج الهجين مقارنة بكل من نموذجي FCAST-SARIMA و NN-FCAST من ناحية متوسط RMSE والتي تساوي ما بين ١٣,٤٪ و ٢٦,١٪. نماذج التنبؤ تستخدم في التتابع لتزويد تحقيق كامل على المحتوى الكامل للإلكترونات لطبقة الأيونوسفير والتي تمت نمذجتها في منطقة باريت راجا الواقعه في ماليزيا.

TABLE OF CONTENTS

CHAPTER 2 LITERATURE REVIEW

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

B	:	backward shift operator
p	:	order of non-seasonal autoregressive
q	:	order of non-seasonal moving average
d	:	order of differencing non-seasonal component
P		order of seasonal autoregressive
Q		Order of seasonal moving average
D	:	order of differencing seasonal component
s	:	length of seasonal period
Φ	:	coefficient of seasonal autoregressive
Θ	:	coefficient of seasonal moving average
ϕ	:	coefficient of non-seasonal autoregressive
θ	:	coefficient of non-seasonal moving average
a_t	:	random errors
n_ϕ	:	phase refractive index
n_g	:	group refractive index
v_ϕ	:	phase velocity
v_g	:	group velocity
$\frac{dn_\phi}{df}$	:	change of phase refractive index with respect to the frequency.
ω_N	:	angular plasma frequency
ω_B	:	electron gyrofrequency

ω	:	angular frequency of the propagating wave,
ω_L	:	longitudinal component of the electron gyrofrequency,
ω_T	:	transverse component of the electron gyrofrequency,
v	:	angular collision frequency between electrons and heavier particles
θ	:	angle between the geomagnetic field vector and the propagation wave direction.
N	:	electron density
e	:	electron charge
ε_0	:	electric permittivity of free space
ΔI_{P_1}	:	pseudo-range ionospheric induced errors (m) at L1 frequency
ΔI_{P_2}	:	pseudo-range ionospheric induced error (m) at L2 frequency
ΔI_{Φ_1}	:	carrier ionospheric induced errors (m) at L1 frequency
ΔI_{Φ_2}	:	carrier ionospheric induced errors (m) at L2 frequency
λ_1	:	wavelength of the signal (m) at L1 frequency
λ_2	:	wavelength of the signal (m) at L2 frequency
N_1	:	carrier phase integer ambiguities between satellite and receiver (cycle) at L1 frequency
N_2	:	carrier phase integer ambiguities between satellite and receiver (cycle) at L2 frequency
bS_{P_1}	:	pseudo-range satellite hardware delays (m) at L1 frequency
bS_{P_2}	:	pseudo-range satellite hardware delays (m) at L2 frequency
bS_{Φ_1}	:	carrier phase satellite hardware delays (m) at L1 frequency
bS_{Φ_2}	:	carrier phase satellite hardware delays (m) at L2 frequency

bR_{P_1} : pseudo-range GPS receiver hardware delays (m) at L1 frequency

bR_{P_2} : pseudo-range GPS receiver hardware delays (m) at L2 frequency

bR_{Φ_1} : carrier phase GPS receiver hardware delays (m) at L1 frequency

bR_{Φ_2} : carrier phase GPS receiver hardware delays (m) at L2 frequency

$\Delta_{multipath,P_1}$, $\Delta_{multipath,P_2}$: pseudo-range multipath effects (m),

$\Delta_{multipath,\Phi_1}$, $\Delta_{multipath,\Phi_2}$: carrier phase multipath effects (m),

$\varepsilon(P_1), \varepsilon(P_2)$: pseudo-range measurement noises (m),

$\varepsilon(\Phi_1), \varepsilon(\Phi_2)$: carrier phase measurement noises (m).

LIST OF APPENDICES

ABBREVIATION

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
ap	Equivalent three hourly planetary amplitude
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BFG	Boyden, Fletcher, Goldfarb and Shanno update
CGB	Powell – Beale restarts conjugate gradient
CGF	Fletcher – Reeves Update conjugate gradient
CGP	Polak-Ribiere Update conjugate gradient
CME	Coronal Mass Ejections
Crel	Relative correction
DN	Day Number
Dst	Disturbance Storm Time
Eabs	Absolute error
EIA	Equatorial Ionization Anomaly
Erel	Relative error
EUV	Extreme Ultra-violet
$F_{10.7}$	Solar Radio Flux
FCAST-NN	Forecast NN
FCAST-SARIMA	Forecast SARIMA
GAGAN	GPS Aided GEO Augmented Navigation
GDA	Batch gradient descent with variable learning rate
GISTM	GPS Ionospheric Scintillation and TEC Monitor
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
HF	High Frequency
HR	Hour
HYBRID SARIMA-NN	Hybrid Seasonal Autoregressive Integrated Moving Average-Neural Network
IGS	International GNSS Service
IPP	Ionospheric Pierce Point
IRI	International Reference Ionosphere
ISIS	International Satellites for Ionospheric Studies
JUPEM	Jabatan Ukur dan Pemetaan Malaysia
LM	Levenberg-Marquardt algorithm
LT/UT	Local or Universal Time
MA	Moving Average
MASS	Malaysia Active GPS System
MLE	Maximum Likehood Estimation
MSE	Mean Square Error

M-SLM	Modify Single Layer Model
MyRTKnet	Malaysia RTK Network
NGDC	National Geophysical Data Centre
NN	Neural Network
PACF	Partial Autocorrelation Function
PPP	Precise Point Positioning
PPS	Precise Positioning Service
PRE	Pre-Reversal Enhancement
RACF	Residual Autocorrelation Function
RINEX	Receiver Independent Exchange Format
RMSE	Root Mean Square Error
ROT	Rate of TEC
RP	Resilient back-propagation algorithm
RTK	Real Time Kinematic
$S_{10.7}$	Solar extreme ultraviolet (EUV) flux
SA	Selectivity Availability
SARIMA	Seasonal Auto Regressive Integrated Moving Average
SC	Solar Cycle
SCG	Scaled conjugate gradient
SEM	Solar EUV Monitor
SLM	Single Layer Model
SOHO	Solar And Heliospheric Observatory
SPS	Standard Positioning Service
SSE	Sum of Squared Errors
SSN	Sunspot Number
sTEC	Slant TEC
TEC	Total Electron Content
TECU	Total Electron Content Unit
URSI	International Union on Radio Science
USDoD	United States Department of Defense
VHF	Very High Frequency
VRS	Virtual Reference Station
vTEC	Vertical TEC
WARAS	Wireless and Radio Science Centre
WDC	World Data Centre