



**DEVELOPMENT OF DATA INTELLIGENT MODELS FOR
ELECTRICITY DEMAND FORECASTING: CASE STUDIES IN
THE STATE OF QUEENSLAND, AUSTRALIA**

A Thesis submitted by

Mohanad Shakir Khalid AL-Musaylh

BSc and MSc (Mathematics)

For the award of

Doctor of Philosophy

2020

Abstract

Electricity demand (G) forecasting is a sustainability management and evaluation task for all energy industries, required to implement effective energy security measures and determine forward planning processes in electricity production and management of consumer demands. Predictive models for G forecasting are utilized as scientific stratagems for such decision-making. The information generated from forecast models can be used to provide the right decisions regarding the operation of National Electricity Markets (NEMs) through a more sustainable electricity pricing system, energy policy, and an evaluation of the feasibility of future energy distribution networks. Data intelligent models are considered as potential forecasting tools, although challenges related to issues of non-stationarity, periodicity, trends, stochastic behaviours in G data and selecting the most relevant model inputs remain a key challenge.

This doctoral thesis presents a novel study on the development of G forecasting models implemented at multiple lead-time forecast horizons utilizing data-intelligent techniques. The study develops predictive models using real G data from Queensland (second largest State in Australia) where the electricity demand continues to elevate. This research is therefore, divided into four primary objectives designed to produce a G forecasting system with data-intelligent models.

In first objective, the development and evaluation of a multivariate adaptive regression splines (MARS), support vector regression (SVR) and autoregressive integrated moving average (ARIMA) model was presented for short-term (30 minutes, hourly and daily) forecasting using Queensland's aggregated G data. MARS outperformed SVR and ARIMA models at 30-minute and hourly horizon, while SVR was the best model for daily G forecasting.

The second objective reported the successful design of SVR model for daily period, including short-term periods (e.g., weekends, working days, and public holidays), and the long-term (monthly) period. Subsequently, the hybrid SVR, with particle swarm optimization (i.e., PSO-SVR) integrated with improved empirical mode decomposition with adaptive noise (ICEEMDAN) tool was constructed where PSO is adopted to optimize SVR parameters and ICEEMDAN was adopted to address

non-linearity and non-stationary in G data. The capability of ICEEMDAN-PSO-SVR to forecast G was benchmarked against ICEEMDAN-MARS and ICEEMDAN-M5 Tree, including traditional PSO-SVR, MARS and M5 model tree methods.

As G is subjected to the influence of exogenous factors (e.g., climate variables), the third objective established a G forecasting model utilizing atmospheric inputs from the Scientific Information for Land Owners (SILO) observed data fields and the European Centre for Medium Range Weather Forecasting outputs. These models were developed using G extracted from the Energen database for eight stations in southeast Queensland for an artificial neural network (ANN) model over 6-hourly and daily forecast horizons.

The final objective was to advance the methods in previous objectives, by applying wavelet transformation (WT) as a decomposition tool to model daily G . Using real data from the University of Southern Queensland (Toowoomba, Ipswich, and Springfield), the maximum overlap discrete wavelet transform (MODWT) was adopted to construct the MODWT-PACF-online sequential extreme learning machine (OS-ELM) model. The results revealed that newly developed MODWT-PACF-OS-ELM (MPOE) model attained superior performance compared to the models without the WT algorithm.

In synopsis, the predictive models developed in this doctoral thesis will provide significant benefits to National Electricity Markets in respect to energy distribution and security, through new and improved energy demand forecasting tools. Energy forecasters can therefore adopt these novel methods, to address the issues of non-linearity and non-stationary in energy usage whilst constructing a real-time forecasting system tailored for energy industries, consumers, governments and other stakeholders.

Certification of Thesis

This Thesis is entirely the work of **Mohanad Shakir Khalid AL-Musaylh** except where otherwise acknowledged, with the majority of the authorship of the papers presented as a Thesis by Publication undertaken by the Student. The work is original and has not previously been submitted for any other award, except where acknowledged.

Associate Professor Ravinesh C Deo

Principal Supervisor

Professor Yan Li

Associate Supervisor

Professor Jan Adamowski

External Supervisor at McGill University, Canada

Student and supervisors signatures of endorsement are held at the University.

Contributions Statement

The details of the student and supervisor contributions for each research paper article presented in this doctoral thesis are as follows:

Article 1: Chapter 3

AL-Musaylh, M. S., Deo, R. C., Adamowski, J. F. and Li, Y., 2018. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics*, Vol. 35, pp.1-16. (Q1, Impact Factor: 3.772 and SINP: 2.387)

DOI: <https://doi.org/10.1016/j.aei.2017.11.002>

Author	Percentage Contributions	Task Performed
Mohanad S. AL-Musaylh (Candidate)	65%	Exploring the methodology in literature, data analysis, programing, preparation of tables and figures, writing and revising of the manuscript
Ravinesh C. Deo (Principal Supervisor)	20%	Supervised and assisted in scientific methodological development, editing and co-authorship of the manuscript.
Jan F. Adamowski (Associate Supervisor)	10%	Editing, proofreading, advised in method, and co-authorship of the manuscript
Yan Li (Associate Supervisor)	5%	Editing, proofreading and co-authorship of the manuscript

Article 2: Chapter 4

AL-Musaylh, M. S., Deo, R. C., Adamowski, J. F. and Li, Y., 2018. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon

electricity demand forecasting. *Applied Energy*, Vol. 217, pp.422-439. (Q1, Impact Factor: 8.426 and SINP: 2.616)

DOI: <https://doi.org/10.1016/j.apenergy.2018.02.140>

Author	Percentage Contributions	Task Performed
Mohanad S. AL-Musaylh (Candidate)	70%	Exploring the methodology in literature, data analysis, programing, preparation of tables and figures, writing and revising of the manuscript
Ravinesh C. Deo (Principal Supervisor)	20%	Supervised in scientific methodological development, editing and co-authorship of the manuscript.
Jan F. Adamowski (Associate Supervisor)	5%	Editing, proofreading, advised in method, and co-authorship of the manuscript
Yan Li (Associate Supervisor)	5%	Editing, proofreading and co-authorship of the manuscript

Article 3: Chapter 5

AL-Musaylh, M. S., Deo, R. C., Adamowski, J. F. and Li, Y., 2019. Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland, Australia. *Renewable and Sustainable Energy Reviews*, Vol. 113, pp.109-293. (Q1, Impact Factor: 10.556 and SINP: 3.694)

DOI: <https://doi.org/10.1016/j.rser.2019.109293>

Author	Percentage Contributions	Task Performed
Mohanad S. AL-Musaylh (Candidate)	70%	Exploring the methodology in literature, data analysis, programing,

		preparation of tables and figures, writing and revising of the manuscript
Ravinesh C. Deo (Principal Supervisor)	20%	Supervised in scientific methodological development, editing and co-authorship of the manuscript.
Jan F. Adamowski (Associate Supervisor)	5%	Editing, proofreading, advised in method, and co-authorship of the manuscript
Yan Li (Associate Supervisor)	5%	Editing, proofreading and co-authorship of the manuscript

Article 4: Chapter 6

AL-Musaylh, M. S., Deo, R. C., and Li, Y., 2020. Electrical energy demand forecasting model development and evaluation with maximum overlap discrete wavelet transform-online sequential extreme learning machines algorithms. *Energies*, Vol. 13, p. 2307. (Q1, Impact Factor: 2.707 and Cite Score: 3.18 Scopus).

DOI: <https://doi.org/10.3390/en13092307>

Author	Percentage Contributions	Task Performed
Mohanad S. AL-Musaylh (Candidate)	75%	Exploring the methodology in literature, data analysis, programing, preparation of tables and figures, writing and revising of the manuscript
Ravinesh C. Deo (Principal Supervisor)	20%	Supervised in scientific methodological development, editing and co-authorship of the manuscript.
Yan Li (Associate Supervisor)	5%	Editing, proofreading and co-authorship of the manuscript

Conference paper: Appendix A

AL-Musaylh, M. S., Deo, R. C., and Li, Y., 2018. Particle swarm optimized–support vector regression hybrid model for daily horizon electricity demand forecasting using climate dataset. *E3S Web of Conferences*, Vol. 64, p. 08001, *EDP Sciences*.

DOI: <https://doi.org/10.1051/e3sconf/20186408001>

Author	Percentage Contributions	Task Performed
Mohanad S. AL-Musaylh (Candidate)	85%	Exploring the methodology in literature, data analysis, programing, preparation of tables and figures, writing and revising of the manuscript
Ravinesh C. Deo (Principal Supervisor)	10%	Supervised in scientific methodological development, editing and co-authorship of the manuscript.
Yan Li (Associate Supervisor)	5%	Editing, proofreading and co-authorship of the manuscript

Acknowledgements

My sincere gratitude and appreciation should go to these individuals and organisations below for their support during my PhD study where it would not have been possible for me to complete the present doctoral thesis without their motivation and encouragement:

- ❖ My Principal Supervisor, **Associate Professor Ravinesh Deo** for his significant support and guidance to successfully complete my PhD project and thesis. His quick responses providing essential advices, suggestions and feedback assisted me widely to present this thesis with high impact factor journals that have reflected his well experience for publications in the field of work.
- ❖ My Associate Supervisors, **Professor Yan Li** and **Professor Jan Adamowski** for their useful comments, editing and proofreading the work. **Professor Adamowski** played a valuable role to improve the performances of the proposed methods while **Professor Li** assisted me with scientific writing presented in this thesis.
- ❖ **The Australian Energy Market Operator (AEMO), Energex, Scientific Information for Land Owners (SILO), the European Centre for Medium-Range Weather Forecasts (ECMWF) and the University of Southern Queensland, Australia** for providing all the required data for this study.
- ❖ **Mrs Alicia Logan** (an environmental manager, campus services, University of Southern Queensland, Australia) for her assistance to get the thesis final objective dataset.
- ❖ **The Ministry of Higher Education and Scientific Research, in the Government of Iraq and the Australian Commonwealth Government** through **the Research Training Program (RTP)** for supporting my PhD project.

- ❖ **Associate Professor Xiaohu Wen** (Cold and Arid Regions Engineering and Environmental Institute, China), **Dr. Mukesh Tiwari** (Department of Irrigation & Drainage Engineering, College of Agricultural Engineering and Technology, Anand Agricultural University, Gujarat, India) and **Dr John Quilty** (Department of Civil and Environmental Engineering, University of Waterloo, Canada) for their advices on the support vector machine model, bootstrap technique and wavelet technique, respectively.
- ❖ **The University of Southern Queensland** for all the facilities, tools and access to different academic services.
- ❖ **The Open Access College**, University of Southern Queensland, Australia, for providing excellence English courses that assisted me to write this thesis in academic way.
- ❖ **Dr. Barbara Harmes** (Language Centre, Open Access College, University of Southern Queensland, Australia), **Ms. Mela Aryal** (Ph.D. student, University of Southern Queensland, Australia), and **Ms. Laura Gilbert** (Ph.D. student in Natural Resource Sciences, McGill University, Canada) for providing help in proof-reading the work.
- ❖ **My colleagues** from Advanced Data Analytics (Environmental Modelling and Simulation Group), which has been administered by the Principal Supervisor, A/Prof Ravinesh Deo, for sharing ideas.
- ❖ My family in Australia including my wife (**Suha AL-Salim**), my daughters (**Jana and Malak**) and my son (**Hamzah**) as well as my extended family (**my mother, brothers and sister**) in my home country, Iraq for their supports, patience and motivations to complete my PhD degree.
- ❖ **My friends and everyone**, who have continually encouraged me to deal with difficulties.

Table of Contents

Abstract	i
Certification of Thesis	iii
Contributions Statement	iv
Acknowledgements	viii
Table of Contents	x
List of Figures	xii
List of Tables	xvii
List of Abbreviations	xxi
Chapter 1: Introduction	1
1.1 Background.....	1
1.2 Research Problem	2
1.3 Research Aim and Objectives.....	4
1.4 Thesis Layout.....	5
Chapter 2: Data and Methodology	7
2.1 Foreword.....	7
2.2 Study Area	7
2.3 Data Description	8
2.4 General Methodology	11
Chapter 3: Short-Term Electricity Demand Forecasting Using Queensland’s Aggregated Data	16
3.1 Foreword.....	16
3.2 Research Highlights	16
3.3 Published Article I	16

Chapter 4: Multiple Electricity Demand Forecasting Horizons Using Hybrid Models	33
4.1 Foreword.....	33
4.2 Research Highlights	33
4.3 Published Article II.....	33
Chapter 5: Using a Wide Range of Data to Investigate the Impact of Climate and Atmospheric Variables on Electricity Demand Forecasting	52
5.1 Foreword.....	52
5.2 Research Highlights	52
5.3 Published Article III.....	53
Chapter 6: Using a Suitable Wavelet Transformation Tool to Achieve Better Electricity Demand Forecasting	76
6.1 Foreword.....	76
6.2 Research Highlights	76
6.3 Published Article IV	77
Chapter 7: Conclusions and Future Scope	97
7.1 Conclusions.....	97
7.2 Limitations and Future Scope	99
References	101
Appendix A: Conference Paper	107

List of Figures

Chapter 1

Figure 1.1 Thesis flow chart.

Chapter 2

Figure 2.1 Types of data intelligent models used in each chapter (objective) in this research thesis.

Chapter 3

Figure 1 Time-series of electricity demand (G) data and various forecasting periods.

Figure 2 Correlation coefficient (r) based on the partial autocorrelation function (PACF) of predictors (i.e., electricity demand, G) used for developing the support vector regression (SVR) and multivariate regression splines (MARS) models. Statistically significant lags at the 95% confidence interval are marked (blue).

Figure 3 Illustration of SVR and MARS model parameters for 1.0 h forecast horizon, (H_4) model.

Figure 4 Scatterplot of the forecasted, G_i^{for} vs. the observed, G_i^{obs} electricity demand data in the testing period for the 0.5 h forecast horizon, (a) SVR(T^b) (b) MARS(T^b) and (c) ARIMA^b. A linear regression line, $y = G_i^{for} = a'G_i^{obs} + b'$ with the correlation coefficient, r is included.

Figure 5 The caption description is the same as that in Figure 4 except for the 1.0 h forecast horizon, (a) SVR(H_4) (b) MARS(H_4) and (c) ARIMA.

Figure 6 The caption description is the same as that in Figure 4 except for the 24 h forecast horizon, (a) SVR(D_3) (b) MARS(D_3) and (c) ARIMA.

Figure 7 The caption description is the same as that in Figure 4 except for the 0.5 h forecast horizon, (a) SVR(T^c) (b) MARS(T^c) and (c) ARIMA^c.

Figure 8 Boxplots of the absolute forecasted error, $|FE| = |G_{FOR,i} - G_{OBS,i}|$ for: (a) 0.5 h, (b) 1.0 h and (c) 24 h forecast horizons using the MARS, SVR and ARIMA models.

Figure 9 Empirical cumulative distribution function (ECDF) of the forecast error, $|FE|$ for: (a) 0.5 h, (b) 1.0 h and (c) 24 h forecast horizons using the MARS, SVR and ARIMA models.

Figure 10 The top ten peak relative forecast errors (%) generated by the MARS, SVR and ARIMA models for: (a) 0.5 h, (b) 1.0 h and (c) 24 h forecast horizons.

Chapter 4

Figure 1 Time-series of the electricity demand (G MW) data and the various forecast periods.

Figure 2 (a) Intrinsic mode functions (IMFs) and residual constructed using ICEEMDAN for the whole week's data. (b) Investigating of the statistically significant lagged data used for developing the models for IMFs and residual using the whole week's data.

Figure 3 The steps of the model development process for both traditional and hybrid models.

Figure 4 Illustration of ICEEMDAN–SVR, ICEEMDAN–MARS and ICEEMDAN–M5 model tree parameters and fitting for public holiday forecast horizon using PACF–residual data in the training period. (a) PSO fitness curve to identify the regulation function (C), kernel widths (σ), and loss function (ϵ). (b) Best number of basis functions based on the lowest generalized cross-validation statistic (GCV). $GCV (\approx 0.15 \cdot 10^{-6} \text{ MW}^2)$ with 2 basis functions and y -intercept, a_0 is indicated. (c) Tree generated by M5 model.

Figure 5 Scatterplot of the G -forecasted vs. G -observed of electricity demand data in the testing period for the (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons using the six models.

The equation of the linear regression line and the coefficient of determination are incorporated.

Figure 6 Boxplots of the absolute forecasted error, $|FE| = |G_{FOR,i} - G_{OBS,i}|$ in the testing period for the (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons using the six models.

Figure 7 Taylor diagram showing the correlation coefficient between G -observed and G -forecasted and standard deviation for each model utilizing (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons.

Figure 8 Empirical cumulative distribution function (ECDF) of the forecast error, $|FE|$ in the testing period for the (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons using the six models.

Figure 9 Boxplots of the absolute forecasted error, $|FE| = |G_{FOR,i} - G_{OBS,i}|$ in the testing period for the monthly forecast horizons using the six models.

Figure 10 Taylor diagram showing the correlation coefficient between G -observed and G -forecasted and standard deviation for each model utilizing monthly forecast horizon.

Figure 11 Relative error analysis over the monthly forecast horizon in the testing period. The six models and the magnitude of relative forecasted errors are indicated on top of each plot and on the radial axis, respectively.

Chapter 5

Figure 1 Map of the present study area showing the locations in southeast Queensland where the energy demand forecasting models were developed and tested using climate-based predictor variables. (a) Energex stations shown as green circles (b) SILO points shown as red triangles and (c) ECMWF (ERA-Interim) Reanalysis grids are shown as pink boxes where each grid contains four points with the mean of

these points used as the data point for that grid. Information in Table 1 was used to generate the spatial map.

- Figure 2** Time-series of the electricity demand (G , MW) data used for the Beerwah study site for 6-hour and daily periods.
- Figure 3** Cross-correlation coefficients (r_{cross}) performed to investigate the covariance between daily electricity demand (G) vs. the predictor variables from SILO (variables 1 to 6) and ERA-Interim Reanalysis dataset (variables 7 to 57) for the Beerwah station in the model training period. The 95% significance boundaries are shown in blue, indicating the variables selected to calculate statistically significant lagged 1 inputs matrix for the models. The selected variables are shown in boldface. Details of the variables are shown in Table 3.
- Figure 4** Statistically significant lags of the electricity demand data (G) analysed in the training phase to construct the model's input variables. (a) The first three lags were selected for daily Beerwah data (b) Lags 2, 4 and 5 were selected for 6-hour Burpengary data (c) Lags 1, 4, 5 and 6 were selected for 6-hour Ipswich South data.
- Figure 5** Selecting the best ANN model in the test period for 6-hour Beerwah forecast horizons based on the statistical indicators (Eqs. 7-13). Only the 50 ANN models with the best parameters, indicated in Table 5, are shown.
- Figure 6** Bar graphs of the Legates & McCabe's Index (ELM) showing the optimal 6-hour models for the eight tested stations in southeast Queensland.
- Figure 7** Boxplots of the 6-hour absolute forecasting error, $|FE| = |G_i^{for} - G_i^{obs}|$ in the testing period for the eight stations.
- Figure 8** The percentage frequency distribution of the 6-hour forecasted error, $|FE|$ for the eight stations in the testing phase.

Figure 9 Scatterplot of the daily forecasted electricity demand (G^{for}) vs. observed electricity demand (G^{obs}) in the testing phase for the eight stations.

Figure 10 Electricity demand (G) observed with 95% prediction bands forecasted using B-hybrid-ANN for all the eight stations for the (a) 6-hour and (b) daily forecasting horizons.

Figure 11 Non-stationary daily time-series of the eight variables from SILO and ECMWF (ERA-Interim) Reanalysis fields.

Chapter 6

Figure 1 Flowchart showing the model development stages.

Figure 2 Daily electricity demand (G , kW) time-series for the three study sites of Toowoomba, Ipswich and Springfield campuses.

Figure 3 The model input variables constructed from the statistically significant lags at a 95% confidence interval from the original daily G -datasets in the training period for the three study sites based on correlation coefficient (r) of predictors (lags) using the partial autocorrelation function (PACF).

Figure 4 Maximum overlap discrete wavelet transform (MODWT) coefficients constructed using lag 1 of the Toowoomba campus dataset with the optimum wavelet (scaling) filter (fk_8) and wavelet level 2. WC (a, b) and SC (c) represent the wavelet and scaling coefficients, respectively.

Figure 5 Observed vs. forecasted G -data in the testing period with the optimal POE and MPOE models for the three study sites.

Figure 6 Scatterplots of the observed and forecasted G -data in the testing phase with the optimal models of POE and MPOE. Equations of linear regression and the coefficient of determination are incorporated. (a) Toowoomba campus. (b) Ipswich campus. (c) Springfield campus.

Figure 7 Boxplots of the absolute forecasted error $|FE|$ in the testing dataset for the three study sites with the optimal models of POE and MPOE.

List of Tables

Chapter 2

Table 2.1 Details of all data used in this study

Chapter 3

Table 1 Descriptive statistics of the electricity demand (G) (MW) data aggregated for the Queensland (QLD) study region.

Table 2 Model designation for the MARS, SVR and ARIMA for 0.5 h, 1.0 h and 24 h forecast horizons.

Table 3 Parameters for the SVR and ARIMA model presented in the training period for 0.5 h, 1.0 h and 24 h forecast horizons.

Table 4 The MARS model forecast equation, $y = a_0 + \sum_{m=1}^M a_m BF_m(X)$ with optimum basis functions (BF_m), and generalized cross validation statistic (GCV) in MW^2 for all horizons, in training period.

Table 5 Evaluation of the optimal models attained for 0.5 h, 1.0 h and 24 h forecast horizons in the test period.

Table 6 The relative root mean square error $RMSE_{\bar{G}}(\%)$, mean absolute percentage error $MAE_{\bar{G}}(\%)$ and Legates & McCabes Index (E_{LM}) for the optimal models in test datasets.

Table 7 Evaluation of the differences in the absolute value of forecast error statistics based on observed and forecasted G in test period for the optimal models.

Chapter 4

Table 1 Descriptive statistics of the total electricity demand (G) (MW) for the Queensland (QLD) study region with different forecast horizons.

Table 2 Root-mean square error ($RMSE$; MW), and Legates & McCabes Index (E_{LM}) for the weekend, working days, whole week, public holiday and monthly forecast horizons in the validation dataset. *The most accurate model is boldfaced.*

- Table 3** The performance of the models in the test period for the short-term measured by Willmott’s index (WI), Nash–Sutcliffe coefficient (E_{NS}), root-mean square error ($RMSE$; MW), and mean absolute error (MAE ; MW). *The most accurate model is boldfaced.*
- Table 4** The relative root-mean square error ($RRMSE$ %), mean absolute percentage error ($MAPE$ %) and Legates & McCabes Index (E_{LM}) for the models in the test datasets for the short-term. *The most accurate model is boldfaced.*
- Table 5** The performance of the models in the test period for the long-term (monthly) measured by Willmott’s index (WI), Nash–Sutcliffe coefficient (E_{NS}), root-mean square error ($RMSE$; MW), and mean absolute error (MAE ; MW). *The most accurate model is boldfaced.*
- Table 6** The relative root-mean square error ($RRMSE$ %), mean absolute percentage error ($MAPE$ %) and Legates & McCabes Index (E_{LM}) for the models in the test datasets for the long-term (monthly) horizon. *The most accurate model is boldfaced.*

Chapter 5

- Table 1** The longitudes and latitudes for data locations used in this study: (a) Energex (b) SILO (c) ECMWF (ERA-Interim) Reanalysis.
- Table 2** Descriptive statistics of the electricity demand (G) (MW) for the eight Energex stations with 6-hour and daily forecast horizons.
- Table 3** Model input variables from SILO and ERA-Interim Reanalysis used for 6-hour and daily G^{for} .
- Table 4** Data splitting technique used for model development and testing.
- Table 5** Optimum model development for ANN, MARS, and ARIMA showing the models’ parameters and predictor datasets for both forecast horizons in the training period.
- Table 6** Optimum model development for hybrid ANN and B-hybrid ANN showing the models’ parameters and predictor datasets for both forecast horizons in the training period.

- Table 7** Root-mean square error (*RMSE*, MW), for all the station datasets for the 6-hour and daily forecast horizons in the training period using the ANN, MARS, MLR, hybrid ANN, and ARIMA models.
- Table 8** Optimum model performance in the test period for the 6-hour forecast horizon as shown by Willmott's index (*WI*), Nash–Sutcliffe efficiency coefficient (*E_{NS}*), root-mean square error (*RMSE*, MW), and mean absolute error (*MAE*, MW).
- Table 9** The mean absolute percentage error (*MAPE*, %), relative root-mean square error (*RRMSE*, %), and Legates & McCabes Index (*E_{LM}*) for the optimum models for the 6-hour forecast horizon for the testing datasets.
- Table 10** Optimum model performance in the test period for the daily forecast horizon by Willmott's index (*WI*), Nash–Sutcliffe efficiency coefficient (*E_{NS}*), root-mean square error (*RMSE*, MW), and mean absolute error (*MAE*, MW).
- Table 11** The mean absolute percentage error (*MAPE*, %), relative root-mean square error (*RRMSE*, %), and Legates & McCabe's Index (*E_{LM}*) for the optimum models with daily forecast horizon in the test datasets.
- Table 12** Optimum hybrid ANN model performance in the test period for both forecast horizons by Willmott's index (*WI*), Nash–Sutcliffe efficiency coefficient (*E_{NS}*), root-mean square error (*RMSE*, MW), mean absolute error (*MAE*, MW), mean absolute percentage error (*MAPE*, %), relative root-mean square error (*RRMSE*, %), and Legates & McCabe's Index (*E_{LM}*).

Chapter 6

- Table 1** Data splitting and descriptive statistics for the *G*-data from the University of Southern Queensland campuses' three stations.
- Table 2** Optimum model performance and parameters in the training and validation phases based on correlation coefficient (*r*) and root-mean square error (*RMSE*; kW), for the three stations with the daily forecast

horizon. The models in **boldface** are the optimal (best performing) models.

Table 3 Optimum model performance in the testing phase for daily forecast horizon based on Willmott's index (*WI*), Nash–Sutcliffe model efficiency coefficient (*E_{NS}*), root-mean square error (*RMSE*; kW), mean absolute error (*MAE*; kW), mean absolute percentage error (*MAPE* %), relative root-mean square error (*RRMSE* %), as well as Legates & McCabes Index (*LM*) for the three stations. The models in **boldface** are the optimal (best performing) models.

Table 4 Wilcoxon Signed-Rank test and *T* test results for the |FE| of the MOPE model vs. the |FE| for the POE model.

List of Abbreviations

MW	Megawatt
G	Electricity Demand (Load)
MARS	Multivariate Adaptive Regression Splines
SVR	Support Vector Regression
ARIMA	Autoregressive Integrated Moving Average
r	Correlation Coefficient
$RMSE$	Root Mean Square Error (MW or kW)
MAE	Mean Absolute Error (MW or kW)
$RMSE_{\bar{G}}$ or $RRMSE$	Relative Root Mean Square Error, %
$MAE_{\bar{G}}$ or $MAPE$	Mean Absolute Percentage Error, %
WI	Willmott's Index of Agreement
E_{NS}	Nash–Sutcliffe Coefficient
E_{LM} or LM	Legates and McCabe Index
ANN	Artificial Neural Network
RBF	Radial Basis Function for SVR
σ	Kernel Width for SVR
C	Regulation for SVR
$BF_m(X)$	Spline Basis Function for MARS
GCV	Generalized Cross-Validation
p	Autoregressive Term in ARIMA
d	Degree of Differencing in ARIMA
q	Moving Average Term in ARIMA
AEMO	Australian Energy Market Operator
NEM	National Electricity Market

ACF	Auto-Correlation Function
PACF	Partial Auto-Correlation Function
MSE	Mean Square Error (MW or kW)
R^2	Coefficient of Determination
AIC	Akaike Information Criterion
L	Log Likelihood
σ^2	Variance
G_i^{for}	i^{th} Forecasted Value of G , MW or kW)
G_i^{obs}	i^{th} Observed Value of G , MW or kW)
Q_{25}	Lower Quartile (25 th Percentile)
Q_{50}	Median Quartile (50 th Percentile)
Q_{75}	Upper Quartile (75 th Percentile)
d	Degree of Differencing in ARIMA
ω	Weighting Factor of PSO
ω_{min} and ω_{max}	The Minimum and Maximum of ω
ICEEMDAN	Improved Version of Empirical Mode Decomposition With Adaptive Noise
ε	Loss Function
PSO	Particle Swarm Optimization
T_{max}	Maximum Number of Iterations in PSO
ECDF	Empirical Cumulative Distribution Function
IMF	Intrinsic Mode Functions
DWT	Discrete Wavelet Transform
c_1 and c_2	PSO Parameters
MODWT	Maximum Overlap Discrete Wavelet Transform
EMD	Empirical Mode Decomposition

WT	Wavelet Transforms
N	The Initial Population of PSO
$\overline{G^{for}}$	The Mean of Forecasted Values
$\overline{G^{obs}}$	The Mean of Observed Values
$ FE $	Forecasted Error Statistics
EEMD	Ensemble EMD
CEEMDAN	Complete EEMD With Adaptive Noise
VMD	Variational Mode Decomposition
SILO	Scientific Information for Land Owners
ECMWF	European Centre for Medium Range Weather Forecasts
MLR	Multiple Linear Regression
<i>trainbfg</i>	Levenberg-Marquardt
<i>trainlm</i>	Broyden-Fletcher-GoldfarbShanno
<i>tansig</i>	Tangent Sigmoid
<i>logsig</i>	Log Sigmoid
<i>purelin</i>	Positive Linear
L	Hidden Neuron Size
B	Bootstrapping Algorithm
BMA	Bayesian Model Averaging
FA	Firefly Algorithm
LH	Log Likelihood
AIC	Akaike Information Criterion
SDGs	Sustainable Development Goals
db_i	Daubechies Wavelet Filter
$f(\cdot)$	SFLM Activation Function

fk_i	Fejer-Korovkin Wavelet Filter
$g_{j,l}$	j^{th} Level Scaling Filter
$h_{j,l}$	j^{th} Level Wavelet Filter
$int(.)$	Nearest Integer Function
k	Number of Hidden Nodes in SFLM
kW	Kilowatts
sym_i	Symlets Wavelet Filter
w_i	Weight Vectors Linking i^{th} Hidden Node With the Input Node
CRO	Coral Reef Optimisation
MRA	Multiresolution Analysis
ELM	Extreme Learning Machine
GGA	Grouping Genetic Algorithm
H	SFLM's Hidden Layer Output Matrix
H^*	Inverse of H Matrix
IIS	Iterative Input Selection
J	Decomposition Level
L_j	Width of the j^{th} Level Filters
M	Hidden Neuron Size
MODIS	Moderate Resolution Imaging Spectroradiometer (NASA)
OS-ELM	Online Sequential Extreme Learning Machine
MPOE	MODWT-PACF-OS-ELM
POE	PACF-OS-ELM
SC	Scaling Coefficient (MODWT)
SFLM	Single Layer Feed-Forward Neural Network
$V_{j,i}$	MODWT Scaling Coefficients

$W_{j,i}$ MODWT Wavelet Coefficients

WC1, WC2 MODWT Wavelet Coefficients

Chapter 1: Introduction

1.1 Background

Electricity demand (G) is subjected to increases and decreases due to the influence of various interacting factors, such as the economy, buildings and population growth, changing weather patterns, increasing use of solar panels, business activities, and social variables. The increasing use of electric cars and air conditioners usage due to the recent developments and global warming are also considered two other significant factors that affect electricity demand. Therefore, a robust forecasting method, used to constantly monitor the usage of electricity, which can take these factors in consideration, is needed to more precisely forecast the G and efficiently support national electricity markets.

Fundamentally, the forecasting of G is a challenging optimisation problem especially in improving the business efficiency of the electricity industry. Currently, it plays a significant role in developing an efficient distribution and effective network system (Hu, Bao & Xiong 2013; Kaytez et al. 2015). The most precise models for G forecasting are necessary to operate company's facilities and decision-making on load switching, grid management and infrastructure in response to growing consumers' needs (Akay & Atak 2007). Forecasting models are also useful planning tools for energy organizations, transmission, distribution and marketing in a globalizing world. An estimate of electricity consumption can reflect the degree of economic growth by an analysis of the causal relationships with power consumptions (Kaboli, Selvaraj & Rahim 2016). Nasr, Badr and Younes (2002); Zareipour, Bhattacharya and Canizares (2006), and Hu, Bao and Xiong (2013) pointed out that many factors including demographic, environmental, social, recreational and seasonal may adversely affect the preciseness of G . Consequently, robust models are required to address any energy-forecasting inaccuracies that may be faced by national electricity markets.

Due to the growth in computational power, the applications of data intelligent models in energy forecasting have risen rapidly (Suganthi & Samuel 2012). These models have adopted soft computing algorithms that are less complex than deterministic models where the only requirement is a set of inputs that are related to past changes in G . Yet, these models are able to forecast the energy demand system that is pervasively imprecise, uncertain and hard to predict precisely (Haida & Muto 1994). In forecasting and predicting of any variable, soft computing parallels the

remarkable ability of machines to reason and learn in a programming environment that encompasses uncertainty and imprecision as an intelligent framework. Such models mimic the ability of human brains to extract patterns and attributes in predicting data that are approximately represented in the temporal evolutions in future. Consequently, data-driven models are showing significantly accurate forecasting results, and as such, have been widely adopted (Florens et al. 2007; Xydias et al. 2016).

Recent advancements in data analytics, including modelling and simulation, have led to improvements in data intelligent models, which have subsequently become a standard methodology for data prediction due to their potential to employ machine learning algorithms to detect the relationships between predictors and objective variables using historical datasets. Using these methods, this PhD thesis explores existing methods and develops a better performance model for G forecasting by suggesting new and advanced predictive models that may, therefore, be of potential use to forecasters in national electricity markets (e.g. Australian Energy Market Operator (AEMO)).

1.2 Research Problem

The research into electricity demand should be undertaken for several reasons, such as making profitable investments, building an efficient system, growing the capability of an existing system and scheduling energy distributions (Türkay & Demren 2011). G estimations are vital for energy industries and governments to respond to competitive electricity markets. The precision (predictive accuracy) of G is very important since just 1% increase in error can lead to a loss of millions of dollars in operational costs (Bunn & Farmer 1985; Haida & Muto 1994; Fan & Chen 2006). According to Erdogdu (2007), if official demand projections are overestimated or underestimated, this can endanger the development of a coherent energy policy and a healthy market. Consequently, there exists a strong need for robust predictive models that are capable of energy forecasting with minimal inaccuracies within the context of the national electricity markets.

Queensland is the Australia's second largest State in area where the end-user demand for electricity continues to increase because people keep moving into the State. The State-averaged energy consumption is relatively high and amounts to 23% of the national (2012–2013) demand (Australian Government 2015). Currently in Queensland, the G data are forecasted by the Australian Energy Market Operator

(AEMO) yielding high errors in forecasts for electricity demand data due to the use of old forecasting techniques. Robust predictive models can provide better performance that can lead to significant outcomes for national electricity markets. However, at present, none are available in Queensland.

Additionally, the G data are affected by different factors, such as climate variables that should be taken into the account when the G data are forecasted. Accordingly, the influences and complex temporal behaviors of non-stationarity features of those variables on G data should be explored to improve models' forecast accuracy. To address these issues, a robust forecasted model from data intelligent models, which can significantly deal with different inputs, needs to be adopted to discover the impact of the climate variables on G data.

A review of the previous and recent studies showed that various machine learning models have been successfully adopted, including, but not limited to, the multivariate adaptive regression splines (MARS) (Zareipour, Bhattacharya & Canizares 2006; Deo, Kisi & Singh 2017), support vector regression (SVR) (Guo et al. 2006; Deo, Wen & Qi 2016), autoregressive integrated moving average (ARIMA) (Contreras et al. 2003), hybrid particle swarm optimization (PSO)-SVR (Li & Li 2016), M5 model tree (Pal & Deswal 2009), artificial neural network (ANN) (Deo & Şahin 2017; Prasad et al. 2017), multiple linear regression (MLR) (Civelekoglu et al. 2007) and online sequential extreme learning machine (OS-ELM) (Ali et al. 2018) models for different data forecasting.

Data decomposition methods, such as improved versions of empirical mode decomposition with adaptive noise (ICEEMDAN) (Li & Li 2016) and maximum overlap discrete wavelet transform (MODWT) (Quilty & Adamowski 2018) are also required to address the non-stationarity features of those input variables. Furthermore, selecting the best models' parameters is another challenge for G forecasting. This can be achieved by a particle swarm optimization (PSO) (Li & Li 2016), grid search (Hsu, Chang & Lin 2003) and trial-and-error techniques. Therefore, this study investigates the ability of the methods above for G forecasting.

A combination of forecasting approaches and uncertainty assessments, which have not been adopted by national electricity markets, can also be used to improve the model's forecast accuracy and assess the model's stability based on the prediction intervals, respectively. Hence, this study develops a hybrid ANN model by combining the forecasts of the ANN, MARS, and MLR models. In addition, a data-driven

technique of bootstrap, which purposes to reduce uncertainties through enormous resampling with replacement (Efron 1992; Efron & Tibshirani 1994), is used and integrated with the ANN model to explore the uncertainty in G forecasting using an ensemble of the hybrid ANN model.

In summary, the PhD thesis aims to address issues associated with G forecasting based on data intelligent techniques and adopting G forecasting models at multiple forecast horizons. The study area focuses on Queensland, Australia's sunshine State and the third most populous State where electricity monitoring and forecasting devices can play a useful role in the State's energy management systems.

1.3 Research Aim and Objectives

The key purpose of this research thesis, presented as a collection of four high quality, $Q1$ publications, is to develop a set of the most high-precision hybrid data-intelligent models for G forecasting based on the Queensland region across short and long-term periods.

To achieve the aim of this study, the following objectives are presented to:

- 1- Develop and evaluate the multivariate adaptive regression splines (MARS), support vector regression (SVR) and autoregressive integrated moving average (ARIMA) models for short-term (30 minutes, hourly and daily) G forecasting using Queensland's aggregated G data. The article has been published in *Advanced Engineering Informatics* (Vol. 35, Pages 1-16).
- 2- Further improve the whole weeks (daily) forecasting plus other multiple-horizons electricity demand forecasting, including short-term for weekends, working days, and public holidays, and long-term for monthly. A novel hybrid particle swarm optimization (PSO)-SVR model integrated with an improved version of empirical mode decomposition with adaptive noise (ICEEMDAN) tool is constructed and its ability benchmarked with alternative techniques of standalone modelling procedures. The article has been published in *Applied energy* (Vol. 217, Pages 422-439).
- 3- Design and establish a relatively new approach for the 6-hour and daily G forecasting horizons based on an artificial neural network (ANN) model utilizing the climate input datasets for the eight stations located in the southeast Queensland area. The article has been published in *Renewable and Sustainable Energy Reviews* (Vol. 113, Pages 109-293).

- 4- Apply wavelet transformation (WT), a very popular decomposition tool that has been incorrectly applied in many recent studies, on a regional area at the University of Southern Queensland (USQ), Australia. A maximum overlap discrete wavelet transform-coupled online sequential extreme learning machine (MODWT-OS-ELM) model is developed in this objective based on the three university campuses datasets of Toowoomba, Ipswich and Springfield for daily G forecasting. The article has been published in the journal of *Energies* (Vol. 13, Page 2307).

To fulfil these objectives, a copy of the above mentioned, four $Q1$ papers, are included as a core contribution of this PhD Thesis by Publications.

1.4 Thesis Layout

The thesis schematic, which is shown in Figure 1.1, is organized into seven chapters as follows:

Chapter 1 presents the introductory background and the statement of the problem pertaining to the research and presents the objectives of this study.

Chapter 2 describes the study area, data and general methodology used in this study and sets the scene for the following chapters. This chapter provides general viewpoints while the specific study area, data and methods are presented in the respective chapters.

Chapter 3 presents as a published journal article in the journal of *Advanced Engineering Informatics* (DOI: <https://doi.org/10.1016/j.aei.2017.11.002>). It explores the application of data intelligent models including MARS, SVR and ARIMA to forecast G data.

Chapter 4 presents as a published journal article in the journal of *Applied Energy* (DOI: <https://doi.org/10.1016/j.apenergy.2018.02.140>). The forecast accuracy of the G data was improved and the forecast error was reduced in this chapter using the ICEEMDAN method.

Chapter 5 presents as a published journal article in the journal of *Renewable and Sustainable Energy Reviews* (DOI: <https://doi.org/10.1016/j.rser.2019.109293>). The

new forecast methodology of incorporating enormous input variables was developed in this chapter with hybrid ANN.

Chapter 6 presents as a published journal article in the journal of *Energies* (DOI: <https://doi.org/10.3390/en13092307>). A suitable wavelet transformation of MODWT was applied and integrated with the OS-ELM model in this chapter.

Chapter 7 presents the summary of the thesis with concluding remarks, limitations, and recommendations for future works.

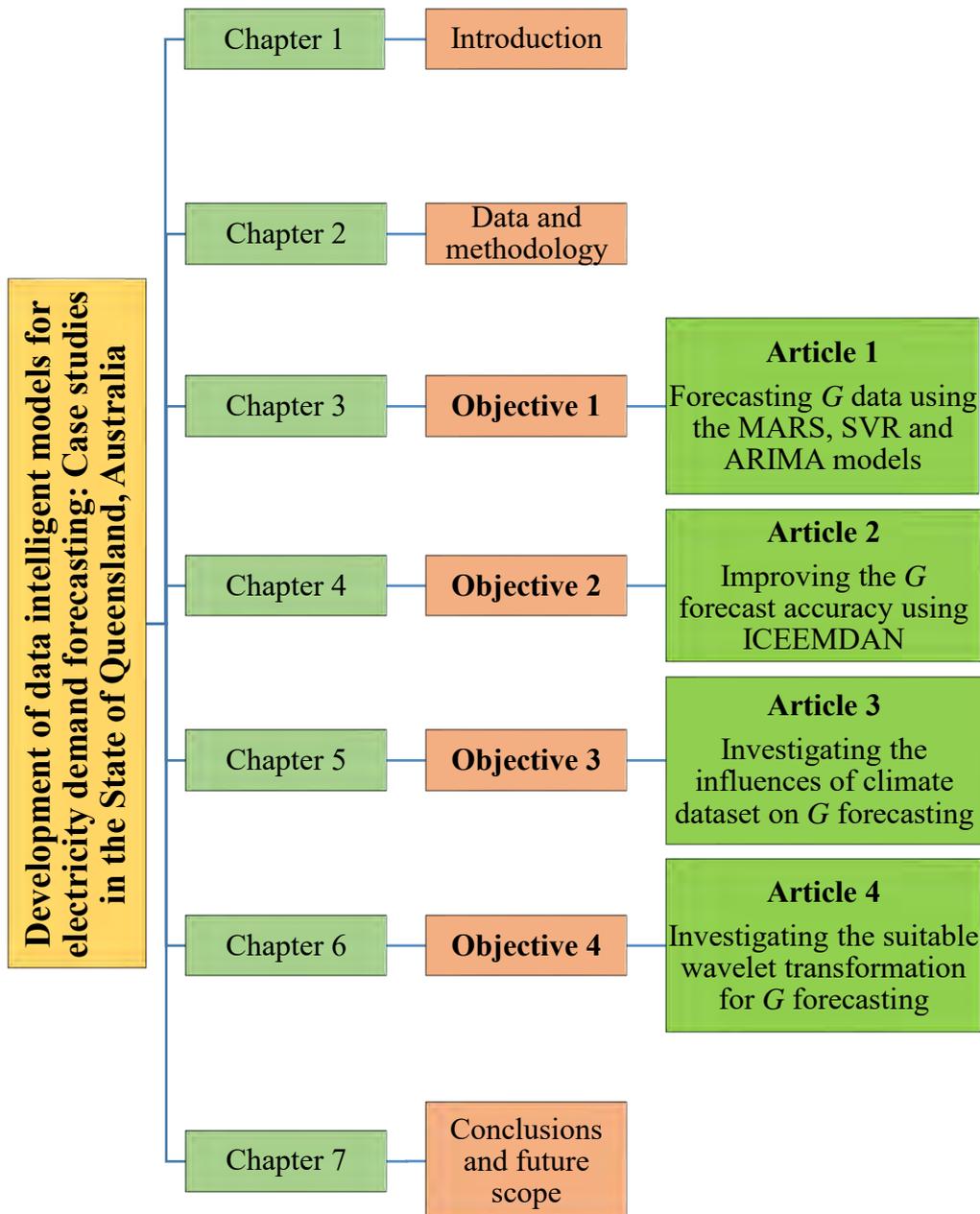


Figure 1.1: Thesis flow chart.

Chapter 2: Data and Methodology

2.1 Foreword

An overview of the study area and the datasets used in developing the hybridised data intelligent G forecasting models for this doctoral study is presented in this chapter. Within Queensland (QLD), Australia, a State with a heavy demand for electricity, different study sites were selected depending on the availability of the data, to achieve the objectives of this PhD thesis. These locations were described in detail in their related chapters while this chapter provides a brief summary for these study areas and the relevant data. A brief and general methodology is also introduced in this chapter whereas the specific methodology of the model development techniques are presented in detail in the other relevant chapters. To sum up, this chapter introduces the general procedure that was used to develop the hybridised data intelligent models for G forecasting.

2.2 Study Area

In order to develop the data intelligent G forecasting models, Queensland, Australia with a population of 4.75 million and an area of 1.85 million km^2 (Grinham et al. 2018) is used as the primary study area. Between 2017 and 2018, Queensland, located between 20.9176° S (latitude) and 142.7028° E (longitude), had the third fastest growing city in Australia (Brisbane) with a 50,100 increase in population (Australian Bureau of Statistics 2019). The same source also pointed out that overseas and internal migrations, as well as natural increase, each represented about one third of Brisbane's population growth. While Queensland is the fastest growing State in Australia, it is also the most decentralised of all the mainland States, with 49% of the population living in the capital city, compared with 68% in other States and it has strong business and employment opportunities that encourage people to live in Queensland (Queensland Government 2019). At the same time, due to the increase in electricity demand, consumers have required the use of larger and more remote generators, so Queensland's network has become more discrete than any other State (Quezada et al. 2014). As the political momentum continues to encourage the building of more generators, the national electricity market, managed by the Australian Energy Market Operator (AEMO) has been established to link all eastern Australian States (Quezada

et al. 2014). Consequently, energy demands of G data are most likely to be affected in Queensland due to the influence of the reasons given above.

Specifically, the first and the second objectives of this study are focussed on the whole of Queensland, which is located on the north-east of Australia. On the other hand, as Queensland has several distinct regions that have different weather affecting the use of the G data, the second objective is based on suburbs in the cities of Brisbane, Gold Coast, Sunshine Coast, Logan, Ipswich, Redlands, and Moreton Bay located in southeast Queensland, which covers more than 200 areas, to investigate the impact of the climate and weather factors on G data. Due to data limitations, such as missing points, zeros, negative values, absent G values for previous or subsequent years, and difficulties in matching G data with corresponding predictor (input) datasets, only eight areas, including Beerwah, Burpengary, Ipswich South, Narangba, Redcliffe, Redland Bay, Wynnum, and Zillmere were selected after examining all the station datasets. Finally, the fourth study objective uses data from three regional campuses (Toowoomba, Ipswich and Springfield) at the University of Southern Queensland (USQ), Australia.

2.3 Data Description

To develop data intelligent G forecasting models, a variety of sources have been used to extract the required data for this study. The details of these data with their respective sources are described in Table 2.1.

For objective 1, the aggregated data from entire State of Queensland (QLD), Australia were used and extracted from the Australian Energy Market Operator (AEMO) (Australian Energy Market Operator 2016) for the period of 01-January-2012 to 31-December-2015. AEMO prepares the seasonal peak G and energy consumption forecasts for the QLD region of the National Electricity Market (NEM). AEMO provides half-hourly QLD G values in megawatt where each day's electricity is divided into 48 periods, which correspond with the NEM settlement periods. Period 1 is from midnight to 0:30 am Eastern Standard Time (EST). To expand the forecasting horizon to 1.0 h and 24 h periods to obtain G values, an arithmetic averaging of the half-hourly data was performed. Consequently, three forecast horizons of 0.5 h, 1.0 h and 24 h were presented utilising data from periods 01-12-2015 to 31-12-2015, 01-11-

2015 to 31-12-2015, and 01-01-2012 to 31-12-2015, respectively. Chapter 3 provides more details of these data.

Table 2.1: Details of all data used in this study.

Objective	Data type	Source	Study period (dd-mm-yyyy)	Forecast horizon
Objective one (Chapter three) (Paper one)	Electricity demand data (<i>G</i>)	Australian Energy Market Operator (2016)	01-12-2015 to 31-12-2015	Half-hourly (0.5 h)
			01-11-2015 to 31-12-2015	Hourly (1.0 h)
			01-01-2012 to 31-12-2015	Daily (24 h)
			01-01-2012 to 31-12-2016	Weekend Working days Whole week
Objective two (Chapter four) (Paper two)			01-01-2000 to 31-12-2016	Public holiday Monthly
Objective three (Chapter five) (Paper three)	Meteorological variables	Scientific Information for Land Owners (SILO) (Jeffrey et al. 2001)	01-07-2014 to 30-06-2017 and	Six-hourly (6 h) and
	Atmospheric Parameters	Interim ERA European Center for medium- range weather forecasting reanalysis (ECMWF)	01-07-2014 to 29-06-2017	Daily

		reanalysis) (Dee et al. 2011)		
	Electricity demand data (<i>G</i>)	Energex (2018)		
			01-01-2013 to 31-12-2014 Toowoomba campus (Main feed)	
Objective four (Chapter six) (Paper four)	Electricity demand data (<i>G</i>)	Campus services of the University of Southern Queensland (USQ), Australia	01-09-2015 to 31-08-2016 Ipswich campus (Main feed)	Daily
			01-09-2015 to 31-08-2016 Springfield campus (A Block)	

For objective 2, the data from the same source as objective 1 were adopted. However, multiple datasets including both short-term [weekend (Saturday to Sunday), working days (Monday to Friday), whole week (Monday to Sunday), and public holiday], and long-term (monthly) horizons were used to forecast *G*. The 30-min data periods were converted to short and long terms by obtaining total values for each day and month, respectively. The datasets for the weekend, working days and whole week were collected from 01-01–2012 to 31–12–2016 and the dataset for the public holidays and monthly horizons were collected from 01-01–2000 to 31–12–2016. Chapter 4 provides more details of these data.

For objective 3, a wide range of datasets extracted from three sources: Scientific Information for Land Owners (SILO) (Jeffrey et al. 2001) and the European Centre for Medium-Range Weather Forecast (ECMWF) models (Dee et al. 2011) for the input variables, and Energex (Energex 2018) for the target variable (*G*) was used. Based on

the eight stations of Burpengary, Ipswich South, Narangba, Redcliffe, Redland Bay, Wynnum, and Zillmere that are located in southeast Queensland, data from 01/07/2014 to 30/06/2017 for the 6-h forecast horizon, and 01/07/2014 to 29/06/ 2017 for the daily forecast horizon were used. The input variables included the six most relevant climate variables from SILO and 51 Reanalysis variables obtained from ECMWF were used to feed the models used in this objective. A complete list of these variables, including their details with other Energex datasets details, are provided in Chapter 5.

For objective 4, the daily electricity demand (G) data were obtained from the USQ, Australia. The data were collected from the three regional university campuses of Toowoomba, Ipswich and Springfield. The historical data were provided by the university campus services for the periods of 01-01-2013 to 31-12-2014 for the main feed of Toowoomba, 01-09-2015 to 31-08-2016 for the main feed and Building A block of Ipswich and Springfield, respectively. The actual data were recorded every 15-minutes (96 times per day) in kilowatts (kW) where the daily data were performed by obtaining the summations of every 15-minute value. Chapter 6 provides more details of these data.

2.4 General Methodology

A number of necessary tasks were applied on the data prior to the model development step. Converting the inputs and the target data into their required forecast horizons and using the partial autocorrelation function (PACF) to select the best statistically significant lags from the target variable (G) were the first two steps used in all objectives of this thesis. Secondly, the data of the second objective were decomposed using an improved version of empirical mode decomposition with adaptive noise (ICEEMDAN) to address the non-stationarity issues associated with the data. The cross-correlation statistic was also adopted between the target (G) and the inputs (SILO and ECMWF data) to select the best input variables in the data of the third objective. For the last objective of this study, a calendar averaging technique and a suitable maximum overlap discrete wavelet transform (MODWT) were applied to fill all the missing values and decompose the input lag datasets into wavelet and scaling coefficients before running the model. Furthermore, the higher frequency data from inputs and target datasets of all study objectives were then normalized between zero and one using Eq. (1) to avoid large numeric ranges from the values of the predictor variables. Finally, the best parameters and boundary conditions of the models

developed in the respective objectives of this study were selected using optimisation and trial-and-error methods. The tasks above were described in detail in their respective chapters.

$$d_{norm} = \frac{d - d_{min}}{d_{max} - d_{min}} \quad (1)$$

where d_{min} and d_{max} are the minimum and maximum values of d , respectively, while d_{norm} is the normalized value of d .

In this thesis, various data intelligent G forecasting models and pre-processing methods were considered to evaluate their abilities to forecast G data over different horizons. The models were multivariate adaptive regression splines (MARS), support vector regression (SVR), autoregressive integrated moving average (ARIMA), M5 model tree, artificial neural network (ANN), multiple linear regression (MLR) and online sequential extreme learning machine (OS-ELM) models. Pre-processing approaches included ICEEMDAN, MODWT, grid search and PSO that were used to handle the non-stationarity features within the inputs and create a hybrid model. These techniques were necessary to increase the forecast accuracy by decomposing the data into low and high pass filters using ICEEMDAN and MODWT and selecting the best input parameters for the SVR model using the grid search and PSO tools. A trial-and-error method was also utilized in this study to select the appropriate parameters of the ARIMA, MARS, ANN and OS-ELM models. Finally, the bootstrap (B) technique was incorporated in this study to create an ensemble model that has a well ability to estimate the forecast uncertainty.

In literature, the MARS model is based on basis functions for each spline and can be developed with linear and cubic regression (Friedman 1991). According to Yu, Chen and Chang (2006), the SVR model, first introduced by (Vapnik 1998), is run by the principle of structural risk minimization (SRM) that aims to decrease the overfitting data by minimizing the expected error of a learning machine. Additionally, the time series ARIMA model, the popular approach for data forecasting, was investigated by the work of Box and Jenkins (1976). Although the MLR model is a statistical technique that examines the cause and effect relationship between objective (G) and predictor variables (Deo & Şahin 2017), the ANN model, a powerful and useful data-driven algorithm, has a better ability to capture and represent the relationships between complex predictor and target variables (Govindaraju 2000;

Abbot & Marohasy 2012; Şahin, Kaya & Uyar 2013; Abbot & Marohasy 2014; Deo & Şahin 2015). In a range of forecasting applications, the OS-ELM model is considered a fast and precise data intelligent approach that can offer better performance compared to other algorithms (Liang et al. 2006; Lan, Soh & Huang 2009). On the other hand, ICEEMDAN, MODWT, PSO and B are robust pre-processing approaches that have been applied in several forecasting studies to improve traditional models, selecting the best parameters for models and generating prediction bands through an ensemble model (Tiwari & Adamowski 2013; Li & Li 2016; Quilty & Adamowski 2018).

For the half-hourly (0.5 h), hourly (1.0 h) and daily (24 h) forecasting horizons presented in Chapter 3 (objective 1), the MARS and SVR models were used and developed against the AEIMA model. In contrast, the two phase ICEEMDAN-PSO-SVR model was adopted and compared with the ICEEMDAN-MARS, ICEEMDAN-M5 model tree, PSO-SVR, MARS, and M5 model tree for the short-term (weekend, working days, whole week, and public holiday), and long-term (monthly) horizons developed in Chapter 4 (objective 2). In Chapter 5 (objective 3) that presents 6.0 hour and daily horizons, the ANN and hybrid ANN models were developed versus the MARS, MLR and ARIMA models. Finally, the MODWT-PACF-OS-ELM (MPOE) model was introduced in comparison with its traditional model of PACF-OS-ELM (POE) for the daily horizon in Chapter 6 (objective 4). Figure 2.1 classifies the models above into their respective chapters while the details of the theoretical backgrounds and methods developments were presented in each related chapter. To sum up, the novel and main hybrid models developed in this study were:

- 1- The MARS model for 30 minute and hourly forecast horizons and the SVR model for the daily forecast horizon presented in Chapter 3. The PACF technique was used to select the significant lag (input) variables from the target variable (G).
- 2- The two phase ICEEMDAN-PSO-SVR for multiple forecasting horizons including whole weeks (daily), weekends, working days, public holidays, and monthly. The ICEEMDAN was adopted to decompose the target data (G) into intrinsic mode functions (IMF) and a residual component while the PSO was used to select the best parameters of the SVR model. The details of the model development steps are clearly shown in Fig. 3 (Chapter 4).

- 3- The ANN and hybrid ANN models integrated with extensive climate and atmospheric datasets for 6-hour and daily G forecasting horizons. A new and novel approach for energy demand forecasting was presented using this method in Chapter 5 where a large number of ANN models were developed.
- 4- The MODWT-PACF-OS-ELM (MPOE) model for the daily forecasting horizon. This model was introduced in Chapter 6 to address objective 4 of this study where the PACF was applied to select the significant (lagged) input variables whereas the MODWT algorithm was utilized before running the OS-ELM model to address non-stationarity problems by decomposing these lag variables into wavelet and scaling coefficients.

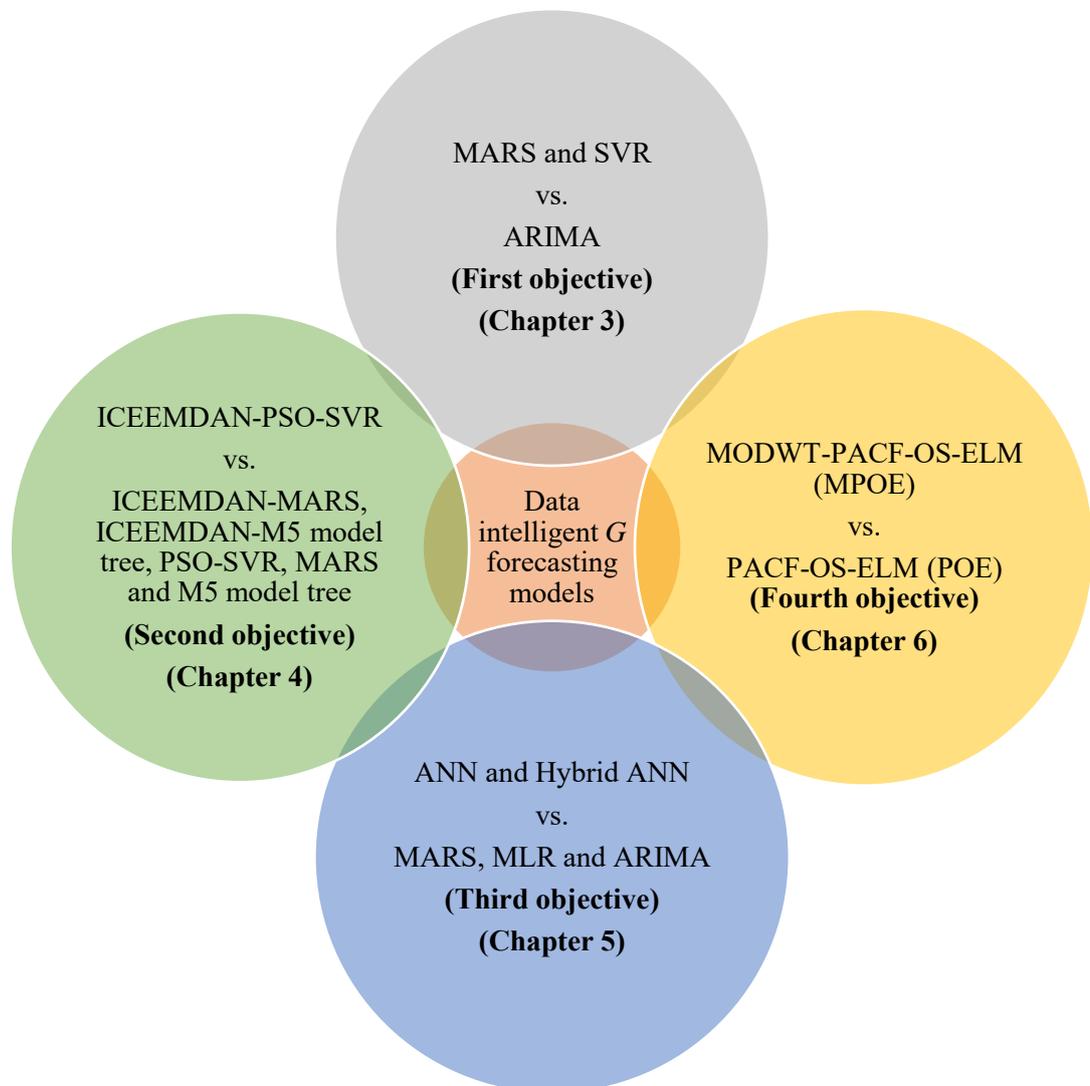


Figure 2.1: Types of data intelligent models developed in each chapter (objective) in this doctoral thesis.

In this doctoral thesis, a wide range of statistical criteria including correlation coefficient (r), root-mean square error ($RMSE$), mean absolute error (MAE), relative root-mean square error ($RRMSE\%$ or $RMSE_{\bar{G}}\%$), and relative mean absolute error ($MAPE\%$ or $MAE_{\bar{G}}\%$), Willmott's Index (WI), Nash–Sutcliffe efficiency coefficient (E_{NS}) and Legates and McCabe's Index (E_{LM} or LM) were employed to evaluate the performance of the models to forecast G data. The details and mathematical equations for these statistical indices are shown in each chapter in this thesis. Additionally, several plots including empirical cumulative distribution function (ECDF), box plots, scatter diagram, histogram, time series plot, relative error analysis plot, bar graphs, and Taylor plots were also introduced to assess the ability of the models developed in this study for G forecasting.

Chapter 3: Short-Term Electricity Demand Forecasting Using Queensland's Aggregated Data

3.1 Foreword

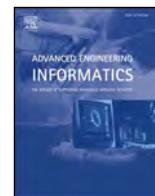
This chapter presents an exact copy of the published article in *Advanced Engineering Informatics* journal (Vol. 35, Pages 1-16).

In this work, the candidate develops and discuss the application of three data intelligent models of multivariate adaptive regression splines (MARS), support vector regression (SVR), and autoregressive integrated moving average (ARIMA) on 30 minutes hourly, and daily forecast horizons. Electricity demand (G) data for whole Queensland, Australia are used to develop and evaluate the models. The partial autocorrelation function (PACF) is adopted to historical G data to select the best inputs for the MARS and SVR while a single input from original G data is used to develop the univariate ARIMA model. A novel contribution and advancement of electricity demand forecasting in Queensland, Australia has been introduced in this chapter by applying the SVR, MARS and ARIMA approaches that have not been used in a forecasting study published before this paper for energy demand in Queensland.

3.2 Research Highlights

- Short-term (0.5, 1.0 and 24h) energy demand is forecasted with MARS, SVR and ARIMA model in Queensland.
- For very short-term forecasting horizon (0.5 & 1.0h), MARS outperforms SVR and ARIMA.
- SVR performs better than MARS and ARIMA for daily (24 h) forecasting horizon.
- SVR and MARS are useful tools for real-time forecasting of energy demand.

3.3 Published Article I



Full length article

Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia



Mohanad S. Al-Musaylh^{a,b,*}, Ravinesh C. Deo^{a,d,*}, Jan F. Adamowski^c, Yan Li^a

^a School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, QLD 4350, Australia

^b Management Technical College, Southern Technical University, Basrah, Iraq

^c Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Québec H9X 3V9, Canada

^d Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China

ARTICLE INFO

Keywords:

Electricity demand forecasting
Machine learning
SVR
MARS
ARIMA

ABSTRACT

Accurate and reliable forecasting models for electricity demand (G) are critical in engineering applications. They assist renewable and conventional energy engineers, electricity providers, end-users, and government entities in addressing energy sustainability challenges for the National Electricity Market (NEM) in Australia, including the expansion of distribution networks, energy pricing, and policy development. In this study, data-driven techniques for forecasting short-term (24-h) G -data are adopted using 0.5 h, 1.0 h, and 24 h forecasting horizons. These techniques are based on the Multivariate Adaptive Regression Spline (MARS), Support Vector Regression (SVR), and Autoregressive Integrated Moving Average (ARIMA) models. This study is focused in Queensland, Australia's second largest state, where end-user demand for energy continues to increase. To determine the MARS and SVR model inputs, the partial autocorrelation function is applied to historical (area aggregated) G data in the training period to discriminate the significant (lagged) inputs. On the other hand, single input G data is used to develop the univariate ARIMA model. The predictors are based on statistically significant lagged inputs and partitioned into training (80%) and testing (20%) subsets to construct the forecasting models. The accuracy of the G forecasts, with respect to the measured G data, is assessed using statistical metrics such as the Pearson Product-Moment Correlation coefficient (r), Root Mean Square Error ($RMSE$), and Mean Absolute Error (MAE). Normalized model assessment metrics based on $RMSE$ and MAE relative to observed means ($RMSE_{\bar{G}}$ and $MAE_{\bar{G}}$), Willmott's Index (WI), Legates and McCabe Index (E_{LM}), and Nash–Sutcliffe coefficients (E_{NS}) are also utilised to assess the models' preciseness. For the 0.5 h and 1.0 h short-term forecasting horizons, the MARS model outperforms the SVR and ARIMA models displaying the largest WI (0.993 and 0.990) and lowest MAE (45.363 and 86.502 MW), respectively. In contrast, the SVR model is superior to the MARS and ARIMA models for the daily (24 h) forecasting horizon demonstrating a greater WI (0.890) and MAE (162.363 MW). Therefore, the MARS and SVR models can be considered more suitable for short-term G forecasting in Queensland, Australia, when compared to the ARIMA model. Accordingly, they are useful scientific tools for further exploration of real-time electricity demand data forecasting.

Abbreviations: MW, Megawatt; G , Electricity load (demand, Mega Watts); MARS, Multivariate Adaptive Regression Splines; SVR, Support Vector Regression; ARIMA, Autoregressive Integrated Moving Average; r , Correlation Coefficient; $RMSE$, Root Mean Square Error (MW); MAE , Mean Absolute Error (MW); $RMSE_{\bar{G}}$, Relative Root Mean Square Error, %; $MAE_{\bar{G}}$, Mean Absolute Percentage Error, %; WI , Willmott's Index of Agreement; E_{NS} , Nash–Sutcliffe Coefficient; E_{LM} , Legates and McCabe Index; ANN, Artificial Neural Network; RBF, Radial Basis Function for SVR; σ , Kernel Width for SVR Model; C , Regularization for SVR Model; $BF_m(X)$, Spline Basis Function for MARS; GCV , Generalized Cross-Validation; p , Autoregressive Term in ARIMA; D , Degree of Differencing in ARIMA; Q , Moving Average Term in ARIMA; AEMO, Australian Energy Market Operator; NEM, National Electricity Market; ACF, Auto-Correlation Function; PACF, Partial Auto-Correlation Function; MSE , Mean Square Error (MW); R^2 , Coefficient of Determination; AIC , Akaike Information Criterion; L , Log Likelihood; σ^2 , Variance; G_t^{for} , t^{th} Forecasted Value of G , MW; G_t^{obs} , t^{th} Observed Value of G , MW; Q_{25} , Lower Quartile (25th Percentile); Q_{50} , Median Quartile (50th Percentile); Q_{75} , Upper Quartile (75th Percentile); d , Degree of Differencing in ARIMA

* Corresponding authors at: School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, QLD 4350, Australia.

E-mail addresses: MohanadShakirKhalid.AL-Musaylh@usq.edu.au, mohanadk21@gmail.com (M.S. Al-Musaylh), ravinesh.deo@usq.edu.au (R.C. Deo).

<https://doi.org/10.1016/j.aei.2017.11.002>

Received 8 April 2017; Received in revised form 18 November 2017; Accepted 20 November 2017

Available online 01 December 2017

1474-0346/ © 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Electricity load forecasting (also referred to as demand and abbreviated as G in this paper, MW) plays an important role in the design of power distribution systems [1,2]. Forecast models are essential for the operation of energy utilities as they influence load switching and power grid management decisions in response to changes in consumers' needs [3]. G forecasts are also valuable for institutions related to the fields of energy generation, transmission, and marketing. The precision of G estimates is critical since a 1% rise in load forecasting error can lead to a loss of millions of dollars [4–6]. Over- or under-projections of G can endanger the development of coherent energy policies and hinder the sustainable operation of a healthy energy market [7]. Furthermore, demographic, climatic, social, recreational, and seasonal factors can impact the accuracy of G estimates [1,8,9]. Therefore, robust forecasting models that can address engineering challenges, such as minimizing predictive inaccuracy in G data forecasting, are needed to, for example, support the sustainable operation of the National Electricity Market (NEM).

Qualitative and quantitative decision-support tools have been useful in G forecasting. Qualitative techniques, including the Delphi curve fitting method and other technological comparisons [6,10,11], accumulate experience in terms of real energy usage to achieve a consensus from different disciplines regarding future demand. On the other hand, quantitative energy forecasting is often applied through physics-based and data-driven (or black box) models that draw upon the inputs related to the antecedent changes in G data. The models' significant computational power has led to a rise in their adoption [12]. Data-driven models, in particular, have the ability to accurately forecast G , which is considered a challenging task [6]. Having achieved a significant level of accuracy, data-driven models have been widely adopted in energy demand forecasting (e.g., [13,14]). Autoregressive Integrated Moving Average (ARIMA) [15], Artificial Neural Network (ANN) [16], Support Vector Regression (SVR) [17], genetic algorithms, fuzzy logic, knowledge-based expert systems [18], and Multivariate Adaptive Regression Splines (MARS) [19] are among the popular forecasting tools used by energy researchers.

The SVR model, utilised as a primary model in this study, is governed by regularization networks for feature extraction. The SVR model does not require iterative tuning of model parameters [20,21]. Its algorithm is based on the structural risk minimization (SRM) principle and aims to reduce overfitting data by minimizing the expected error of a learning machine [21]. In the last decades, this technique has been recognized and applied throughout engineering, including in forecasting (or regression analysis), decision-making (or classification works) processes and real-life engineering problems [22]. Additionally, the SVR models have been shown to be powerful tools when a time-series (e.g., G) needs to be forecasted using a matrix of multiple predictors. As a result, their applications have continued to grow in the energy forecasting field. For example, in Turkey (Istanbul), several investigators have used the SVR model with a radial Basis Kernel Function (RBF) to forecast G data [23]. In eastern Saudi Arabia, the SVR model generated more accurate hourly G forecasts than a baseline autoregressive (AR) model [24]. In addition, different SVR models were applied by Sivapragasam and Liong [25] in Taiwan to forecast daily loads in high, medium, and low regions. In their study, the SVR model provided better predictive performance than an ANN approach for forecasting regional electric loads [29]. Except for one study that confirmed SVR models' ability to forecast global solar radiation [17], to the best of the authors' knowledge, a robust SVR forecasting model has been limitedly applied for energy demand. Thus, additional studies are needed to explore SVR modelling in comparison to other models applied in G forecasting.

Contrary to the SVR model, the MARS model has not been widely tested for G forecasting. It is designed to adopt piecewise (linear or cubic) basis functions [26,27]. In general, the model is a fast and flexible statistical tool that operates through an integrated linear and non-linear modelling approach [28]. More importantly, it has the capability of

employing a set of basic functions using several predictor variables to assess their relationship with the objective variable through non-linear and multi-collinear analysis. This is important for demand forecasting based on interactions between different variables and the demand data. Although the literature on MARS models applied in the field of G forecasting is very scarce, this model has proven to be highly accurate in several estimation engineering challenges. Examples may be drawn from studies that discuss doweled pavement performance modelling, determination of ultimate capacity of driven piles in cohesionless soil, and analysis of geotechnical engineering systems [29–31]. In Ontario (Canada), the MARS model was applied, through a semiparametric approach, for forecasting short-term oil prices [32] and investigating the behaviour of short-term (hourly) energy price (HOEP) data through lagged input combinations [8]. Sigauke and Chikobvu [19] tested the MARS model for G forecasting in South Africa; this demonstrated its capability of yielding a significantly lower Root Mean Square Error (RMSE) when compared to piecewise regression-based models. However, despite its growing global applicability (e.g., [26,27,33–35]), the MARS model remains to be explored for G forecasting in the present study region.

In the literature, the ARIMA model has generated satisfactory results for engineering challenges including the forecasting of electricity load data [15], oil [32], and gas demand [36]. A study in Turkey applied a cointegration method with an ARIMA model for G -estimation and compared results with official projections. It concluded that approximately 34% of the load was overestimated when compared to measured data from the ARIMA model [8]. Several studies have indicated that the ARIMA model tends to generate large errors for long-range forecasting horizons. For example, a comparison of the ARIMA model, the hybrid Grey Model (GM-ARIMA), and the Grey Model (GM(1, 1)) for forecasting G in China showed that GM (1, 1) outperformed the ARIMA model [37]. Similarly, a univariate ARAR model (i.e., a modified version of the ARIMA model) outperformed a classical ARIMA model in Malaysia [38]. However, to the best of the authors' knowledge, a comparison of the MARS, SVR, and ARIMA methods, each having their own merits and weaknesses, has not been undertaken in the field of G forecasting.

To explore opportunities in G forecasting, this paper discusses the versatility of data-driven techniques (multivariate MARS and SVR models and the univariate ARIMA model) for short-term half-hourly (0.5 h), hourly (1.0 h) and daily (24 h) horizon data. The study is beneficial to the field of power systems engineering and management since energy usage in Queensland continues to face significant challenges, particularly as it represents a large fraction (i.e., 23%) of the national 2012–2013 averaged energy demand [39]. The objectives of the study are as follows: (1) To develop and optimise the MARS, SVR, and ARIMA models for G forecasting using lagged combinations of the state-aggregated G data as the predictor variable; (2) To validate the optimal MARS, SVR, and ARIMA models for their ability to generate G forecasts at multiple forecasting horizons (i.e., 0.5, 1.0 and 24 h); and (3) To evaluate the models' preciseness over a recent period, [01-01-2012 to 31-12-2015 (dd-mm-yyyy)], by employing robust statistical metrics comparing forecasted and observed G data obtained from the Australian Energy Market Operator (AEMO) [40]. To evaluate and reach these objectives, this paper is divided into the following sections: Section 2 describes the theory of SVR, MARS, and ARIMA models; Section 3 presents the materials and methods including the G data and model development and evaluation; Section 4 presents the results and discussion; and Section 5 further discusses the results, research opportunities, and limitations. The final section summarizes the research findings and key considerations for future work.

2. Theoretical background

2.1. Support Vector regression

An SVR model can provide solutions to regression problems with multiple predictors $X = \{x_i\}_{i=1}^n$, where n is the number of predictor

variables and each x_i has N variables. These are linked to an objective variable $y = \{y_i\}_{i=1}^N$. The matrix X is converted to a higher-dimensional feature space, in accordance with the original, but constitutes a lower-dimensional input space [41,42]. With an SVR model, a non-linear regression problem is defined as [43]:

$$y = f(X) = \omega \cdot \varnothing(X) + b \tag{1}$$

where b is a constant, ω is the weighted vector, and $\varnothing(X)$ denotes the mapping function employed in the feature space. The coefficients ω and b are estimated by the minimisation process below [43]:

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi_i^*) \tag{2}$$

$$\text{Subject to } \begin{cases} |y_i - (w \cdot x_i + b)| \geq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{3}$$

where C and ε are the model's prescribed parameters. The term of $\frac{1}{2} \|w\|^2$ measures the smoothness of the function and C evaluates the trade-off between the empirical risk and smoothness. ξ and ξ^* are positive slack variables representing the distance between actual and corresponding boundary values in the ε -tube model of function approximation.

After applying Lagrangian multipliers and optimising conditions, a non-linear regression function is obtained [43]:

$$f(X) = \sum_{i=1}^{i=N} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \tag{4}$$

where α_i and α_i^* are Lagrangian multipliers and the term $K(x_i, x_j)$ is the kernel function describing the inner product in D -dimensional feature space, x_i and $x_j \in X$ [43]. Under Kuhn-Tucker conditions, a limited number of α_i and α_i^* coefficients will be non-zero [17]. The associated data points, termed the "support vectors", lie the closest to the decision surface (or hyperplane) [17]. The radial basis function (RBF) employed in developing the SVR model in this study, can be expressed as [44]:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{5}$$

where x_i and x_j are the inputs in the i^{th} and j^{th} respective dimensions and σ is the kernel width. Over the training period, the support vectors' area of influence with respect to input data space is determined by kernel width (σ) and regulation (C). Deducing these can represent a critical task for achieving superior model accuracy [17]. This is performed through a grid-search procedure (Section 3.2).

2.2. Multivariate adaptive regression splines

The MARS model, first introduced by Friedman [28], implements the piecewise regression process for feature identification of the input dataset. In addition, it has the capability to flexibly and efficiently analyse the relationships between a given predictand (i.e., the G in context of the present study) and a set of predictor variables (i.e., the lagged combinations of G). In general, the MARS model can analyse non-linearities in predictor-predictand relationships when forecasting a given predictand [45].

Assuming two variable matrices, X and y , where X is a matrix of descriptive variables (predictors) over a domain $D \subset \mathbb{R}^n$, $X = \{x_i\}_{i=1}^n$, and y is a target variable (predictand), there are then N realizations of the process $\{y_i, x_{i1}, x_{i2}, \dots, x_{in}\}_1^N$ [8]. Consequently, the MARS model relationship between X and y is demonstrated below [28]:

$$y = \hat{f}(X) = a_0 + \sum_{m=1}^M a_m BF_m \tag{6}$$

where a_0 is a constant, $\{a_m\}_1^M$ are the model coefficients estimated to produce data-relevant results, M is the number of subregions $R_m \subset D$ or

the equivalent basis functions in MARS, and $BF_m(X)$ is a spline function defined as $C(X|s, t_1, t_2)$. In the latter, $t_1 < t < t_2$, and s have a value of $+1$ or -1 for a spline basis function or its mirror image, respectively.

The Generalized Cross-Validation criterion (GCV) used by the MARS model assesses the lack-of-fit of the basis functions through the Mean Square Error (MSE) [28] and is expressed as:

$$GCV = MSE / \left[1 - \frac{\tilde{G}(M)}{N}\right]^2 \tag{7}$$

where $MSE = \frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}(X_i)]^2$ and $\left[1 - \frac{\tilde{G}(M)}{N}\right]^2$ is a penalty that accounts for an increasing variance from a complex model. Furthermore, $\tilde{G}(M)$ is defined as [28]:

$$\tilde{G}(M) = C(M) + v \cdot M \tag{8}$$

where v is a penalty factor with a characteristic value of $v = 3$ and $C(M)$ is the number of parameters being fitted.

The MARS model with the lowest value of the GCV for the training dataset is considered the optimal model.

2.3. Autoregressive integrated moving average

Relying on the antecedent data to forecast G , the ARIMA model constitutes a simplistic, yet popular approach applied for time-series forecasting. ARIMA was popularized by the work of Box and Jenkins [46]. To develop the ARIMA model, two types of linear-regressions are integrated: the Autoregressive (AR) and the Moving Average (MA) [46].

The AR model is written as [46]:

$$y_t = c + a_1 y_{t-1} + \dots + a_p y_{t-p} + u_t \tag{9}$$

where a_1, \dots, a_p are the AR parameters, c is a constant, p is the order of the AR, and u_t is the white noise.

Likewise, the MA model can be written as [46]:

$$y_t = \mu + u_t + m_1 u_{t-1} + \dots + m_q u_{t-q} \tag{10}$$

where m_1, \dots, m_q are the MA parameters, q is the order of MA, $u_t, u_{t-1}, \dots, u_{t-q}$ are the white noise (error) terms, and μ is the expectation of y_t .

By integrating these models with the same training data, the ARIMA model [ARIMA (p, q)] becomes [46]:

$$y_t = c + a_1 y_{t-1} + \dots + a_p y_{t-p} + u_t + m_1 u_{t-1} + \dots + m_q u_{t-q} \tag{11}$$

where p and q are the autoregressive and moving average terms, respectively.

The basic premise of this model is that time-series data incorporates statistical stationarity which implies that measured statistical properties, such as the mean, variance, and autocorrelation remain constant over time [47]. However, if the training data displays non-stationarity, as is the case with real-life predictor signals (e.g., G data), the ARIMA model requires differenced data to transform it to stationarity. This is denoted as ARIMA (p, d, q) where d is the degree of differencing [37].

3. Materials and methods

3.1. Electricity demand data

In this study, a suite of data-driven models was developed for short-term G forecasting in Queensland, Australia. The predictor data, comprised of half-hourly (48 times per day) G records for a period between 01-01-2012 to 31-12-2015 (dd-mm-yyyy), was acquired from the Australian Energy Market Operator (AEMO) [40]. The AEMO database aims to provide G data, in terms of relevant energy consumption, for the Queensland region of the NEM. Hence, these data have been previously used in various forecasting applications (e.g., [48,49]). However, they have not been employed in machine learning models as attempted in the present study.

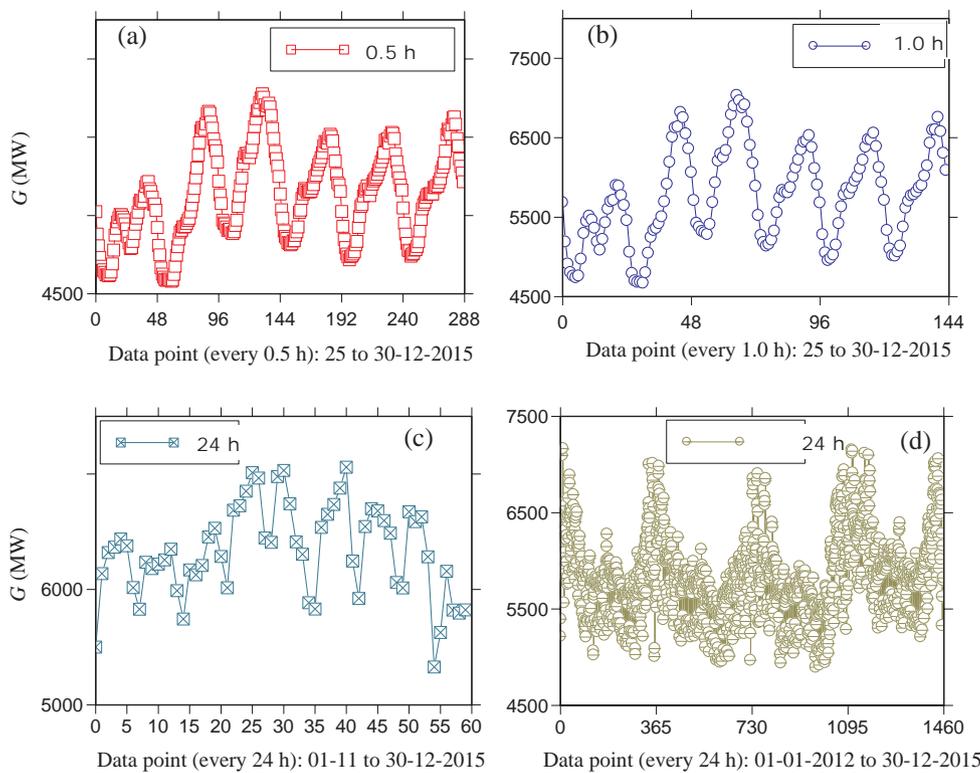


Fig. 1. Time-series of electricity demand (G) data and various forecasting periods.

In the present study, the 0.5 h time-step corresponds to the NEM settlement periods 1 (0:00 h–0:30 h) through 48 (23:30 h–24:00 h). The 0.5 h interval readings, reported in other research works (e.g., [48,49]), were thus used for short-term forecasting of the G data. To expand the forecasting horizon to 1.0 h and 24 h periods to obtain G values, an arithmetic averaging of the half-hourly data was performed. The MARS, SVR, and ARIMA models considered in this paper, developed and evaluated 0.5 h, 1.0 h and 24 h forecasts utilising data from periods 01-12-2015 to 31-12-2015, 01-11-2015 to 31-12-2015, and 01-01-2012 to 31-12-2015, respectively. In principle, the number of predictive features remained similar throughout (i.e., approximately 1460 data points for each horizon).

Fig. 1(a–d) depicts plots of the aggregated G data for the Queensland region, whereas Table 1 provides its associated descriptive statistics. The stochastic components, present in G data at the 0.5 h and 1.0 h time-scales, exhibit fluctuations due to the change in consumer electricity demands. This is confirmed by the large standard deviation and high degree of skewness observed for the 0.5 h and 1.0 h scale when compared to those associated with the 24 h scale in Table 1.

3.2. Forecast model development

Data-driven models incorporate historical G data to forecast future G values. The initial selection of (lagged) input variables to determine the predictors is critical for developing a robust multivariate (SVR or MARS) model [17,26]. The literature outlines two input selection methods for determining the sequential time series of lagged G values that provide an optimal performance. These are (i) trial and error and (ii) an auto-correlation function (ACF) or partial auto-correlation

function (PACF) approach. For this study, patterns were analysed in historical G data from the training period using the ACF and PACF to extract correlation statistics [50–52]. This approach employed time-lagged information to analyse the period between current and antecedent G values at specific points in the past (i.e., applying a time lag) and assessed any temporal dependencies existing in the time-series. Subsequently, inputs for each time lag (0.5 h, 1.0 h, 24 h) were identified by statistical verification of lagged G combinations and their respective correlation coefficient (r).

The PACF for G data, depicted in Fig. 2, aided in identifying potential inputs for data-driven models. The method computed a time-series regression against its n -lagged-in-time values that removed the dependency on intermediate elements and identified the extent to which G was correlated to the antecedent timescale value. Consequently, the statistically correlated signal $G(t)$ and the respective n -lagged signals were selected. This procedure developed forecast models that considered the role of memory (i.e., antecedent G) in forecasting the current G . The 15 modelling scenarios, presented in Table 2, were developed based on the MARS and SVR algorithms.

For the 0.5 h and 24 h forecasting horizons, the models employed half-hourly and daily data from the 1-12-2015 to 31-12-2015 (≈ 1488 data points) and 1-1-2012 to 31-12-2015 (≈ 1461 data points) time periods, respectively. The MARS and SVR models were built with 1–3 statistically significant lagged input combinations (3 representing the maximum number of lags of significant G data) and denoted as T_1, T_2 and T_3 for 0.5 h, and D_1, D_2 and D_3 for 24 h, respectively. Similarly, the 1.0 h forecasting horizon for the MARS and SVR models were constructed from data over the period 1-11-2015 to 31-12-2015 (≈ 1464 data points), built with 1–6 statistically significant lagged input

Table 1
Descriptive statistics of the electricity demand (G) (MW) data aggregated for the Queensland (QLD) study region.

Forecast horizon (h)	Data Period (dd-mm-yyyy)	Minimum (MW)	Maximum (MW)	Mean (MW)	Standard deviation (MW)	Skewness	Flatness
0.5	01-12 to 31-12-2015	4660.55	8402.56	6318.42	802.67	0.17	−0.85
1.0	01-11 to 31-12-2015	4668.66	8393.81	6323.48	806.06	0.11	−0.83
24	01-01-2012 to 31-12-2015	4896.05	7165.54	5827.85	414.81	0.54	0.36

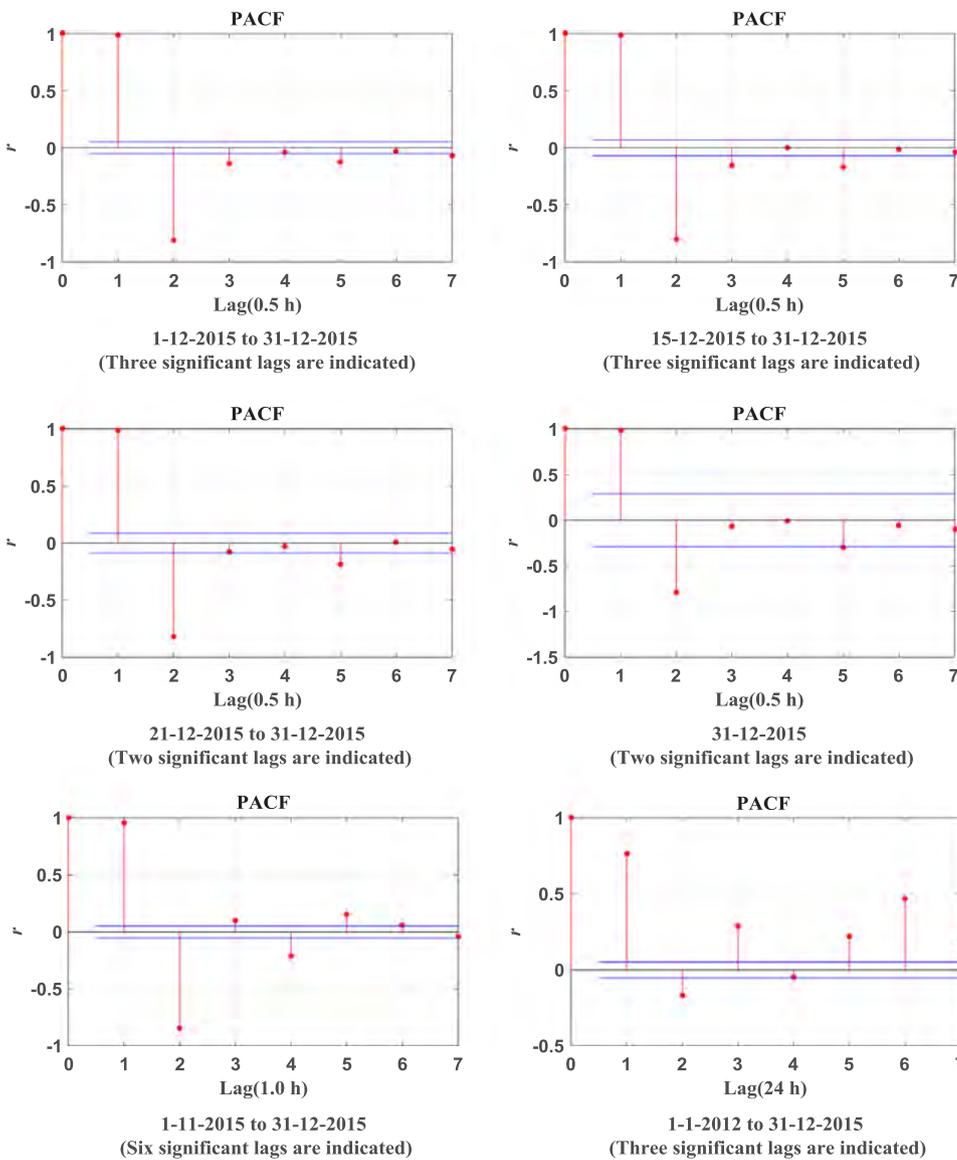


Fig. 2. Correlation coefficient (r) based on the partial autocorrelation function (PACF) of predictors (*i.e.*, electricity demand, G) used for developing the support vector regression (SVR) and multivariate regression splines (MARS) models. Statistically significant lags at the 95% confidence interval are marked (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

combinations (6 representing the maximum number of significant lagged G values), and denoted as H_1, \dots, H_6 , respectively.

To determine the effect of data length, the short-term (0.5 h) forecasting horizon scenario was studied using data from the 15-12-2015 to 31-12-2015 period for the SVR and MARS models. A total of 817 data points with 1–3 statistically significant lags were applied and denoted as the T^a model. Furthermore, the T^b and T^c models used data from period 21-12-2015 to 31-12-2015 and single-day data for 31-12-2015 which consisted of 529 data points and 48 data points with 1 or 2 statistically significant lags, respectively.

On the other hand, the univariate ARIMA model’s mechanism differs as it creates its own lagged data through the p and q parameters developed in its identification phase seen in Table 3. Therefore, all historical G data were used as a single input (with no lags) to identify the ARIMA model for all forecasting horizons.

Table 2 and Fig. 2 contain further details regarding the forecast models and their nominal designation. It should be noted that for the baseline models, the input variables had a total of 1461–1488 data points.

There is no single method for dividing training and evaluation data [17]. To deduce optimal models for G forecasting, data were split into subsets as follows: 80% for training and 20% for evaluation (testing). Given the chaotic nature of the input where changes in G seem to occur at a higher frequency, the trained data required appropriate scaling to

avoid predictor values (and associated patterns/attributes) with large numeric ranges from dominating attributes with narrower ones [53,54]. Data were therefore normalized and bounded by zero and one through the following expression [17]:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (12)$$

where x is any given data value (input or target), x_{min} is the minimum value of x , x_{max} is the maximum value of x , and x_{norm} is the normalized value of the data.

The SVR models were developed by the MATLAB-based Libsvm toolbox (version 3.1.2) [55]. The RBF (Eq. (5)) was used to map non-linear input samples onto a high dimensional feature space because it examines the non-linearities between target and input data [53,54] and outperforms linear-kernel-based models in terms of accuracy [42,56]. The RBF is also faster in the training phase [57,58] as demonstrated in [41]. An alternative linear kernel is a special case of the RBF [56], whereas the sigmoid kernel behaves as the RBF kernel for some model parameters [54].

Furthermore, the selection of C and σ values is crucial to obtain an accurate model [59]. For this reason, a grid search procedure, over a wide range of values seeking the smallest MSE , was used to establish the optimal parameters [53]. Fig. 3(a) illustrates a surface plot of the MSE

Table 2
Model designation for the MARS, SVR and ARIMA for 0.5 h, 1.0 h and 24 h forecast horizons.

Model	Period of <i>G</i> data studied (dd-mm-yyyy) No. data points						No. significant lags (* = all lags)					
	1-1-2012 to 31-12-2015	1-11-2015 to 31-12-2015	31-12-2015	21-12-2015 to 31-12-2015	15-12-2015 to 31-12-2015	1-12-2015 to 31-12-2015	1	2	3	4	5	6
	1461	1464	48	529	817	1488						
<i>Half-hourly (0.5 h) forecast horizon</i>												
MARS and SVR	T_1							×		×		
	T_2							×		×		
	T_3							×			×	*
	T^a					×					×	*
	T^b			×						×	*	
	T^c			×						×	*	
	ARIMA							×				
	ARIMA ^a				×							
	ARIMA ^b			×								
	ARIMA ^c			×								
<i>Hourly (1.0 h) forecast horizon</i>												
MARS and SVR	H_1	×								×		
	H_2	×								×		
	H_3	×									×	
	H_4	×										×
	H_5	×										×
	H_6	×										×
	ARIMA	×										×
<i>Daily (24 h) forecast horizon</i>												
MARS and SVR	D_1	×								×		
	D_2	×									×	
	D_3	×										×
	ARIMA	×										×

with respect to different regularisation constants *C* and σ (kernel width) values for the SVR model used in 1.0 h forecasting. In this case, the optimal model H_4 attained an *MSE* of $\cong 0.0001 \text{ MW}^2$ for $C = 1.00$ and $\sigma = 48.50$. Table 3 lists the optimal values of *C* and σ that are unique to each SVR model.

Alternatively, the MARS model adopted the MATLAB-based ARESLab toolbox (version 1.13.0) [60]. Two types of MARS models are possible and employ cubic or linear piecewise formula as their basis functions. In this study, a piecewise cubic model was adopted since it

provided a smoother response in comparison to a linear function [61]. Moreover, generalized recursive partitioning regression was adopted for function approximation given its capacity to handle multiple predictors [8]. Optimisation operated in two phases: forward selection and backward deletion. In the forward phase, the algorithm ran with an initial ‘naïve’ model consisting of only the intercept term. It iteratively added the reflected pair(s) of basis functions to yield the largest reduction in training the *MSE*. The forward phase was executed until one of the following conditions was satisfied [62]:

Table 3
Parameters for the SVR and ARIMA model presented in the training period for 0.5 h, 1.0 h and 24 h forecast horizons.

SVR [*]	<i>C</i>	σ	<i>MSE</i> (MW^2)	ARIMA ^{**}	<i>p</i>	<i>d</i>	<i>q</i>	R^2	σ^2	<i>L</i>	<i>AIC</i>	<i>RMSE</i> (MW)	<i>MAPE</i> (%)
<i>0.5 h Forecast horizon</i>													
T_1	0.19	256.0	0.0012										
T_2	1.74	256.0	0.0004										
T_3	1.00	147.0	0.0004	ARIMA	2	1	6	0.993	4966	-6738.5	13494.9	70.44	0.829
T^a	1.00	84.5	0.0005	ARIMA ^a	5	1	6	0.993	4042	-3623.2	7270.4	63.53	0.785
T^b	0.57	147.0	0.0004	ARIMA ^b	6	1	3	0.994	3553	-2319.6	4659.2	59.54	0.768
T^c	1.00	9.2	0.0011	ARIMA ^c	6	0	1	0.991	2170	-203.7	425.3	46.59	0.660
<i>1.0 h Forecast horizon</i>													
H_1	0.19	147.0	0.0041										
H_2	0.57	256.0	0.0010										
H_3	0.33	147.0	0.0010										
H_4	1.00	48.5	0.0001										
H_5	0.57	48.5	0.0008										
H_6	0.33	27.9	0.0007	ARIMA	5	1	5	0.981	12613	-7159.7	14341.2	112.26	1.366
<i>24 h Forecast horizon</i>													
D_1	0.06	3.0	0.0134										
D_2	0.19	27.9	0.0122										
D_3	0.33	27.9	0.0093	ARIMA	8	1	3	0.805	34015	-7736.7	15497.3	184.35	2.298

* *C* = cost function, σ = kernel width.
** *d* = degree of differencing, *p* = autoregressive term, *q* = moving average term, R^2 = coefficient of determination, σ^2 = variance, *L* = log likelihood, *AIC* = Akaike information criterion, *MAPE* = mean absolute percentage error, *RMSE* = root mean square error.

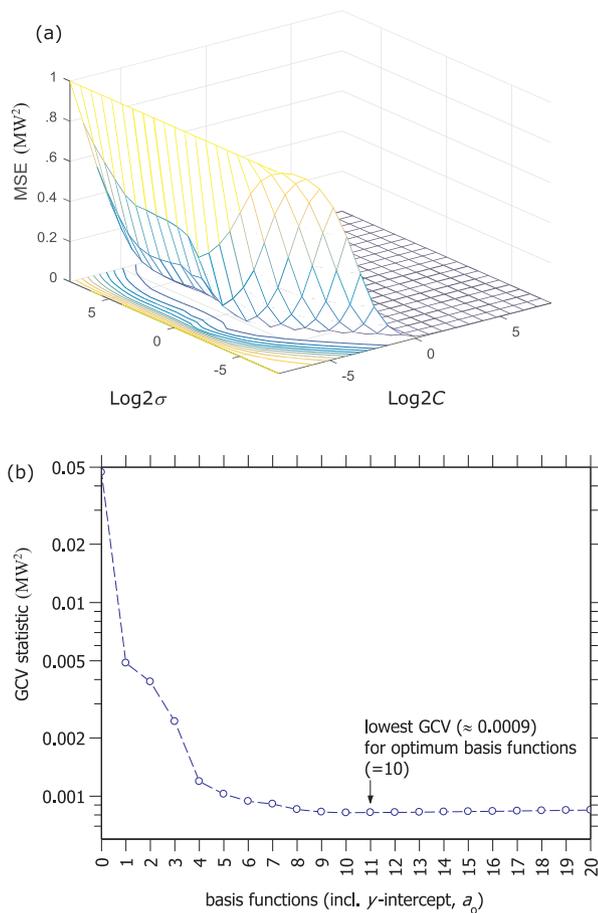


Fig. 3. Illustration of SVR and MARS model parameters for 1.0 h forecast horizon, (H_4) model.

Table 4

The MARS model forecast equation, $y = a_0 + \sum_{m=1}^M a_m BF_m(X)$ with optimum basis functions (BF_m), and generalized cross validation statistic (GCV) in MW^2 for all horizons, in the training period.

MARS model	Model Equation	Opt. Basis Functions	GCV (MW^2)
0.5 h Forecast horizon			
T_1	$y = 0.461 + 0.992BF_1 - 0.984BF_2$	2	0.00109
T_2	$y = 0.456 + 1.67BF_1 - 1.911BF_2 - 0.681BF_3 + 0.944BF_4$	4	0.00037
T_3	$y = 0.480 + 1.587BF_1 - 1.834BF_2 - 0.484BF_3 + 0.790BF_4 - 0.104BF_5$	5	0.00036
T^a	$y = 1.251 + 1.475BF_1 - 1.641BF_2 - 0.481BF_3 + 0.525BF_4 - 0.176BF_5 + 0.110BF_6$	6	0.00043
T^b	$y = 1.035 + 1.711BF_1 - 1.806BF_2 - 0.791BF_3 + 0.834BF_4$	4	0.00038
T^c	$y = 0.656 - 1.689BF_1 + 0.857BF_2 + 0.785BF_3$	3	0.00168
1.0 h Forecast horizon			
H_1	$y = 0.236 + 0.47BF_1 + 1.784BF_2 - 1.453BF_3 - 0.849BF_4$	4	0.0039
H_2	$y = 0.139 + 0.277BF_1 - 0.837BF_2 + 1.319BF_3 + 1.538BF_4 - 2.171BF_5 - 0.324BF_6$	6	0.0010
H_3	$y = 0.926 + 3.131BF_1 - 3.417BF_2 + 1.011BF_3 - 0.339BF_4 - 1.616BF_5 + 2.092BF_6 + 0.201BF_7 - 0.465BF_8 + 2.773BF_9 - 1.417BF_{10} - 1.829BF_{11}$	11	0.0009
H_4	$y = -0.144 + 0.537BF_1 + 1.702BF_2 - 2.298BF_3 - 0.194BF_4 - 0.434BF_5 + 0.437BF_6 + 0.063BF_7 - 0.749BF_8 + 0.896BF_9 - 0.261BF_{10}$	10	0.0009
H_5	$y = -0.021 + 0.010BF_1 - 0.932BF_2 + 1.371BF_3 + 0.618BF_4 - 0.771BF_5 + 1.707BF_6 - 2.332BF_7 + 0.522BF_8 - 0.213BF_9 - 0.544BF_{10} + 0.314BF_{11} + 0.016BF_{12} + 0.113BF_{13} - 0.555BF_{14}$	14	0.0009
H_6	$y = 0.686 + 2.418BF_1 - 2.417BF_2 - 0.792BF_3 + 1.655BF_4 - 0.288BF_5 - 0.721BF_6 + 0.432BF_7 - 0.826BF_8 - 0.391BF_9 + 0.581BF_{10} - 0.076BF_{11} - 0.058BF_{12}$	12	0.0008
24 h Forecast horizon			
D_1	$y = 0.176 + 0.617BF_1 + 0.749BF_2 - 0.448BF_3$	3	0.01339
D_2	$y = 0.214 + 0.98BF_1 + 0.486BF_2 - 1.183BF_3 - 0.538BF_4 - 0.764BF_5 - 0.158BF_6 + 1.820BF_7$	7	0.01269
D_3	$y = 0.092 + 1.106BF_1 - 0.487BF_2 - 0.387BF_3 + 0.592BF_4 + 1.872BF_5 - 0.864BF_6 + 0.400BF_7 + 0.750BF_8 - 0.819BF_9 - 1.197BF_{10} - 1.528BF_{11}$	11	0.01187

- (i) the maximum number of basis functions reached threshold rule $\min [200, \max(20, 2n) + 1]$, where $n =$ the number of inputs;
- (ii) adding a new basis function changed the coefficient of determination (R^2) by less than 1×10^{-4} ;
- (iii) R^2 reached ≈ 1 ;
- (iv) the number of basis functions including the intercept term reached the number of data observations; or
- (v) the effective number of parameters reached the number of observed data points.

In the deletion phase, the large model, which typically over-fits the data, was pruned back one-at-a-time to reduce RMSE until only the model's intercept term remained. Subsequently, the model with lowest Generalized Cross-Validation (GCV) was selected. The MARS model (H_4) used for the 1.0 h forecasting horizon had 20 basis functions and the lowest GCV at the pruning stage was indicated with 10 functions (Fig. 3(b)). Table 4 shows the forecasting equations (in training periods) with optimum basis functions (BF_m) and the GCV for all forecast horizons. A MARS model's GCV statistic after the pruning stage should be relatively small.

To offer a comparative framework for the SVR and MARS models, the ARIMA model was developed using the R package [46,63]. Table 3 displays the ARIMA model's architecture. Since many model identification methods exist, a selection technique was implemented that considered the coefficient of determination (R^2), Akaike information criterion (AIC) [64], log likelihood (L) [64] and the lowest variance (σ^2).

Since G data was non-stationary as observed in Fig. 2, a differencing process was applied to convert the G data to stationary and satisfy the ARIMA model's input requirements as previously mentioned [46,63]. The requirement was confirmed by ensuring the results of *autoarima* (AR) function [65] obtained the lowest standard deviation and AIC with the highest L .

Additionally, the autoregressive (p), differencing (d), and moving average terms (q) were determined iteratively [46]. The estimates of p and q were obtained by testing reasonable values and evaluating how the criteria, L , AIC, σ , and R^2 , were satisfied. The fitted ARIMA model was then optimised with 'trial' values of p, d , and q . The training

performance was unique for each forecasting horizon and in accordance with the goodness-of-fit parameters shown in Table 3.

3.3. Model performance evaluation

Error criteria were adopted to establish the accuracy of the data-driven models. [66–71]. These include the Mean Absolute Error (MAE), RMSE, relative error (%) based on MAE and RMSE values ($MAE_{\bar{G}}$ and $RMSE_{\bar{G}}$), correlation coefficient (r), Willmott's Index (WI), the Nash–Sutcliffe coefficient (E_{NS}), and Legates and McCabe Index (E_{LM}) [41,67–69,72–74] represented below:

$$r = \frac{\sum_{i=1}^{i=n} [(G_i^{obs} - \bar{G}^{obs})(G_i^{for} - \bar{G}^{for})]}{\sqrt{\sum_{i=1}^{i=n} (G_i^{obs} - \bar{G}^{obs})^2} \cdot \sqrt{\sum_{i=1}^{i=n} (G_i^{for} - \bar{G}^{for})^2}} \quad (13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2} \quad (14)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |G_i^{for} - G_i^{obs}| \quad (15)$$

$$RMSE_{\bar{G}} = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}}{\bar{G}^{obs}} \quad (16)$$

$$MAE_{\bar{G}} = 100 \times \frac{1}{n} \sum_{i=1}^{i=n} \left| \frac{G_i^{for} - G_i^{obs}}{\bar{G}^{obs}} \right| \quad (17)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (|G_i^{for} - \bar{G}^{obs}| + |G_i^{obs} - \bar{G}^{obs}|)^2} \right], \text{ and } 0 \leq WI \leq 1 \quad (18)$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (G_i^{obs} - \bar{G}^{obs})^2} \right], \text{ and } \infty \leq E_{NS} \leq 1 \quad (19)$$

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^{i=n} |G_i^{obs} - G_i^{for}|}{\sum_{i=1}^{i=n} |G_i^{obs} - \bar{G}^{obs}|} \right], \text{ and } (\infty \leq E_{LM} \leq 1) \quad (20)$$

where n is the total number of observed (and forecasted) values of G , G_i^{for} is the i^{th} forecasted value of G , \bar{G}^{for} is the mean of forecasted values, G_i^{obs} is the i^{th} observed value of G , \bar{G}^{obs} is the mean of observed values.

The model statistics, obtained through equations (13)–(20), aimed to assess the accuracy of the G forecasts with respect to observed G values. For instance, the covariance-based metric r served to analyse the statistical association between G_i^{for} and G_i^{obs} where $r = 1$ represents an absolute positive (ideal) correlation; $r = -1$, an absolute negative correlation; and $r = 0$, a lack of any linear relationship between G_i^{for} and G_i^{obs} data. According to the work of Chai and Draxler [70], the $RMSE$ is more representative than the MAE when the error distribution is Gaussian. However, when it is not the case, the use of MAE , $RMSE$, and their relative expressions, $MAE_{\bar{G}}$ and $RMSE_{\bar{G}}$, can yield complementary evaluations. Since other metrics can also assess model performance [70], the E_{NS} and WI were also calculated. A value of E_{NS} and WI near 1.0 represents a perfect match between G_i^{for} and G_i^{obs} , while a complete mismatch between the G_i^{for} and G_i^{obs} results in values of ∞ and 0, respectively. For example, when E_{NS} , which is the ratio of the mean square error to the variance in the observed data, equals 0.0, it indicates that \bar{G}^{obs} is as good a predictor as G_i^{for} , however, if E_{NS} is less than 0.0, the square of the differences between G_i^{for} and G_i^{obs} is as large as the variability in G_i^{obs} and indicates that \bar{G}^{obs} is a better predictor than G_i^{obs} [74,75]. As a result, using a modified version of WI , which is the Legates and McCabe Index ($\infty \leq E_{LM} \leq 1$) [74], can be more advantageous than the traditional WI , when relatively high values are expected as a result of squaring of differences [68,73]. On the other

hand, the $MAE_{\bar{G}}$ and $RMSE_{\bar{G}}$ were applied to compare forecasts at different timescales that yield errors of different magnitudes (e.g., Fig. 2). According to [41,42,76,77], a model can be considered excellent when $RMSE_{\bar{G}} < 10\%$, good if the model satisfies $10\% < RMSE_{\bar{G}} < 20\%$, fair if it satisfies $20\% < RMSE_{\bar{G}} < 30\%$, and poor if $RMSE_{\bar{G}} > 30\%$.

4. Results and discussion

Evaluation of the data-driven models' ability to forecast the electricity demand (G) data for the 0.5 h, 1.0 h, and 24 h horizons is presented in this section using the statistical metrics from Eqs. (13)–(20). Only optimum models with lowest MAE and largest r and WI are shown in Table 5. Between the SVR and ARIMA models, the MARS model yielded better G forecasting results for the 0.5 h and 1.0 h horizons. This was evident when comparing the MARS (T^b) model's accuracy statistics ($r = 0.993$, $WI = 0.997$, and $MAE = 45.363$ MW) with the equivalent SVR (T^b) and ARIMA^b models' results ($r = 0.990$, $WI = 0.995$ and $MAE = 55.915$ MW) and ($r = 0.423$, $WI = 0.498$ and $MAE = 362.860$ MW), respectively.

While both the MARS and SVR models yielded accurate G forecasts when predictor variables were trained for the data period from 21-12-2015 to 31-12-2015, the ARIMA model attained the highest accuracy for data trained in period 31-12-2015 (i.e., model ARIMA^a; $r = 0.976$, $WI = 0.702$ and $MAE = 237.746$ MW). Despite being significantly inferior to the MARS and SVR models for longer periods, the ARIMA models' performance improved when a shorter data set (i.e., 31-12-2015) was utilised. When the four ARIMA models for 0.5 h forecasting horizons (developed in Table 3) were evaluated, an increase in the correlation coefficient (0.128–0.976) was identified. In addition, a respective decrease was observed in MAE and $RMSE$ values (475.087–237.746 MW) and (569.282–256.565 MW) respectively, with parallel changes in WI and E_{NS} values.

The analysis based on Fig. 1(a) confirmed that the ARIMA model was most responsive in forecasting G data when input conditions had lower variance, as detected in single day's data (31-12-2015) in comparison to longer periods (1–12 to 31-12-2015). Therefore, the SVR and MARS models had a distinct advantage over the ARIMA model when a lengthy database was used for G forecasting. Furthermore, when models were evaluated for the 1.0 h forecasting horizon (Table 5), the MARS and SVR models (H_4), with four sets of lagged input combinations, were the most accurate and outperformed the best ARIMA model.

The MARS model was significantly superior to the SVR and ARIMA

Table 5
Evaluation of the optimal models attained for 0.5 h, 1.0 h and 24 h forecast horizons in the test period.

Model	Model Accuracy Statistics ^a				
	r	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)
<i>0.5 h Forecast horizon</i>					
MARS(T^b)	0.993	0.997	0.986	57.969	45.363
SVR(T^b)	0.990	0.995	0.980	70.909	55.915
ARIMA ^b	0.423	0.498	0.080	476.835	362.860
ARIMA ^a	0.976	0.702	-0.233	256.565	237.746
<i>1.0 h Forecast horizon</i>					
MARS(H_4)	0.990	0.994	0.978	106.503	86.502
SVR(H_4)	0.972	0.981	0.930	189.703	124.453
ARIMA	0.401	0.381	0.144	665.757	555.637
<i>24 h Forecast horizon</i>					
MARS(D_3)	0.753	0.859	0.543	256.000	200.426
SVR(D_3)	0.806	0.890	0.647	225.125	162.363
ARIMA	0.289	0.459	-1.018	538.124	474.390

^a r = correlation coefficient, E_{NS} = Nash–Sutcliffe coefficient, MAE = mean absolute error, $RMSE$ = root mean square error, WI = Willmott's index.

models for the 1.0 h forecasting horizon. Based on the r , WI , and MAE metrics, the MARS model ($r = 0.990$, $WI = 0.994$ and $MAE = 86.502$ MW) outperformed the SVR model ($r = 0.972$, $WI = 0.981$ and $MAE = 124.453$ MW). The MARS model's WI , a more robust statistic than the linear dependence measured by r [66], was 1.33% greater than the SVR model's. This was supported by the MARS model's lower $RMSE$ and MAE values, 78.12% and 43.87%, respectively. In contrast, the ARIMA model displayed an inferior performance ($r = 0.401$, $WI = 0.381$ and $MAE = 555.637$ MW) as seen in Table 5.

For a 24 h forecasting horizon, the SVR ($r = 0.806$, $WI = 0.890$ and $MAE = 162.363$ MW) outperformed the MARS model (D_3) by a small margin ($r = 0.753$, $WI = 0.859$ and $MAE = 200.426$ MW) (Table 5). Similarly to the hourly scenario, the ARIMA model performed poorly ($r = 0.289$, $WI = 0.459$ and $MAE = 474.390$ MW). It is important to consider that the ARIMA models for hourly and daily forecasting horizons were developed using the long time-series: 1-11-2015 to 31-12-2015 and 1-1-2012 to 31-12-2015, respectively. The predictor (historical G) data exhibited significant fluctuations over these long-term periods compared to the single day G data of 31-12-2015 (ARIMA^c).

In conjunction with statistical metrics and visual plots of forecasted vs. observed G data, the $MAE_{\bar{G}}$, $RMSE_{\bar{G}}$, and E_{LM} (e.g., [17,41,42,78]) are used to show the alternative 'goodness-of-fit' of the model-generated G in relation to observed G data. The MARS model yielded relatively high precision (lowest $MAE_{\bar{G}}$ and $RMSE_{\bar{G}}$ and the highest E_{LM}) followed by the SVR and ARIMA models (Table 6). For the MARS model, $MAE_{\bar{G}}/RMSE_{\bar{G}}$ for the 0.5 h and 1.0 h forecasting horizons were 0.77/0.99% (T^b) and 1.45/1.76% (H_4), respectively. On the other hand, the SVR model resulted in 0.95/1.21% (T^b) and 2.19/3.13% (H_4). Likewise, E_{LM} was utilised in combination with other performance metrics for a robust assessment of models [74]. The respective value for both 0.5 h and 1.0 h forecasting horizons was determined to be greater for the MARS model (0.887/0.857) than for the SVR model (0.861/0.794). Although the MARS models outperformed the SVR models for the 0.5 h and 1.0 h horizons, the SVR model surpassed the MARS model for the 24 h horizon (13.73%/23.63% lower $RMSE_{\bar{G}}/MAE_{\bar{G}}$ and 45.42% higher E_{LM}). It is evident that both the MARS and SVR models, adapted for G forecasting in the state of Queensland, exceeded the performance of the ARIMA model and thus, should be further explored for use in electricity demand estimation.

Nevertheless, despite the ARIMA model faring slightly worse for most of the G forecasting scenarios in this paper, specifically for the case of 1.0 h and 24 h horizons ($RMSE_{\bar{G}} = 11.0\%$ and 9.04% , respectively), its performance for the 0.5 h horizon using a single day's data (ARIMA^c) exhibited good results. This is supported by an $RMSE_{\bar{G}}$ value of approximately 4.18% (Table 6). Therefore, it is possible that a large degree of fluctuation in the longer training dataset could have led the ARIMA model's autoregressive mechanism to be more prone to cumulative errors than to a situation with a shorter data span.

Table 6

The relative root mean square error $RMSE_{\bar{G}}(\%)$, mean absolute percentage error $MAE_{\bar{G}}(\%)$ and Legates & McCabes Index (E_{LM}) for the optimal models in the test datasets.

Model	E_{LM}	$MAE_{\bar{G}}(\%)$	$RMSE_{\bar{G}}(\%)$
0.5 h Forecast horizon			
MARS(T^b)	0.887	0.765	0.990
SVR(T^b)	0.861	0.945	1.211
ARIMA ^b	0.098	6.487	8.140
ARIMA ^c	-0.238	3.939	4.184
1.0 h Forecast horizon			
MARS(H_4)	0.857	1.446	1.760
SVR(H_4)	0.794	2.192	3.134
ARIMA	0.080	9.350	11.000
24 h Forecast horizon			
MARS(D_3)	0.295	3.359	4.300
SVR(D_3)	0.429	2.717	3.781
ARIMA	-0.668	8.193	9.039

In contrast to previous studies on the MARS, SVR, or ARIMA models, the forecasting models developed in this study achieved a relatively high precision for short-term G forecasting. For example, a study that forecasted daily G data for South Africa using the MARS model attained an $RMSE$ of 446.01 MW [19], whereas the present study's MARS model resulted in an $RMSE$ of 256.00 MW (see MARS(D_3) in Table 5). Likewise, 24 h lead time forecasts of G in Istanbul (Turkey) using an RBF-based SVR model [23] yielded an $MAE_{\bar{G}}$ of 3.67%, whereas the $MAE_{\bar{G}}$ value obtained in the present study was 2.72% (see SVR(D_3) in Table 6). For the same forecast horizon, the ARIMA model

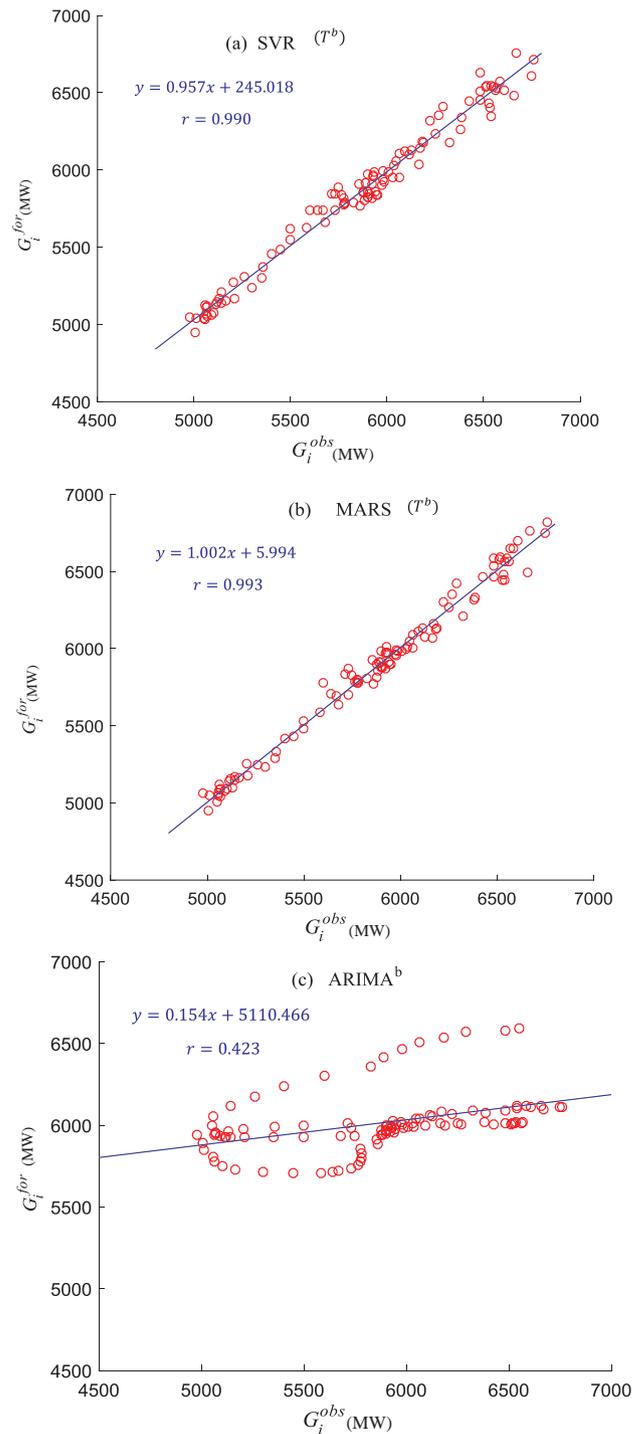


Fig. 4. Scatterplot of the forecasted, G_i^{for} vs. the observed, G_i^{obs} electricity demand data in the testing period for the 0.5 h forecast horizon, (a) SVR(T^b) (b) MARS(T^b) and (c) ARIMA^b. A linear regression line, $y = G_i^{for} = a'G_i^{obs} + b'$ with the correlation coefficient, r is included.

reported in [38], denoted as $(p,d,q) = (4,1,4)$, yielded an $RMSE$ value of 584.72 MW compared to a lower $RMSE$ of 538.12 MW achieved with the present ARIMA model denoted as $(p,d,q) = (8,1,3)$. Furthermore, a study that forecasted G data in New South Wales, Queensland and Singapore [79], used singular spectrum analysis, gravitational search, and adaptive particle swarm optimization following a gravitational search algorithm (APSOGSA) to forecast G . The APSOGSA model yielded an $MAE/RMSE$ of 115.59/133.99 MW and an $MAE_{\bar{G}}$ of 2.32%. Equivalent models in this study seem to exceed the others' performance

as evidenced in Tables 5 and 6. The analysis for $MARS(T^b)$ and $SVR(T^b)$ resulted in an $MAE/RMSE$ of 45.36/57.97 and 55.92/70.91 MW and $MAE_{\bar{G}}$ of 0.77 and 0.95%, respectively.

Separately, Figs. 4–6 depict scatterplots of G_i^{for} vs. G_i^{obs} for the 0.5 h, 1.0 h and 24 h forecasting horizons using optimal MARS, SVR, and ARIMA models (see Table 5). A least square regression line, $y = G_i^{for} = a'G_i^{obs} + b'$, and r value are used to illustrate the relationship between G_i^{for} and G_i^{obs} data, where a' is the slope and b' is the y -intercept. Both are used to describe the model's accuracy [17].

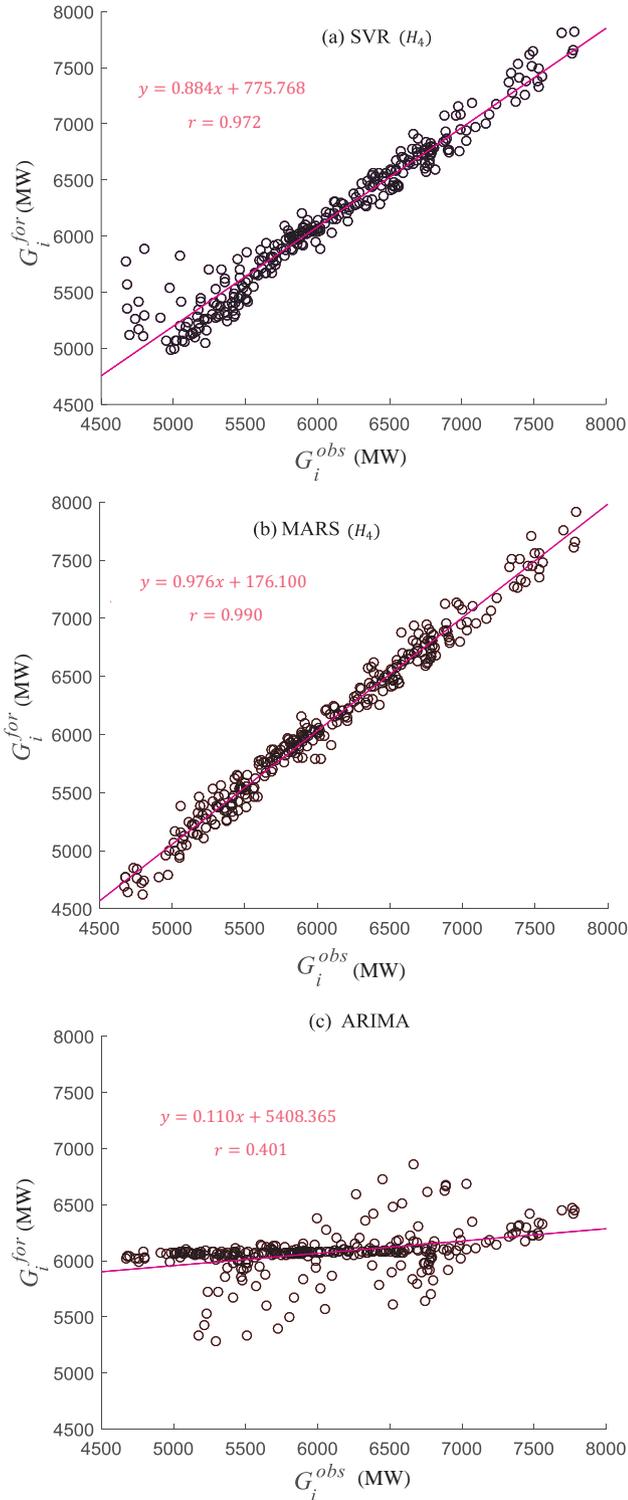


Fig. 5. The caption description is the same as that in Fig. 4 except for the 1.0 h forecast horizon, (a) SVR(H_4), (b) MARS(H_4) and (c) ARIMA.

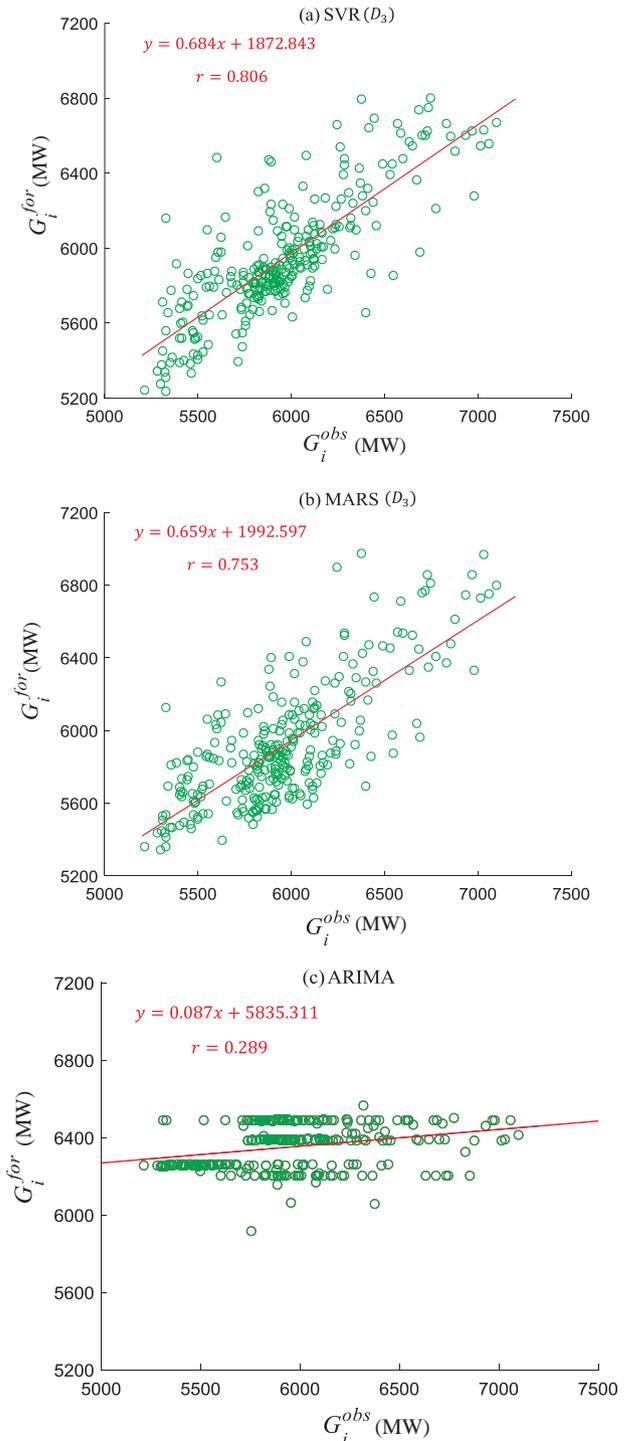


Fig. 6. The caption description is the same as that in Fig. 4 except for the 24 h forecast horizon, (a) SVR(D_3) (b) MARS(D_3) and (c) ARIMA.

For the 0.5 h horizon, the optimal SVR and MARS models yielded near unity a' values of 0.957 and 1.002, respectively. On the contrary, the a' for the ARIMA model (ARIMA^b) was 0.154 deviating significantly from an ideal value of 1 (Fig. 4a–c). The deviation of forecasted G data from observations (i.e., 1:1 line or reference a' -value of 1) was largest in the case of the ARIMA model, approximately 0.846. In the case of the SVR and MARS models, these deviations were 0.043 and 0.002, respectively.

Consistent with the level of scattering, the r value for the MARS model exceeded the SVR and ARIMA models' values. In concordance

with the r value trends, similar results for a' values were attained for 1.0 h forecasting where the optimum MARS, SVR, and ARIMA models (Fig. 5a, b and c) yielded 0.976, 0.884 and 0.110, respectively. Additionally, for the 24 h forecasting horizon (Fig. 6a, b and c), the SVR model ($r = 0.806$, $a' = 0.684$, $b' = 1872.843$) outperformed the MARS model ($r = 0.753$, $a' = 0.659$, $b' = 1992.597$). Both models provided

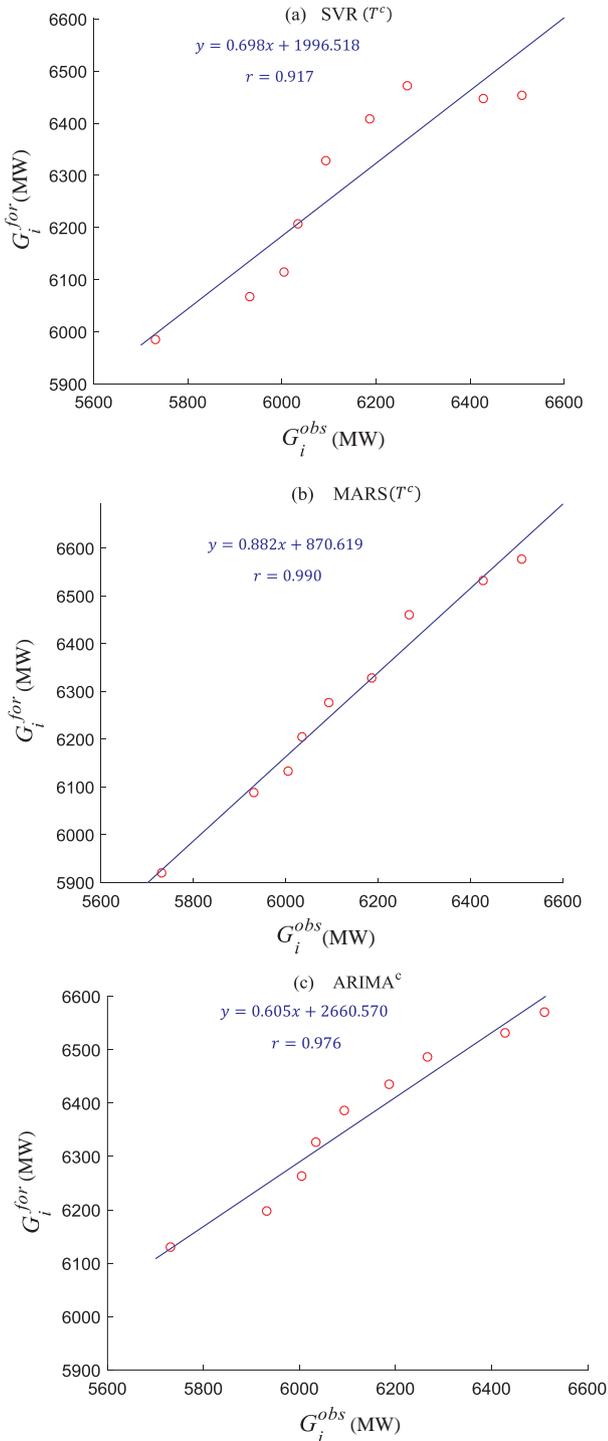


Fig. 7. The caption description is the same as that in Fig. 4 except for the 0.5 h forecast horizon, (a) SVR(T^c), (b) MARS(T^c) and (c) ARIMA^c.

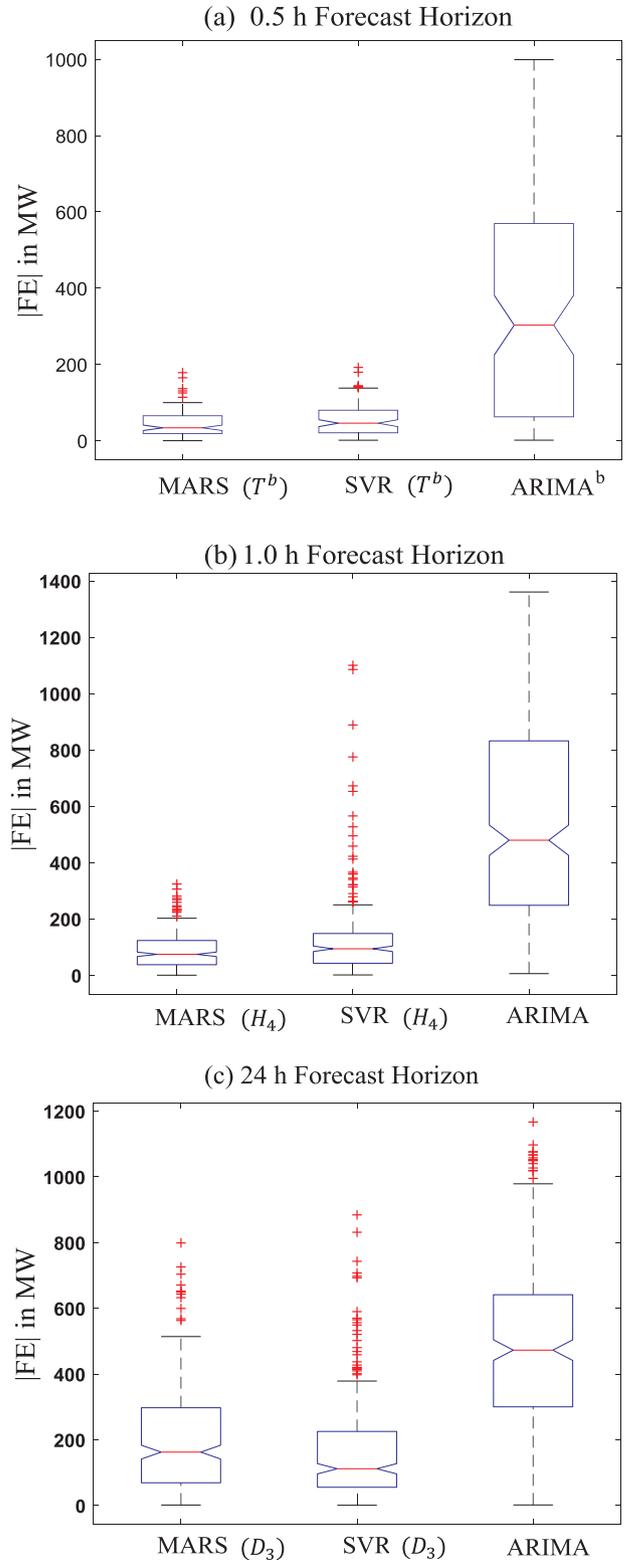


Fig. 8. Boxplots of the absolute forecasted error, $|FE| = |G_{FOR,i} - G_{OBS,i}|$ for: (a) 0.5 h, (b) 1.0 h and (c) 24 h forecast horizons using the MARS, SVR and ARIMA models.

better results than the ARIMA model ($r = 0.289$, $a' = 0.087$ and $b' = 5835.311$).

On the other hand, Fig. 7 compares the performance for the shortest horizon (0.5 h) using G data gathered over a single day (i.e., 31-12-2015) partitioned into training and testing phases. The MARS model ($r = 0.99$) outperformed the SVR ($r = 0.917$) and ARIMA ($r = 0.976$) models. However, it is important to note that the performance of the ARIMA, for the shorter dataset, was better than its performance for longer datasets (Table 5). This suggests that the ARIMA model's performance deteriorated as the forecasting period increased. This concurs with its auto-regressive and integrated averaging nature since the sum of preceding errors is used for forecasting the next G value [46]. Although the cause is not yet clear, the ARIMA^c model's better performance could be attributed to greater fluctuations in longer-term predictor data drawn upon in the hourly and daily models (Table 2 and Fig. 2).

Boxplots showing the error distribution for absolute values of forecasted error statistics, $|FE| = |G_i^{for} - G_i^{obs}|$, reveal a greater amount of detail about the models' precision, where the whiskers (Fig. 8) represent the extremes of the forecasted and the observed G values. The lower end of each boxplot represents the lower quartile, Q_{25} (25th percentile); the upper end shows the upper quartile, Q_{75} (75th percentile); and the central line shows the second quartile, Q_{50} (i.e., 50th percentile) or the median value. Two horizontal whiskers are also extended from Q_{25} to the smallest non-outlier and from Q_{75} to the largest non-outlier, respectively. Based on the box plots, Table 7 summarizes statistical properties of the forecasted and observed G data.

For all forecasting horizons considered, the MARS and SVR models performed better than the ARIMA model and therefore, demonstrated significant differences. In terms of the maximum absolute error, the MARS model was most precise for the 0.5 h horizon. For example, the MARS(T^b) resulted in a maximum $|FE|$ of 178.54 MW (Fig. 8a and Table 7) and the smallest median value ($Q_{50} \approx 33.77$ MW) relative to any other model. Similarly, for the 1.0 forecasting scenario, statistics indicated the superiority of the MARS model over the SVR and ARIMA models (Table 7; Fig. 8b).

When the errors for the 24 h forecasting horizon were analysed, the MARS and SVR resulted in similar maximum values but distinctly lower than for the ARIMA model. When the median errors were compared, the SVR model (111.76 MW) generated more accurate forecasts than the MARS model (162.41 MW). These median errors differed significantly from those of the ARIMA model (479.66 MW; Table 7; Fig. 8c).

Fig. 9(a–c) illustrates the percentage of the absolute value of forecasted error statistics ($|FE|$) encountered through the empirical cumulative distribution function (ECDF) for optimal models at

different forecasting horizons. With respect to the percentage of errors located in the smallest error bracket (i.e., 0 to ± 50 MW), the ECDF demonstrated that the MARS and SVR models outperformed the ARIMA model for all forecasting horizons. Based on this error bracket, the MARS performed slightly better than the SVR model (i.e., about 60% vs. 57% and 34% vs. 28% for 0.5 h and 1.0 h forecasts, respectively). Within the error bracket of 0 to ± 100 MW for the 0.5 h horizon, the MARS model recorded about 94% of all forecasted errors, whereas the SVR model only 83%. Additionally, for the 1.0 h horizon, the MARS model performed better than the SVR model (i.e., about 63% vs. 53% of errors within the 0 to ± 100 MW bracket). However, data for the 24 h horizon recorded comparable values between the two in the smaller error bracket. Nonetheless, better percentage was yielded for the SVR (about 49%) against MARS (about 38%) in the 0 to ± 100 MW bracket.

Since the MARS and SVR models illustrated similar performance in several cases, a statistical t -test was utilised to demonstrate whether the differences in the mean of $|FE|$ were significant. For the 0.5 h, 1.0 h, and 24 h forecasting horizons, we could reject the null-hypothesis that the means are the same (p -value < 0.05). Consequently, the differences in the means are statistically significant for the absolute values of the forecasted errors generated by the MARS and the SVR model.

Based on Table 5, the ARIMA model proved highly inaccurate for the short-term 0.5 h G forecasting horizon as nearly 60% of the errors in the testing period fell in the error range magnitude of greater than 100 MW (Fig. 9a). Similar observations were evident for about 90% of the hourly and daily ARIMA forecasts (Fig. 9b and c). The ARIMA models' forecasting accuracy for the 0.5 h horizon exceeded those for 1.0 h or 24 h horizons as the percentage of errors received from ECDF in the smallest category (0.5 h) was nearly double. This concurred with earlier results (Table 5) where overall evaluation metrics demonstrated the greatest correlation between the observed and ARIMA-forecasted G , including higher WI and E_{NS} and lower $RMSE/MAE$ values.

Ultimately, the versatility of data-driven models was also examined with respect to the forecasting errors for peaks in G by plotting the ten greatest relative errors (Fig. 10). Except for one data point, it was apparent that the MARS models consistently yielded the lowest percent errors for the 0.5 h and 1.0 h forecasting horizons compared to the SVR or ARIMA models (Fig. 10a and b). In contrast, for the 24 h forecasting horizon, the ten highest relative error values were very similar between the MARS and SVR models, but dramatically lower for the ARIMA model (Fig. 10c). The accuracy of the present data-driven models appeared to deteriorate as the forecasting period was extended. This was demonstrated by the relative performance errors (Table 6), the top error values (Fig. 10), and the statistical distribution of the errors (Fig. 8 and Table 7).

Table 7

Evaluation of the differences in the absolute value of forecast error statistics based on observed and forecasted G in the test period for the optimal models.

Error Statistic ^a (MW)	Forecast horizon (h)								
	0.5 h			1.0 h			24 h		
	MARS(T^b)	SVR(T^b)	ARIMA ^b	MARS(H_4)	SVR(H_4)	ARIMA	MARS(D_3)	SVR(D_3)	ARIMA
Maximum	178.54	192.20	999.10	324.09	1100.50	1360.80	798.97	884.25	1177.40
Minimum	0.02	1.21	1.34	0.27	1.20	5.85	1.41	0.78	2.72
Q_{25}	18.54	20.61	62.54	37.47	42.65	248.39	69.15	55.47	270.17
Q_{50}	33.77	45.97	302.97	74.47	94.03	479.54	162.41	111.76	479.66
Q_{75}	65.37	79.99	569.29	123.46	148.31	832.24	297.44	225.36	667.07
Range	178.52	190.99	997.76	323.82	1099.30	1355.00	797.56	883.47	1174.70
Skewness	1.23	0.91	0.46	0.99	3.68	0.45	1.77	1.77	0.09
Flatness	4.59	3.14	1.87	3.98	21.13	2.15	3.82	6.61	2.34

^a Lower quartile (Q_{25}), median (Q_{50}), upper quartile (Q_{75}).

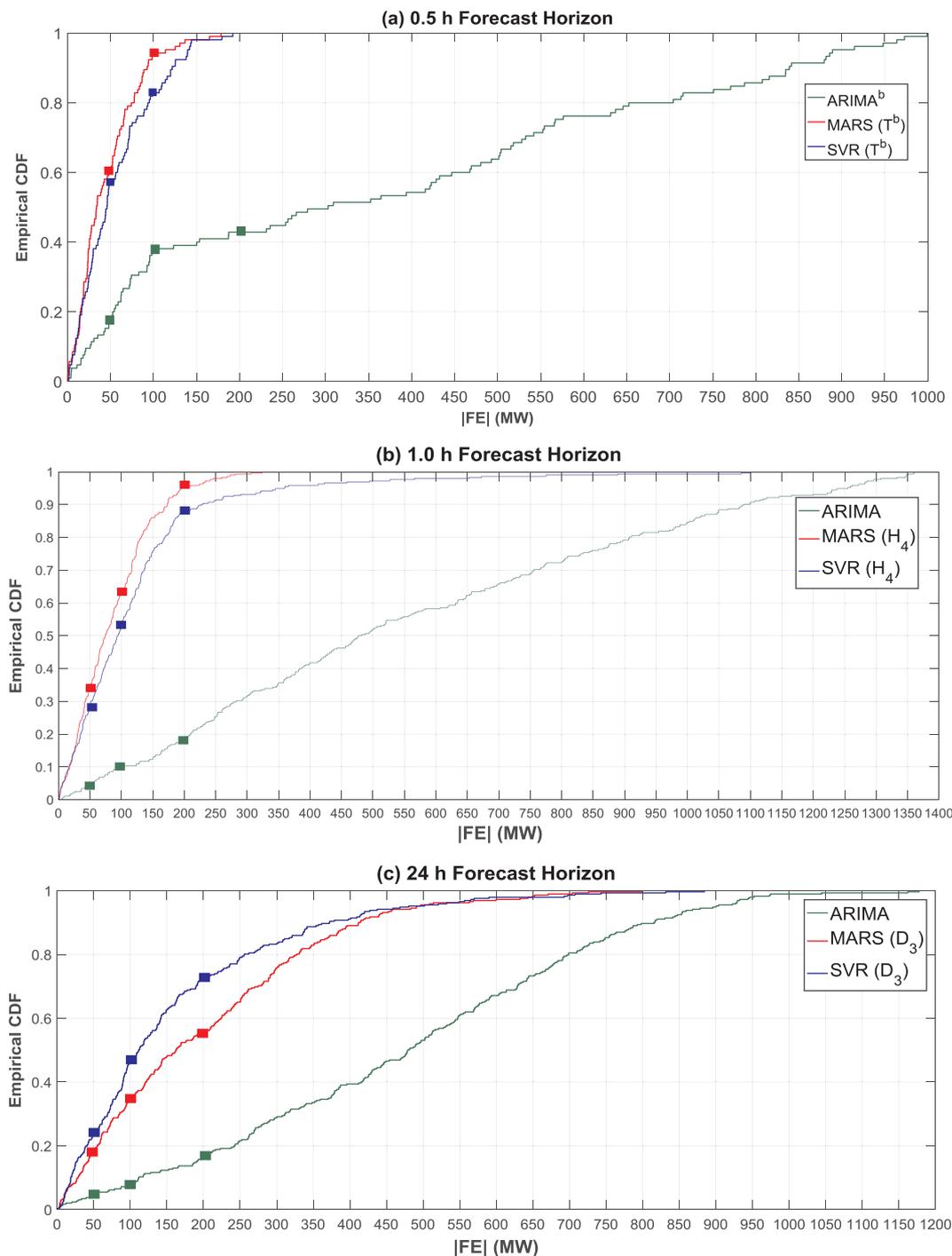


Fig. 9. Empirical cumulative distribution function (ECDF) of the forecast error, $|FE|$ for: (a) 0.5 h, (b) 1.0 h and (c) 24 h forecast horizons using the MARS, SVR and ARIMA models.

5. Further discussion, limitations and opportunities for future research

Data-driven models applied for G forecasting over multiple forecast horizons were evaluated. The SVR models were constructed by optimizing regulation constants (minimizing the training error) and radial basis function width (Table 3). The MARS models were tuned with a piecewise multivariate regression function based on the lowest GCV statistic, while the ARIMA models were optimised by a trial and error process (Tables 3 and 4). A comprehensive evaluation showed a greater accuracy of the MARS models when compared to the SVR and ARIMA models for 0.5 h and 1.0 h forecasting horizons. However, for 24 h

forecasting horizon, the SVR performed considerably better (Tables 5 and 6).

Given the importance of accurately forecasting G data to meet engineering and energy demand challenges, including the sustainable operation of the NEM, this research paper has highlighted the potential utility of further exploring the MARS and SVR models to improve G forecasting accuracy. Particularly, this research study established the distinct advantage of the MARS model if employed in real-time G forecasting. In terms of greater speed, simplicity of development, and efficiency in performance, the MARS model was best adapted to such forecasts given the SVR models' requirements for tedious modelling phases (*i.e.*, identifying the regulation and kernel width parameters via

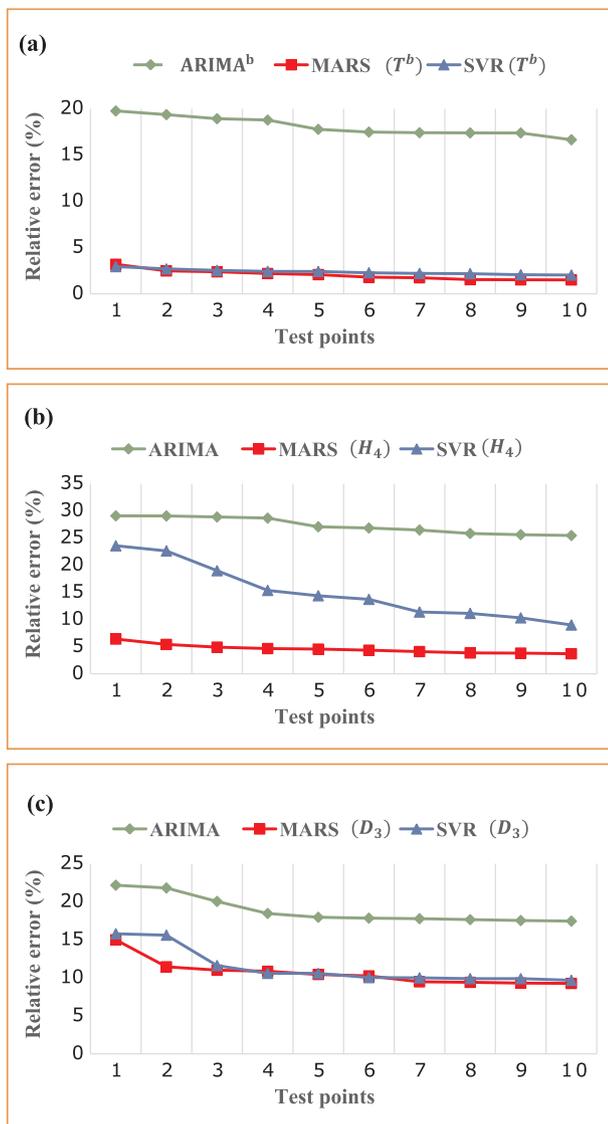


Fig. 10. The top ten peak relative forecast errors (%) generated by the MARS, SVR and ARIMA models for: (a) 0.5 h, (b) 1.0 h and (c) 24 h forecast horizons.

a grid search approach).

Comparable to existing studies in Australia (e.g., [48,49,80,81]), this research has revealed the greater accuracy of the proposed models employed for forecasting. For instance, the SVR model [SVR(D_3)], applied for daily forecasting, attained an $RMSE_{\bar{G}}$ of 3.781% (Table 6), which is similar to 2.42% (whole weekly forecast) reported in [80]. Likewise, MAE and $RMSE$ values for the weekly-average data forecasted in the same study were 224.18 MW and 311.04 MW, whereas for the SVR(D_3) model they resulted in 162.363 MW and 225.125 MW. Also, an adaptive neuro-wavelet model employed for G forecasting, in Queensland, showed a $0.16\% < MAE_{\bar{G}} < 0.99\%$ over 7 days in the test period [81]. Comparably, the $MAE_{\bar{G}}$ values were 0.355 for the MARS and 0.502 for the SVR models for the 0.5 h forecasting horizon in the present study. Moreover, recent studies [48,49] have adopted statistical approaches for 0.5 h forecasting to support the Australian Energy Market Operator; they have used the drivers of energy use (e.g., temperatures, calendar effects, demographic and economic variables) in combination with demand and time of the year to forecast G . Differently to these studies, which adopted a semi-parametric additive model, the developed MARS and SVR models were an improvement as data assumptions or linear considerations were not employed. These models

were guided by a fully data-driven modelling process.

Although this study was the first to evaluate the MARS and SVR models for short-term G forecasting in Queensland, multiple limitations should be addressed in future research. In this paper, the only predictor data used was time-lagged (historical) G . Alternative models for short-term horizons can also incorporate climate data (e.g., temperature, rainfall, humidity and solar radiation) that modulate electricity demand influenced by consumers' needs under different conditions. According to previous work (e.g., [19,82]), climatic factors can have an influence on G . For instance, an inverse relationship exists between electricity demand and ambient air temperature in wintertime, when lighting and heating usage are likely to increase. Similarly, this relationship can also occur in the summer when an increase in temperature can lead to increased air-conditioning demand [83]. Therefore, in a follow-up study, the MARS and SVR models could utilise seasonal data (both G and climatic factors). While this study provided accurate aggregated data models for Queensland, distinctive regions in the state are likely to exhibit different conditions.

In this study, a radial basis function was used to develop the SVR models employing a grid search to identify the parameters (C and σ). Despite the grid search demonstrating good performance, it is envisaged that a genetic algorithm (GA) [84] could serve to identify appropriate parameters for the model. GAs have been extensively applied to optimization problems [85–87]. According to [88], a GA-SVR was able to outperform other comparable models and yield high forecasting accuracy.

It is important to note that the MARS and SVR models could be improved by wavelet transformation (WT) and ensemble-based uncertainty testing via a bootstrapping procedure. This procedure uses a Bayesian Model Averaging (BMA) framework to assess the models' stability [52,89,90]. Many studies (e.g., [17,91,92]) have suggested that WT could deliver benefits by decomposing predictor time series into time and frequency domains. Also, non-stationarity features in real data can be encapsulated by partitioning them with low and high pass filters. For example, very good results were obtained by a WT-SVR model for short and long-term solar forecasting when compared to the standard SVR model [17]. In addition, the data-driven technique of bootstrapping can also serve as an ensemble framework to reduce parametric uncertainties through resampling of inputs [93,94]. A hybrid wavelet-bootstrap-neural network model could be explored since such a model has outperformed non-WT models for water demand forecasting [51]. The use of the BMA also resulted in a better understanding of model uncertainty compared to a simple equal-weighted forecasting averaging method [95]. In addition to WT-based models, empirical model decomposition, applied for G forecasting in New South Wales (Australia), could similarly be employed in the present region to improve the MARS and SVR models. Considering other work [17,96–98], it is recommended that future research applies the WT, ensembles, and BMA to explore their usefulness for G forecasting.

6. Concluding remarks

Data-driven models based on the MARS, SVR and ARIMA algorithms were evaluated for short-term G forecasting using Queensland's area-aggregated data from the Australian Energy Market Operator. To demonstrate their feasibility for real-time applications, partial auto-correlation functions were applied to G data to identify significant inputs for three forecast horizons: 0.5 h, 1.0 h, and 24 h, with an identical number of predictive features (Table 2 and Fig. 2).

The versatility of the trained models for shorter span predictor data (31-12-2015) was investigated. Performances were assessed via correlation coefficient (r) between observed and forecasted G data in the testing period along with other performance metrics such as root mean square error ($RMSE$), mean absolute error (MAE), relative $RMSE$ and MAE (%), Willmott's Index (WI), Nash–Sutcliffe coefficient (E_{NS}), and Legates and McCabe Index (E_{LM}). In terms of the statistical metrics, the

MARS model yielded the most accurate results for 0.5 h and 1.0 h forecasts, whereas the SVR models were better for a 24 h horizon. As expected, given its linear formulation in the modelling process, the ARIMA model's performance was lower for all forecasting horizons as it generated very high forecast errors.

Although this paper has advanced the work of previous studies (e.g., [48,49,80,81]), it is also a pilot study in the context of the present study region (i.e., Queensland). Future studies will employ Energen G data coupled with exogenous climate predictors for identified substations in the metropolitan Queensland area with the largest populations (i.e., Brisbane, Gold Coast, Sunshine Coast, Logan, Ipswich, Redlands and Moreton Bay). The aim is to apply the MARS and SVR models via wavelet transformation and incorporate an ensemble framework and BMA to explore a coherent mechanism for uncertainty in forecasting models.

To summarize, the MARS and SVR models represent useful data-driven tools that can be used for G forecasting, and as such, they should be explored by forecasters working in the National Electricity Market (e.g., AEMO). In particular, this study found that the MARS models provide a powerful, yet simple and fast forecasting framework when compared to the SVR models. However, the incorporation of a data pre-processing scheme (e.g., wavelet transformation or empirical mode decomposition) as well as model uncertainty tools (e.g., Bayesian and ensemble models) are alternative tools that could be explored for energy demand forecasting for engineering applications in an independent follow-up study.

Acknowledgments

Data were acquired from the Australian Energy Market Operator (AEMO). We thank Associate Professor Xiaohu Wen (Cold and Arid Regions Engineering and Environmental Institute, China) for his initial advice on the SVR model and the Ministry of Higher Education and Scientific Research, in the Government of Iraq for funding the first author's PhD project. We extend thanks to Dr. Georges Dodds for editing. Dr Ravinesh Deo acknowledges the Academic Development and Outside Studies Program and CAS Presidential Fellowship. We thank both Reviewers and the Handling Editor Professor Timo Hartmann for their constructive criticisms that helped improve the clarity of the paper.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.aei.2017.11.002>.

References

- [1] Z. Hu, Y. Bao, T. Xiong, Electricity load forecasting using support vector regression with memetic algorithms, *Scient. World J.* 2013 (2013).
- [2] F. Kaytez, M.C. Taplamacioglu, E. Cam, F. Hardalac, Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines, *Int. J. Electr. Power Energy Syst.* 67 (2015) 431–438.
- [3] D. Akay, M. Atak, Grey prediction with rolling mechanism for electricity demand forecasting of Turkey, *Energy* 32 (2007) 1670–1675.
- [4] S. Fan, L. Chen, Short-term load forecasting based on an adaptive hybrid method, *IEEE Trans. Power Syst.* 21 (2006) 392–401.
- [5] D. Bunn, E.D. Farmer, Comparative models for electrical load forecasting, 1985.
- [6] T. Haida, S. Muto, Regression based peak load forecasting using a transformation technique, *IEEE Trans. Power Syst.* 9 (1994) 1788–1794.
- [7] E. Erdogan, Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey, *Energy Policy* 35 (2007) 1129–1146.
- [8] H. Zareipour, K. Bhattacharya, C. Canizares, Forecasting the hourly Ontario energy price by multivariate adaptive regression splines, 2006 IEEE Power Engineering Society General Meeting, IEEE, 2006, p. 7.
- [9] G. Nasr, E. Badr, M. Younes, Neural networks in forecasting electrical energy consumption: univariate and multivariate approaches, *Int. J. Energy Res.* 26 (2002) 67–78.
- [10] L. Suganthi, T. Jagadeesan, Energy substitution methodology for optimum demand variation using Delphi technique, *Int. J. Energy Res.* 16 (1992) 917–928.
- [11] N. Dalkey, O. Helmer, An experimental application of the Delphi method to the use of experts, *Manage. Sci.* 9 (1963) 458–467.
- [12] L. Suganthi, A.A. Samuel, Energy models for demand forecasting—a review, *Renew. Sustain. Energy Rev.* 16 (2012) 1223–1240.
- [13] E. Xydas, C. Marmaras, L.M. Cipcigan, N. Jenkins, S. Carroll, M. Barker, A data-driven approach for characterising the charging demand of electric vehicles: A UK case study, *Appl. Energy* 162 (2016) 763–771.
- [14] M. Florens, I.H. Cairns, S. Knock, P. Robinson, Data-driven solar wind model and prediction of type II bursts, *Geophys. Res. Lett.* 34 (2007).
- [15] J. Contreras, R. Espinola, F.J. Nogales, A.J. Conejo, ARIMA models to predict next-day electricity prices, *IEEE Trans. Power Syst.* 18 (2003) 1014–1020.
- [16] A. Sözen, M. Ali Akçayol, Modelling (using artificial neural-networks) the performance parameters of a solar-driven ejector-absorption cycle, *Appl. Energy* 79 (2004) 309–325.
- [17] R.C. Deo, X. Wen, F. Qi, A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset, *Appl. Energy* 168 (2016) 568–593.
- [18] A.K. Singh, S.K. Ibraheem, M. Muazzam, D. Chaturvedi, An overview of electricity demand forecasting techniques, *Network Complex Syst.* 3 (2013) 38–48.
- [19] C. Sigauke, D. Chikobvu, Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach, *ORiON*. 26 (2010) 97.
- [20] A.J. Smola, B. Schölkopf, *Learning with Kernels*, Citeseer, 1998.
- [21] P.-S. Yu, S.-T. Chen, I.-F. Chang, Support vector regression for real-time flood stage forecasting, *J. Hydrol.* 328 (2006) 704–716.
- [22] S. Salcedo-Sanz, J.L. Rojo-Álvarez, M. Martínez-Ramón, G. Camps-Valls, Support vector machines in engineering: an overview, *Wiley Interdiscipl. Rev.: Data Min. Knowl. Discov.* 4 (2014) 234–267.
- [23] B.E. Türkay, D. Demren, Electrical load forecasting using support vector machines, 2011 7th International Conference on Electrical and Electronics Engineering (ELECO), IEEE, 2011, pp. 1–49–1–53.
- [24] M. Mohandes, Support vector machines for short-term electrical load forecasting, *Int. J. Energy Res.* 26 (2002) 335–345.
- [25] C. Sivapragasam, S.-Y. Liong, Flow categorization model for improving forecasting, *Hydrol. Res.* 36 (2005) 37–48.
- [26] R.C. Deo, P. Samui, D. Kim, Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models, *Stoch. Env. Res. Risk Assess.* 1–16 (2015).
- [27] V. Sharda, S. Prasher, R. Patel, P. Ojasvi, C. Prakash, Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data/Performances de régressions par splines multiples et adaptives (MARS) pour la prévision d'écoulement au sein de micro-bassins versants Himalayens d'altitudes intermédiaires avec peu de données, *Hydrol. Sci. J.* 53 (2008) 1165–1175.
- [28] J.H. Friedman, Multivariate adaptive regression splines, *Annals Stat.* (1991) 1–67.
- [29] N.O. Attoh-Okin, K. Cooger, S. Mensah, Multivariate adaptive regression (MARS) and hinged hyperplanes (HHP) for doveled pavement performance modeling, *Constr. Build. Mater.* 23 (2009) 3020–3023.
- [30] P. Samui, Determination of ultimate capacity of driven piles in cohesionless soil: a multivariate adaptive regression spline approach, *Int. J. Numer. Anal. Meth. Geomech.* 36 (2012) 1434–1439.
- [31] W. Zhang, A.T.C. Goh, Multivariate adaptive regression splines for analysis of geotechnical engineering systems, *Comput. Geotech.* 48 (2013) 82–95.
- [32] C. Morana, A semiparametric approach to short-term oil price forecasting, *Energy Econ.* 23 (2001) 325–338.
- [33] R.C. Deo, O. Kisi, V.P. Singh, Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model, *Atmos Res.* 184 (2017) 149–175, <http://dx.doi.org/10.1016/j.atmosres.2016.10.004>.
- [34] P. Sephton, Forecasting recessions: Can we do better on mars, *Federal Reserve Bank of St Louis Rev.* 83 (2001).
- [35] A. Abraham, D. Steinberg, Is Neural Network a Reliable Forecaster on Earth? A MARS query! *Bio-Inspired Applications of Connectionism*, Springer, 2001, pp. 679–686.
- [36] W.K. Buchanan, P. Hodges, J. Theis, Which way the natural gas price: an attempt to predict the direction of natural gas spot price movements using trader positions, *Energy Econ.* 23 (2001) 279–293.
- [37] C. Yuan, S. Liu, Z. Fang, Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model, *Energy* 100 (2016) 384–390.
- [38] N.H. Miswan, N.H. Hussin, R.M. Said, K. Hamzah, E.Z. Ahmad, ARAR Algorithm in Forecasting Electricity Load Demand in Malaysia, *Glob. J. Pure Appl. Math.* 12 (2016) 361–367.
- [39] AEU, Australian Energy Update. Department of Industry and Science, 2015, Australian energy update, Canberra, August, 2015.
- [40] AEMO, Aggregated Price and Demand Data - Historical. Australia: The Australian Energy Market Operator, 2016. p. The Australian Energy Market Operator is responsible for operating Australia's largest gas and electricity markets and power systems.
- [41] K. Mohammadi, S. Shamshirband, M.H. Anisi, K.A. Alam, D. Petković, Support vector regression based prediction of global solar radiation on a horizontal surface, *Energy Convers. Manage.* 91 (2015) 433–441.
- [42] K. Mohammadi, S. Shamshirband, C.W. Tong, M. Arif, D. Petković, S. Ch, A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation, *Energy Convers. Manage.* 92 (2015) 162–171.
- [43] V.N. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [44] J.A.K. Suykens, J. De Brabanter, L. Lukas, J. Vandewalle, Weighted least squares support vector machines: robustness and sparse approximation, *Neurocomputing.* 48 (2002) 85–105.

- [45] P.A.W. Lewis, J.G. Stevens, Nonlinear modeling of time series using multivariate adaptive regression splines (MARS), *J. Am. Stat. Assoc.* 86 (1991) 864–877.
- [46] G.E. Box, G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, revised ed., Holden-Day, 1976.
- [47] R.E. Rosca, Stationary and non-stationary time series, *USV Ann. Econ. Public Administr.* 10 (2011) 177–186.
- [48] S. Fan, R.J. Hyndman, *Forecasting long-term peak half-hourly electricity demand for Queensland. Business & Economic Forecasting Unit (Monash University): Report for The Australian Energy Market Operator (AEMO)*, 2015.
- [49] S. Fan, R.J. Hyndman, *Monash Electricity Forecasting Model (Version 2015.1). Business & Economic Forecasting Unit (Monash University): Report for The Australian Energy Market Operator (AEMO)*, 2015.
- [50] K. Sudheer, A. Gosain, K. Ramasastri, A data-driven algorithm for constructing artificial neural network rainfall-runoff models, *Hydrol. Process.* 16 (2002) 1325–1330.
- [51] M.K. Tiwari, J. Adamowski, Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models, *Water Resour. Res.* 49 (2013) 6486–6507.
- [52] M.K. Tiwari, C. Chatterjee, A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting, *J. Hydroinform.* 13 (2011) 500–519.
- [53] C.-W. Hsu, C.-C. Chang, C.-J. Lin, *A practical guide to support vector classification*, 2003.
- [54] H.-T. Lin, C.-J. Lin, A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods, *Neural Comput.* (2003) 1–32.
- [55] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, *ACM Trans. Intell. Syst. Technol. (TIST)*. 2 (2011) 27.
- [56] S.S. Keerthi, C.-J. Lin, Asymptotic behaviors of support vector machines with Gaussian kernel, *Neural Comput.* 15 (2003) 1667–1689.
- [57] R. Maity, P.P. Bhagwat, A. Bhatnagar, Potential of support vector regression for prediction of monthly streamflow using endogenous property, *Hydrol. Process.* 24 (2010) 917–923.
- [58] M.K. Goyal, B. Bharti, J. Quilty, J. Adamowski, A. Pandey, Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS, *Exp. Syst. Appl.* 41 (2014) 5267–5276.
- [59] N.-D. Hoang, A.-D. Pham, M.-T. Cao, A novel time series prediction approach based on a hybridization of least squares support vector regression and swarm intelligence, *Appl. Comput. Intell. Soft Comput.* 14 (2014) 15.
- [60] G. Jekabsons, *Adaptive Regression Splines toolbox for Matlab/Octave*. Version. 2013;1:72.
- [61] C. Kooperberg, D.B. Clarkson, Hazard regression with interval-censored data, *Biometrics* (1997) 1485–1494.
- [62] S. Milborrow, *Multivariate Adaptive Regression Splines. Package 'earth': Derived from mda:mars by Trevor Hastie and Rob Tibshirani Uses Alan Miller's Fortran utilities with Thomas Lumley's leaps wrapper*, 2016 < <http://www.milbo.users.sonic.net/earth> > .
- [63] J. Adamowski, H. Fung Chan, S.O. Prasher, B. Ozga-Zielinski, A. Sliusarieva, Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada, *Water Resour. Res.* 48 (2012).
- [64] S. Hu, *Akaike Information Criterion*, Center for Research in Scientific Computation, 2007.
- [65] Hyndman R, Khandakar Y. *Automatic Time Series Forecasting: The Forecast Package for R 7*. 2008. < <https://www.wjstatsoft.org/article/view/v027i03> > , 2007.
- [66] P. Krause, D. Boyle, F. Bäse, Comparison of different efficiency criteria for hydrological model assessment, *Adv. Geosci.* 5 (2005) 89–97.
- [67] C.J. Willmott, Some comments on the evaluation of model performance, *Bull. Am. Meteorol. Soc.* 63 (1982) 1309–1313.
- [68] C.J. Willmott, On the validation of models, *Phys. Geogr.* 2 (1981) 184–194.
- [69] C.W. Dawson, R.J. Abraham, L.M. See, HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts, *Environ. Modell. Software* 22 (2007) 1034–1052.
- [70] T. Chai, R.R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature, *Geoscient. Model Develop.* 7 (2014) 1247–1250.
- [71] J. Nash, J. Sutcliffe, River flow forecasting through conceptual models part I—a discussion of principles, *J. Hydrol.* 10 (1970) 282–290.
- [72] C.J. Willmott, S.M. Robeson, K. Matsuura, A refined index of model performance, *Int. J. Climatol.* 32 (2012) 2088–2094.
- [73] C.J. Willmott, On the Evaluation of Model Performance in Physical Geography. *Spatial Statistics and Models*, Springer, 1984, pp. 443–460.
- [74] D.R. Legates, G.J. McCabe, Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation, *Water Resour. Res.* 35 (1999) 233–241.
- [75] B.P. Wilcox, W. Rawls, D. Brakensiek, J.R. Wight, Predicting runoff from rangeland catchments: a comparison of two models, *Water Resour. Res.* 26 (1990) 2401–2410.
- [76] M.-F. Li, X.-P. Tang, W. Wu, H.-B. Liu, General models for estimating daily global solar radiation for different solar radiation zones in mainland China, *Energy Convers. Manage.* 70 (2013) 139–148.
- [77] A.B. Heinemann, P.A. van Oort, D.S. Fernandes, A.d.H.N. Maia, Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation, *Bragantia* 71 (2012) 572–582.
- [78] R. Deo, N. Downs, A. Parisi, J. Adamowski, J. Quilty, Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle, *Environ. Res.* 155 (2017) 141.
- [79] M. Niu, S. Sun, J. Wu, L. Yu, J. Wang, An innovative integrated model using the singular spectrum analysis and nonlinear multi-layer perceptron network optimized by hybrid intelligent algorithm for short-term load forecasting, *Appl. Math. Model.* 40 (2016) 4079–4093.
- [80] N. An, W. Zhao, J. Wang, D. Shang, E. Zhao, Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting, *Energy* 49 (2013) 279–288.
- [81] B.-L. Zhang, Z.-Y. Dong, An adaptive neural-wavelet model for short term load forecasting, *Electr. Power Syst. Res.* 59 (2001) 121–129.
- [82] C.-L. Hor, S.J. Watson, S. Majithia, Analyzing the impact of weather variables on monthly electricity demand, *IEEE Trans. Power Syst.* 20 (2005) 2078–2085.
- [83] UKCIP, *London's warming - The Impacts of Climate Change on London. Technical report. The UK Climate Impacts Programme (UKCIP)*, 2002.
- [84] D. Zhang, W. Liu, A. Wang, Q. Deng, Parameter optimization for support vector regression based on genetic algorithm with simplex crossover operator, *J. Inform. Comput. Sci.* 8 (2011) 911–920.
- [85] D.E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. Addison Wesley, 1989;1989:102.
- [86] J.J. Grefenstette, Optimization of control parameters for genetic algorithms, *IEEE Trans. Syst., Man, Cybernet.* 16 (1986) 122–128.
- [87] D.B. Fogel, An introduction to simulated evolutionary optimization, *IEEE Trans. Neural Networks* 5 (1994) 3–14.
- [88] C.-H. Wu, G.-H. Tzeng, Y.-J. Goo, W.-C. Fang, A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy, *Exp. Syst. Appl.* 32 (2007) 397–408.
- [89] J.M. Slaughter, T. Gneiting, A.E. Raftery, Probabilistic wind speed forecasting using ensembles and Bayesian model averaging, *J. Am. Stat. Assoc.* 105 (2010) 25–35.
- [90] Q. Feng, X. Wen, J. Li, Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions, *Water Resour. Manage.* 29 (2015) 1049–1065.
- [91] P. Areekul, T. Senjyu, N. Urasaki, A. Yona, Neural-wavelet approach for short term price forecasting in deregulated power market, *J. Int. Council Electr. Eng.* 1 (2011) 331–338.
- [92] Z. Tan, J. Zhang, J. Wang, J. Xu, Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models, *Appl. Energy* 87 (2010) 3606–3610.
- [93] B. Efron, *Bootstrap Methods: Another Look At the Jackknife*. Breakthroughs in Statistics, Springer, 1992, pp. 569–593.
- [94] B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap*, CRC Press, 1994.
- [95] J.H. Wright, Forecasting US inflation by Bayesian model averaging, *J. Forecast.* 28 (2009) 131–144.
- [96] J. Kim, B.P. Mohanty, Y. Shin, Effective soil moisture estimate and its uncertainty using multimodel simulation based on Bayesian Model Averaging, *J. Geophys. Res.: Atmos.* 120 (2015) 8023–8042.
- [97] S. Crimp, K.S. Bakar, P. Kokic, H. Jin, N. Nicholls, M. Howden, Bayesian space-time model to analyse frost risk for agriculture in Southeast Australia, *Int. J. Climatol.* 35 (2015) 2092–2108.
- [98] R. Maheswaran, R. Khosa, Long term forecasting of groundwater levels with evidence of non-stationary and nonlinear characteristics, *Comput. Geosci.* 52 (2013) 422–436.

Chapter 4: Multiple Electricity Demand Forecasting Horizons Using Hybrid Models

4.1 Foreword

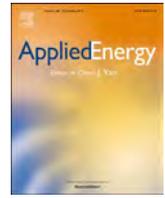
This chapter presents an exact copy of the published article in *Applied Energy* journal (Vol. 217, Pages 422-439).

This study investigates an improvement in forecast method and accuracy developed in Chapter 3 by addressing the behaviours of non-stationarity associated with the target (G) data. A hybrid model is developed utilizing the support vector regression (SVR) model integrated with two-phase method employing an improved version of empirical mode decomposition with adaptive noise (ICEEMDAN) tool and the particle swarm optimization (PSO) algorithm. After applying ICEEMDAN and the partial autocorrelation function (PACF) to decompose the original G data into high and low frequency sub-series and extract the significant lags (inputs) for each series, respectively, the PSO is used to select the best parameters of the SVR models developed for each series individually. Electricity demand (G) data with multiple horizons of short-term (weekends, working days, whole weeks, and public holidays) and long-term (monthly) for whole Queensland, Australia were adopted to evaluate the ICEEMDAN-PSO-SVR model against ICEEMDAN- multivariate adaptive regression spline (MARS), ICEEMDAN-M5 model tree and their traditional models. A novel contribution and innovation of this work is shown by developing a new energy forecasting model with two-phase hybrid machine learning.

4.2 Research Highlights

- Hybrid two-phase PSO-SVR is integrated with ICEEMDAN multi-resolution tool for demand forecasting.
- ICEEMDAN-PSO-SVR is evaluated against single-phase hybrid and standalone models.
- ICEEMDAN-PSO-SVR outperforms several benchmark models at multiple-horizons.
- Two-phase hybrid model has potential applications in energy management systems.

4.3 Published Article II



Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting

Mohanad S. AL-Musaylh^{a,b,*}, Ravinesh C. Deo^{a,*}, Yan Li^a, Jan F. Adamowski^c

^a School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, QLD 4350, Australia

^b Management Technical College, Southern Technical University, Basrah, Iraq

^c Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Québec H9X 3V9, Canada

HIGHLIGHTS

- Hybrid two-phase PSO-SVR is integrated with ICEEMDAN multi-resolution tool for demand forecasting.
- ICEEMDAN-PSO-SVR is evaluated against single-phase hybrid and standalone models.
- ICEEMDAN-PSO-SVR outperforms several benchmark models at multiple-horizons.
- Two-phase hybrid model has potential applications in energy management systems.

ARTICLE INFO

Keywords:

SVR
PSO
Improved CEEMDAN
Electricity demand
MARS
M5 model tree
Energy management system

ABSTRACT

Real-time energy management systems that are designed to support consumer supply and demand spectrums of electrical energy continue to face challenges with respect to designing accurate and reliable real-time forecasts due to the stochasticity of model construction data and the model's inability to disseminate both the short- and the long-term electrical energy demand (G) predictions. Using real G data from Queensland, Australia's second largest state, and employing the support vector regression (SVR) model integrated with an improved version of empirical mode decomposition with adaptive noise (ICEEMDAN) tool, this study aims to propose a novel hybrid model: ICEEMDAN-PSO-SVR. Optimization of the model's weights and biases was performed using the particle swarm optimization (PSO) algorithm. ICEEMDAN was applied to improve the hybrid model's forecasting accuracy, addressing non-linear and non-stationary issues in time series inputs by decomposing statistically significant historical G data into intrinsic mode functions (IMF) and a residual component. The ICEEMDAN-PSO-SVR model was then individually constructed to forecast IMFs and the residual datasets and the final G forecasts were obtained by aggregating the IMF and residual forecasted series. The performance of the ICEEMDAN-PSO-SVR technique was compared with alternative approaches: ICEEMDAN-multivariate adaptive regression spline (MARS) and ICEEMDAN-M5 model tree, as well as traditional modelling approaches: PSO-SVR, MARS and M5 model tree algorithms. To develop the models, data were partitioned into different subsets: training (70%), validation (15%), and testing (15%), and the tuned forecasting models with near global optimum solutions were applied and evaluated at multiple horizons: short-term (i.e., weekends, working days, whole weeks, and public holidays), and long-term (monthly). Statistical metrics including the root-mean square error (RMSE), mean absolute error (MAE) and their relative to observed means (RRMSE and MAPE), Willmott's Index (WI), the Legates and McCabe Index (E_{LM}) and Nash–Sutcliffe coefficients (E_{NS}), were used to assess model accuracy in the independent (testing) period. Empirical results showed that the ICEEMDAN-PSO-SVR model performed well for all forecasting horizons, outperforming the alternative comparison approaches: ICEEMDAN-MARS and ICEEMDAN-M5 model tree and the PSO-SVR, PSO-MARS and PSO-M5 model tree algorithm. Due to its high predictive utility, the two-phase ICEEMDAN-PSO-SVR hybrid model was particularly appropriate for whole week forecasts ($E_{NS} = 0.95$, $MAPE = 0.89\%$, $RRMSE = 1.22\%$, and $E_{LM} = 0.79$), and monthly forecasts ($E_{NS} = 0.70$, $MAPE = 2.18\%$, $RRMSE = 3.18\%$, and $E_{LM} = 0.56$). The excellent performance of the ICEEMDAN-PSO-SVR hybrid model indicates that the two-phase hybrid model should be explored for potential applications in real-time energy management systems.

* Corresponding authors at: School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and Environment (IAg&E), University of Southern Queensland, QLD 4350, Australia.

E-mail addresses: MohanadShakirKhalid.AL-Musaylh@usq.edu.au (M.S. AL-Musaylh), ravinesh.deo@usq.edu.au (R.C. Deo).

<https://doi.org/10.1016/j.apenergy.2018.02.140>

Received 13 November 2017; Received in revised form 13 February 2018; Accepted 21 February 2018

Available online 06 March 2018

0306-2619/© 2018 Elsevier Ltd. All rights reserved.

Nomenclature

MW	megawatt	ECDF	empirical cumulative distribution function
G	electricity load (demand; mega watts)	IMF	intrinsic mode functions
MARS	multivariate adaptive regression splines	ARIMA	autoregressive integrated moving average
SVR	support vector regression	ANN	artificial neural network
RMSE	root-mean square error	AEMO	australian energy market operator
MAE	mean absolute error	DWT	discrete wavelet transform
RRMSE	relative root-mean square error, %	c_1 and c_2	PSO parameters
MAPE	mean absolute percentage error, %	PACF	partial auto-correlation function
WI	Willmott's index	MSE	mean square error
E_{NS}	Nash–Sutcliffe coefficient	R^2	coefficient of Determination
E_{LM}	Legates and McCabe Index	MODWT	maximum overlap discrete wavelet transform
ω	weighting factor of PSO	EMD	empirical mode decomposition
ω_{min} and ω_{max}	the minimum and maximum of ω	WT	wavelet transforms
RBF	radial basis function for SVR	N	the initial population of PSO
ICEEMDAN	improved version of empirical mode decomposition with adaptive noise	G_i^{for}	i th forecasted value of G , (MW)
ϵ	loss function	G_i^{obs}	i th observed value of G , (MW)
σ	kernel width for SVR model	$\frac{G^{for}}{G^{obs}}$	the mean of forecasted values
C	regulation for SVR model	$ FE $	forecasted error statistics
PSO	particle swarm optimization	EEMD	ensemble EMD
GCV	generalized cross-validation	CEEMDAN	complete EEMD with adaptive noise
T_{max}	maximum number of iterations in PSO	r	correlation coefficient
		VMD	variational mode decomposition

1. Introduction

Electricity demand (G) forecasting can provide essential information that is likely to be utilized for energy transactions in competitive electricity markets [1–3]. Policies addressing energy distribution and pricing and providing energy security to a growing population requires accurate forecasting of G data, especially for short-term periods (e.g., daily). Estimating G is a very sensitive task as an error in under- or over-estimation of even just 1% can lead to millions of dollars in losses affecting the whole energy policy and management system [3–5]. As such, to estimate G , a very accurate near real-time (i.e., short-term), as well as a foresight (i.e., long-term), forecasting model is a useful tool.

In recent years, data-driven models, such as autoregressive integrated moving average (ARIMA) [6], artificial neural network (ANN) [7], support vector regression (SVR) [8], genetic algorithms, fuzzy logic, knowledge-based expert systems [9], M5 model tree [10,11] and multivariate adaptive regression splines (MARS) [12] have been widely adopted in energy demand forecasting studies. Based on structural risk minimization (SRM), the SVR model is able to reduce overfitting data through the minimization of expected error of a learning algorithm [13]. For example, the SVR model with a radial basis kernel function (RBF) has been used for G forecasting [14]. The parameters of the SVR model can be selected by different optimization techniques, such as a grid search procedure [15], particle swarm optimization (PSO) [16] and a genetic algorithm [17]. The PSO algorithm can be considered an effective method to solve engineering challenges and can also be used to provide better performance when used to screen the near global optimum set of SVR parameters [18,19]. On the other hand, the MARS model is a fast and flexible statistical tool that can be developed to adopt a piecewise (linear or cubic) basis function [20,21]. In the literature, a significantly lower root-mean-square error (RMSE) was found for the MARS model when compared with the piecewise regression-based model used for G forecasting [12]. A piecewise linear function in a M5 model tree [22] has also been used in different studies including wave [10] and solar energy studies [11].

However, these types of traditional machine learning methods often have challenges addressing non-stationary time series [22,23]. Non-stationary time series problems can be addressed by different model input data decomposition methods; for example, the discrete wavelet

transform (DWT) [8], maximum overlap discrete wavelet transform (MODWT) [24], empirical mode decomposition (EMD) [25], ensemble EMD (EEMD) [26], complete EEMD with adaptive noise (CEEMDAN) [27] or improved CEEMDAN (termed hereafter as the 'ICEEMDAN' algorithm) [28]. These algorithms resolve the frequency components present in input series prior to using them in the model development process. These techniques are powerful tools as they can be used to decompose the original data into high and low frequency sub-series to address the issues of non-stationary, repeats/periods and jump-type perturbations before such data are utilized for prediction purposes.

In spite of the many applications of wavelet transforms (WT) (e.g., [29–34]), recent studies show major weaknesses in WT-based models, particularly in their forecasting ability, which is limited by the adoption of non-causal filters constructed with DWT algorithms. It should be noted that DWT can induce the decimation effect in model input sub-series coefficients, and therefore generate half the coefficients of the detailed signal at the current level, while the other half of the smooth version can be recursively processed by the high pass and low pass filters at a coarser temporal resolution [35]. Although the problem of the decimation effect in DWT can be solved by the more advanced MODWT algorithm, selection of the mother wavelet is a still major issue as there is no rule to select a near global optimum wavelet other than applying an iterative trial and error process [24]. However, there is an alternative decomposition tool available to address such issues: the self-adaptive EMD algorithm that splits data into several intrinsic mode functions (IMFs) and a residual data subset. While the frequent appearance of mode mixing in the EMD algorithm is problematic [26], it can be addressed by the EEMD-based model, which is able to obtain a true number of IMFs. However, when a signal is reconstructed using the EEMD process, different numbers of IMFs can be obtained which generates a new problem [27]. In order to resolve all of these issues, CEEMDAN was developed [27] to precisely reconstruct the original time series data and give a better spectral separation of the IMFs at a lower computational cost. Some of the residual noise within IMFs and the slower performance of the algorithm compared to the EEMD are two major issues associated with the CEEMDAN algorithm [28]. Hence, the ICEEMDAN algorithm was developed to address issues of model input decomposition [28].

Few studies (e.g. [16,36,37]) have applied the ICEEMDAN for

Table 1
Descriptive statistics of the total electricity demand (G) (MW) for the Queensland (QLD) study region with different forecast horizons.

Forecast horizon	Time	Data period (dd-mm-yyyy)	No. data points	Minimum (MW)	Maximum (MW)	Mean (MW)	Standard deviation (MW)	Training (70%)	Validation (15%)	Testing (15%)	No. IMFs
Weekend	Saturday to Sunday	01-01-2012 to 31-12-2016	522	234990.82	335747.41	268757.98	18595.84	366	78	78	7
Working days	Monday to Friday		1305	238435.95	368972.19	290105.15	19584.54	913	196	196	7
Whole week	Monday to Sunday		1827	234990.82	368972.19	284005.96	21578.34	1279	274	274	8
Public holiday	Each holiday only	01-01-2000 to 31-12-2016	153	198984.14	319144.80	253174.38	25060.00	107	23	23	6
Monthly	Total monthly		204	6690812.27	9766378.01	8320704.42	657174.70	144	30	30	5

forecasting different kinds of data. The ability of the ICEEMDAN as a data decomposition and modelling tool was tested in an application to predict near-infrared non-invasive glucose detection with the PSO-SVR model [16]. Results confirmed that the hybrid ICEEMDAN-PSO-SVR model generated more accurate and relevant results with good predictive ability compare to the PSO-SVR model [16].The ICEEMDAN algorithm successfully reconstructed near-infrared spectra data and as a result, improved prediction when compared to other reconstructed signal methods, such as wavelets [37]. Successful estimation of the decay ratio of boiling water reactors employing the ICEEMDAN algorithm has also been reported [36]. In spite of this use of the ICEEMDAN algorithm, it has yet to be applied to *G* forecasting where the data contain highly stochastic and non-stationary features.

In recent years, many studies (e.g. [38–40]) have focused on improving energy markets in various ways, for example through renewable energy resources, minimization of the production cost, performance optimization, and forming collaborations between two distributed energy resources in energy markets. In addition, other studies (e.g. [41–46]) have focused on maximizing operational efficiency and minimizing operational cost, system stability and schedule, as well as the imbalance between the generated power and the load demand of an energy system. The uncertainty of the energy demand, which impacts an energy system, has also been explored (e.g. [47–50]) using appropriate statistical models, ANN combined with Markov-chain (MC) (ANN-MC), Taguchi’s orthogonal array testing approach, and the discrete Markov model, respectively. Finally, other studies have focused on developing hybrid and efficient forecasting models (e.g. [51,52]).

To develop a high-performance forecasting model for applications in electricity demand forecasting, this study built on earlier studies (e.g. [51]) to create a new hybrid model. This paper proposes a hybrid model that is designed by integrating a robust and efficient PSO-SVR technique with the ICEEMDAN algorithm. This constructs a two-phase hybrid ICEEMDAN-PSO-SVR model where the model inputs are firstly decomposed by the ICEEMDAN algorithm for better frequency resolution within the predictor data, and then the PSO algorithm is used to tune the weights from the model’s input features with the near global optimum. The contributions of this research are that we, for the first time, integrate the ICEEMDAN algorithm with the PSO-SVR model, to feed important frequency-based input information to predict energy demand. Subsequently, we incorporate large sets of data in a real problem of *G* forecasting to test the model at multiple lead times: short-term period (weekend, working days, whole week, and public holidays), and long-term (monthly) period using independent (real *G*) data. The newly proposed hybrid model was explored for Queensland, Australia’s second largest state, which is expected to experience significant growth and demand for consumer energy. The two-phase hybrid model was then benchmarked with alternative models: ICEEMDAN-MARS, ICEEMDAN-M5 model tree, and several traditional machine learning methods (PSO-SVR, MARS and M5 model tree).

This paper is organized into six different sections. We first provide a general introduction of the topic (Section 1) followed by Section 2 which provides a theoretical background for the models, and describes the details of the ICEEMDAN and PSO algorithms. We then provide a materials and methods section that describes the *G* data, model development, validation and evaluation processes in Section 3. Section 4 presents the results and discussion while limitations and future research opportunities are presented in Section 5. Finally the last section summarizes the research findings and key considerations.

2. Theoretical background

Only ICEEMDAN and PSO are explained in detail here as these methods are considered relatively new for *G* forecasting. The theoretical explanations of SVR, MARS and M5 model tree are explained elsewhere since they are well known approaches. A detailed background of SVR, a machine learning method, is provided by Vapnik [53]

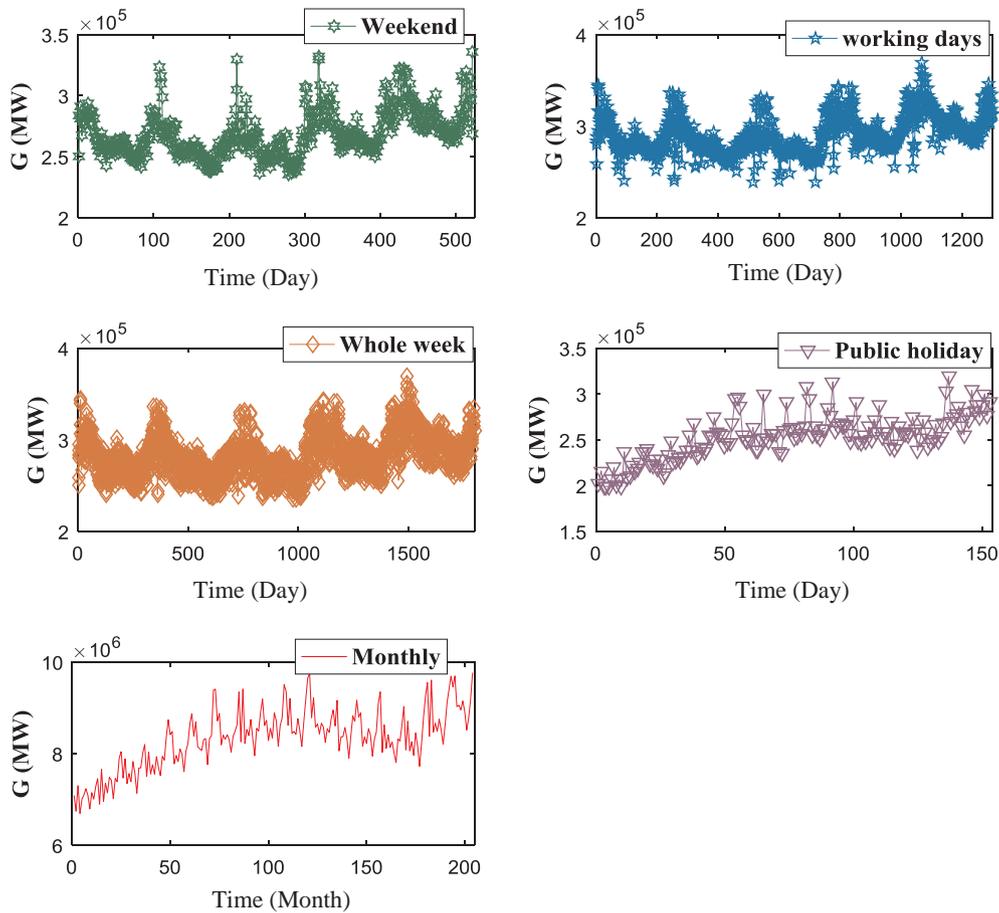


Fig. 1. Time-series of electricity demand (G MW) data and various forecasting periods.

and other studies explain its application in terms of solar radiation and G forecasting [8,14]. Details of the MARS method is given by Friedman [54]. The M5 model tree [55] requires two different steps which are well explained by Rahimikhoob et al. [56].

2.1. Improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN)

A signal $x(t)$ can be decomposed using the EMD algorithm, pioneered by Huang et al. [25], into several modes or intrinsic mode functions (IMFs). Based on the EMD process, the ICEEMDAN technique is used to decompose $x(t)$ into a number of IMFs and one residual component as follows [28]:

1. Consider the operator $E_j(\cdot)$ as the j th mode of the signal $x(t)$ decomposed by the EMD method and the operator $M(\cdot)$ produces the local mean of $x(t)$. Let ω^i be the white Gaussian noise with the mean and variance equal to 0 and 1, respectively.
2. The local means of n realizations $x^i(t) = x(t) + \beta_0 E_1(\omega^i(t))$ are calculated using EMD to obtain the first residue $r_1(t) = \langle M(x^i(t)) \rangle$, where $\beta_0 = \varepsilon_0 \text{std}(x) / \text{std}(E_1(\omega^i(t)))$ is used to avoid the fraction of the energy of the noise when the algorithm begins.
3. For $k = 1$, calculate the first mode: $\text{IMF}_1(t) = x(t) - r_1(t)$.
4. The average of the local means of the realizations $r_1(t) + \beta_1 E_2(\omega^i(t))$ is estimated as the second residue ($r_2(t)$). As a result, the second mode ($\text{IMF}_2(t)$) is obtained as:

$$\text{IMF}_2(t) = r_1(t) - r_2(t) = r_1(t) - \langle M(r_1(t) + \beta_1 E_2(\omega^i(t))) \rangle \quad (1)$$

5. For $k = 3, \dots, K, k^{\text{th}}$ for the residues and modes can be calculated as follows:

$$r_k(t) = \langle M(r_{k-1}(t) + \beta_{k-1} E_k(\omega^i(t))) \rangle \quad (2)$$

$$\text{IMF}_k(t) = r_{k-1}(t) - r_k(t) \quad (3)$$

6. For the next k , repeat step 5.

As the SNR between the added noise and the residue rises with k in the process of EEMD, the term $\beta_k = \varepsilon_k \text{std}(r_k(t)), k \geq 1$ is selected to control the system above and to reduce the amplitudes in the noise realizations [28].

2.2. Particle swarm optimization (PSO)

A fast convergence speed and the capability of dealing with high-dimensional problems is provided by the PSO algorithm [16]. The PSO process, first developed by Kennedy and Eberhart [57,58], is used in this study to optimize the 3D parameters of the SVR model, which are the regulation function (C), kernel width (σ) and loss function (ε). The iterative formula of PSO is presented below [16,59,60]:

$$V_{ij}(k+1) = \omega * V_{ij}(k) + c_1 * \text{rand}() * (P_{\text{best}_{ij}}(k) - X_{ij}(k)) + c_2 * \text{rand}() * (g_{\text{best}_j}(k) - X_{ij}(k)) \quad (4)$$

$$X_{ij}(k+1) = X_{ij}(k) + V_{ij}(k+1) \quad (5)$$

where $X_i = (X_{i1}, X_{i2}, \dots, X_{iD})^T$ is the i th particle from the initial population (swarm) of size $i = 1, 2, \dots, N$ and dimension $j = 1, 2, \dots, D$, and $V_i = (V_{i1}, V_{i2}, \dots, V_{iD})^T$ is the velocity of each particle X_i in the population. $P_{\text{best}_{ij}}$ and g_{best_j} are the individual and global extreme values, respectively, while $\text{rand}()$ is a random number between zero and one [16,60].

According to Alam, [59], the two values of c_1 and c_2 are the learning (acceleration) factors, which are usually within [2, 2.05], whereas ω is the weighting factor that keeps the swarm under control by slowly decreasing the velocity of particles [61]. It can be defined as follows [16]:

$$\omega = \omega_{min} + \frac{(T_{max}-T) * (\omega_{max}-\omega_{min})}{T_{max}} \quad (6)$$

where ω_{min} and ω_{max} are the minimum and maximum of ω that are usually determined to be 0.4 and 0.9, respectively. T and T_{max} are the current and maximum iteration numbers, respectively [16].

To determine the best three parameters (C, σ, ε) for the PSO-SVR, the mean square error (MSE) is employed as the fitness function in this algorithm [16]. Below, some stages that need to be considered to select the near global optimum parameters in PSO-SVR [16,59,60,62] are listed:

1. Set initial parameters of C, σ , and ε .
2. Set iteration $T = 1$.
3. Set parameters $c_1, c_2, \omega_{min}, \omega_{max}, T_{max}$ and population size.
4. Have positions X_i and velocities V_i vectors to produce particles.
5. Compute the fitness for each generated particle.
6. Update $P_{best_{i,j}}$ if the value of fitness is better than $P_{best_{i,j}}$.
7. Update g_{best_j} if the value of the updated $P_{best_{i,j}}$ in step 5 is better than $P_{best_{i,j}}$. Compute and update the velocity and position for each particle using Eqs. (1) and (2), respectively.
8. Go to step 4 and set $T = T + 1$, then continue until the system gets $T = T_{max}$.
9. Print the best SVR parameters with the lowest MSE.

3. Materials and method

3.1. Electricity demand data (G)

In this study, the G data was recorded half-hourly (48 times per day) in megawatts (MW) for the state of Queensland, and was acquired from the Australian Energy Market Operator (AEMO) [63] for the period of 01-01-2000 to 31-12-2016 (dd-mm-yyyy). Multiple datasets including both short-term [weekend (Saturday to Sunday), working days (Monday to Friday), whole week (Monday to Sunday), and public holiday], and long-term (monthly) horizons were used to forecast G . The 30-min data periods were converted to short and long terms by obtaining total values for each day and month, respectively. The datasets for the weekend, working days and whole week were collected from 01 to 01-2012 to 31-12-2016 and the dataset for the public holidays and monthly horizons were collected from 01 to 01-2000 to 31-12-2016.

For the various forecast horizons, descriptive statistics of the original data are found in Table 1, and Fig. 1 shows the plots of the actual data series. The data in Fig. 1 clearly exhibited large fluctuations due to the variations in consumer use of G . This was confirmed by the large values of the standard deviation, particularly for the public holiday and monthly periods (Table 1).

3.2. Forecast model development and validation

Historical G data were used to develop six models: ICEEMDAN-PSO-SVR, ICEEMDAN-MARS, ICEEMDAN-M5 model tree, and their traditional variants (PSO-SVR, MARS and M5 model tree). The original data

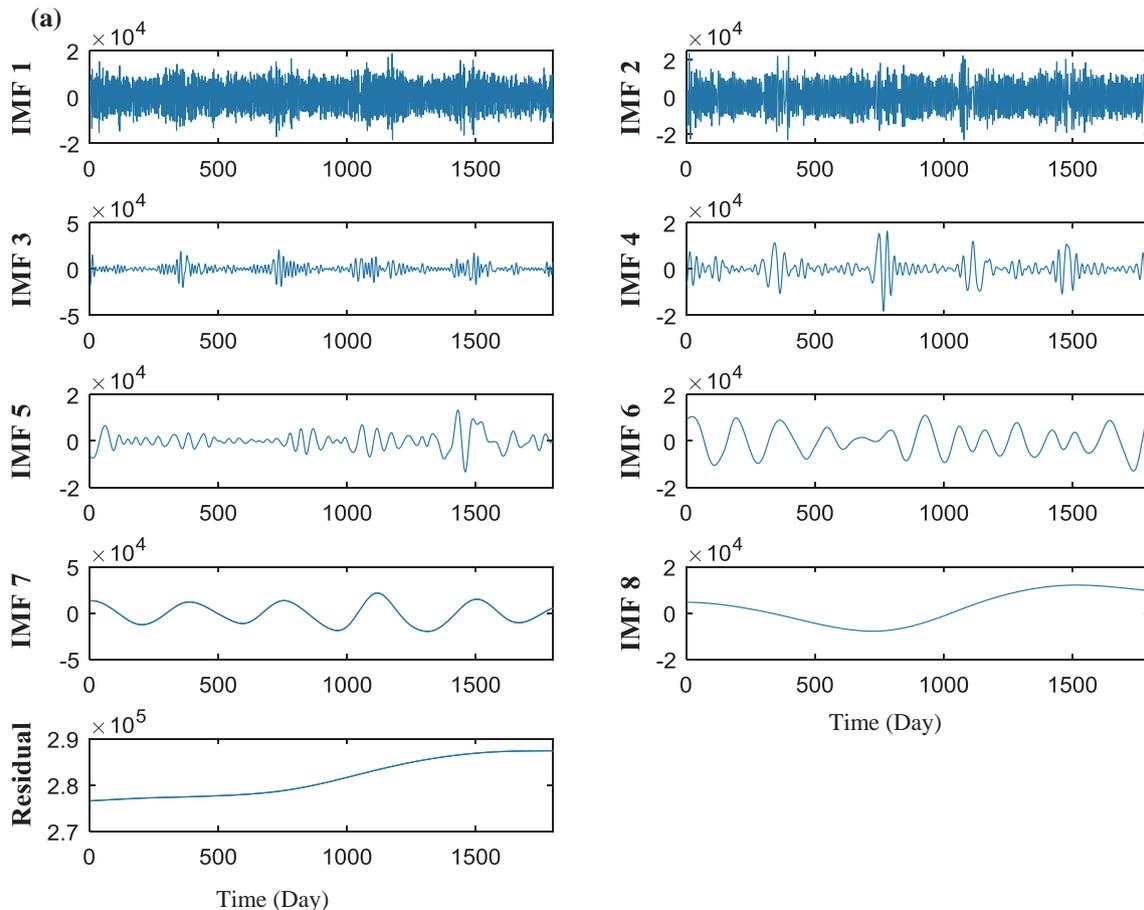


Fig. 2. (a) Intrinsic mode functions (IMFs) and residual constructed using ICEEMDAN for the whole week's data. (b) Statistical significant lags used for developing the models for IMFs and residual using the whole week's data.

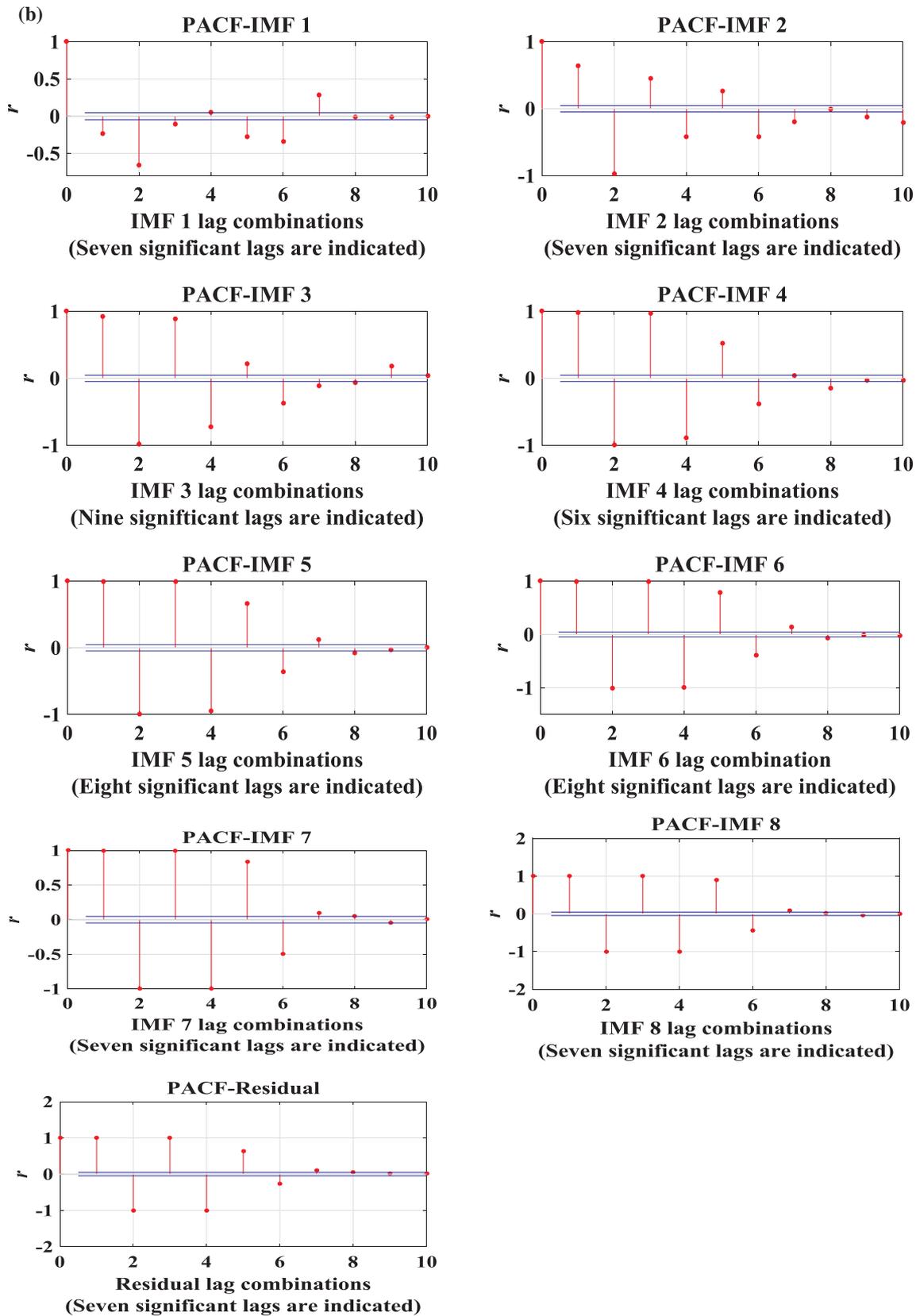


Fig. 2. (continued)

(non-ICEEMDAN data) with the statistically significant lagged variables through the partial autocorrelation function (PACF) were employed for the latter three models' development, whereas the ICEEMDAN was utilized to decompose the original data for the first three models before

applying the PACF. The non-stationary and non-linear time series was addressed by the ICEEMDAN technique that decomposed the original G data from the highest to the lowest frequencies named as IMF1 and residual, respectively [22]. A high-level flowchart of the hybrid and

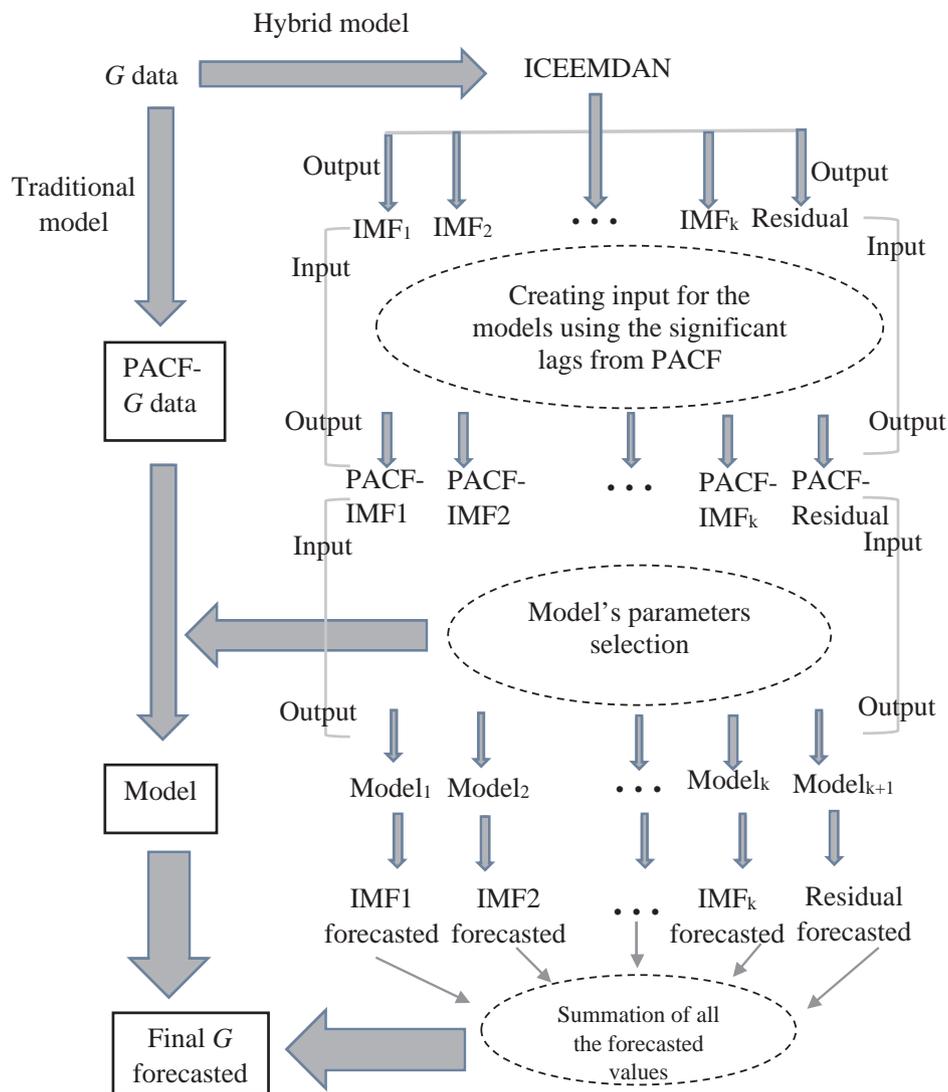


Fig. 3. The steps of the model development for both transitional and hybrid models.

traditional models is shown in Fig. 3. In this paper, different periods of G data (weekend, working days, whole week, public holiday and monthly) were decomposed using ICEEMDAN. Fig. 2a illustrates the whole week horizon where the datasets were analyzed into eight IMFs and a residual, while the statistically significant lags utilizing PACF are depicted in Fig. 2b for the IMFs and residual components. The number of IMFs for the various forecasting horizons is shown in Table 1.

As there is no single method for splitting data into training, validation and testing [8], the data were divided into subsets as follows: 70% for training, 15% for validation and 15% for testing (Table 1). Due to the higher frequency in G data, the training data were normalized, which converted the data between zero and one, before the models were run using Eq. (7) to avoid input values with large numeric ranges [15,64]:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

where x_{min} and x_{max} are the minimum and maximum values of x , respectively, while x_{norm} is the normalized value of x .

The models were developed in this study using MATLAB software running over an Intel i7, 3.60 GHz processor. The mechanism of the ICEEMDAN-models can be summarized in four steps. Firstly, the ICEEMDAN was used to decompose the original data into the IMFs and residual subsets (Fig. 2a). Secondly, each subset was reconstructed to

the number of significant lagged values using PACF (Fig. 2b). Those values were then individually employed to forecast each IMF and residual components that constructed the ICEEMDAN-models. Finally, the forecasted G utilizing the ICEEMDAN-model was obtained by aggregating those forecasted values. The stages above are clearly shown in Fig. 3.

MATLAB-based Libsvm toolbox (version 3.22), developed by Chang and Lin [65], was used to build the SVR model in this study. The radial basis function (RBF), which is expressed in Eq. (8), was employed in developing the SVR model [66]. Non-linearities between the objective and predictor variables were examined by the RBF [15,64], which considers a faster function in the training part [67–69]. The sensitive step to develop an accurate SVR model in the training datasets was to identify the 3D parameters (C , σ and ϵ) [70]. For this reason, the PSO algorithm, (Section 2.2), was used to select the near global optimum parameters based on the smallest value of MSE . In this paper, the PSO parameters, c_1 and c_2 , were tested according to Alam [59] with values in the range of [2, 2.05], whereas ω_{min} and ω_{max} were 0.4 and 0.9, respectively [16]. The initial population (N) was explored between 10 and 100 in increments of 10 [59], and the maximum number of iterations (T_{max}) was extended further [59,60,71] to be between 200 and 1000 in increments of 100. Fig. 4a illustrates the public holiday forecast horizon using PACF-Residual data. As a result, the near global optimum SVR model reached an MSE of $\cong 7.4210^{-5} \text{ MW}^2$ for $C = 7.809$, $\sigma = 0.01$

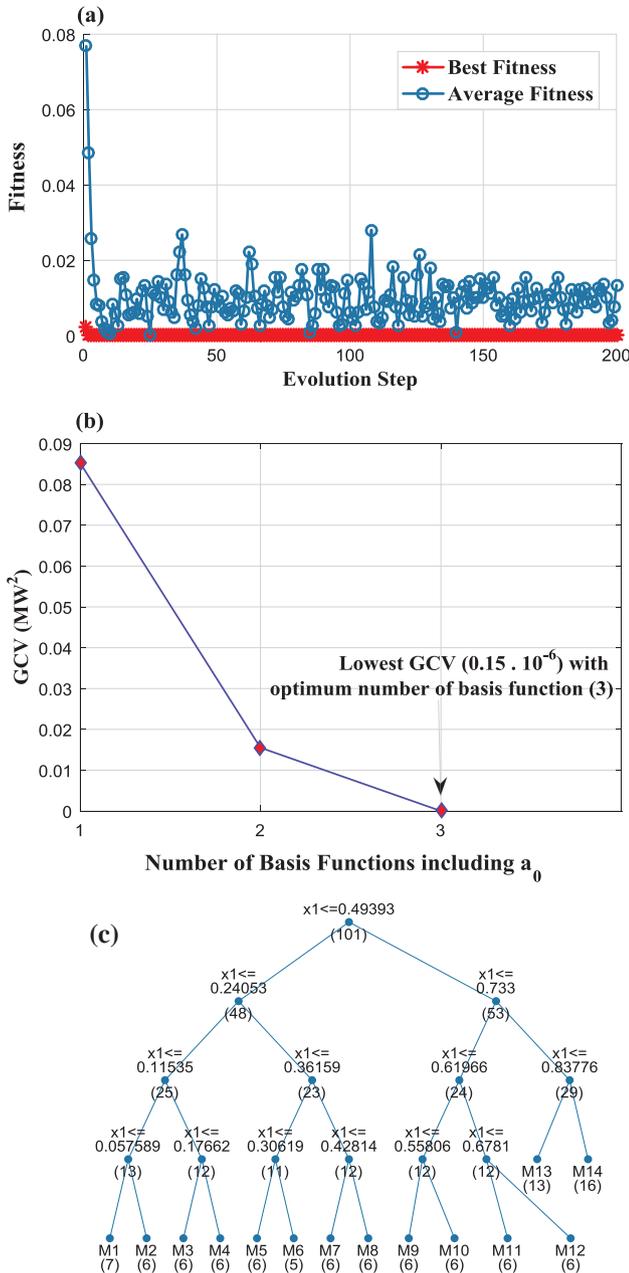


Fig. 4. Illustration of ICEEMDAN-SVR, ICEEMDAN-MARS and ICEEMDAN-M5 model tree parameters and fitting for public holiday forecast horizon using PACF-residual data in the training period. (a) PSO fitness curve to identify the basis function (C), kernel widths (σ), and loss function (ε). (b) Best number of basis functions based on the lowest generalized cross-validation statistic (GCV). $GCV (\approx 0.15 \cdot 10^{-6} \text{ MW}^2)$ with 2 basis functions and y-intercept, a_0 is indicated. (c) Tree generated by M5 model.

and $\varepsilon = 0.01$, based on $c_1 = 2$, and $c_2 = 2$, $N = 20$ and $T_{max} = 200$.

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (8)$$

where x_i and x_j are the inputs in the i th and j th respective dimensions and σ is the kernel width.

To evaluate the accuracy of the SVR model, MARS and M5 model trees were constructed with and without the ICEEMDAN method utilizing piecewise cubic and linear regression functions, respectively. For both models in the training phase, the software packages (version 1.13.0 for MARS and version 1.7.0 for the M5 model tree) were employed [72,73]. The best MARS model was selected based on the lowest Generalized Cross-Validation (GCV) [54], while the best M5 model tree

was selected based on the optimum number of decision-trees (M) that was identified by attaining the maximum splits for data attributes through minimizing the value of the standard deviation reduction (SDR) [74]. Fig. 4b–c demonstrates the optimum MARS and the M5 model tree utilizing PACF-residual data for the public holiday forecast horizon. As a result, $GCV \approx 0.15 \cdot 10^{-6} \text{ MW}^2$ including two basis functions and y-intercept a_0 for the MARS model (Fig. 4b) and $M = 14$ for M5 model tree (Fig. 4c) were obtained.

The models were validated using the following statistics (Eqs. (9) and (15) Section 3.3): root-mean square error (RMSE) and Legates and McCabe Index (E_{LM}). For all forecasting horizons, the ICEEMDAN-PSO-SVR model yielded the lowest RMSE and the highest E_{LM} , which indicated the best accuracy compared to the other models. Table 2 summarizes the evaluation of the models in the validation period.

3.3. Model performance evaluation

To establish whether or not the models were useful for G forecasting, this study adopted a wide range of statistical indicators: mean absolute error (MAE), root-mean square error (RMSE), relative error (%) based on MAE and RMSE values (MAPE and RRMSE), Willmott’s Index (WI), the Nash–Sutcliffe coefficient (E_{NS}), and the Legates and McCabe Index (E_{LM}) [69,75–83], represented below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2} \quad (9)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |G_i^{for} - G_i^{obs}| \quad (10)$$

$$RRMSE = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}}{\overline{G^{obs}}} \quad (11)$$

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^{i=n} \left| \frac{G_i^{for} - G_i^{obs}}{G_i^{obs}} \right| \quad (12)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (|G_i^{for} - \overline{G^{obs}}| + |G_i^{obs} - \overline{G^{obs}}|)^2} \right], \text{ and } 0 \leq WI \leq 1 \quad (13)$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (G_i^{obs} - \overline{G^{obs}})^2} \right], \text{ and } \infty \leq E_{NS} \leq 1 \quad (14)$$

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^{i=n} |G_i^{obs} - G_i^{for}|}{\sum_{i=1}^{i=n} |G_i^{obs} - \overline{G^{obs}}|} \right], \text{ and } (\infty \leq E_{LM} \leq 1) \quad (15)$$

where n is the total number of G -observed (or G -forecasted) values in the testing period, G_i^{for} and G_i^{obs} are the i th forecasted and observed values of G , respectively; $\overline{G^{for}}$ and $\overline{G^{obs}}$ are the means of forecasted and observed values, respectively.

Various combinations of performance indicators were used in this study (Eqs. (9)–(15)) as the model’s advantages and weaknesses could not be assessed by a single metric [24,80–82]. For the model accuracy when the error distribution is Gaussian, the RMSE is a more suitable metric than the MAE, otherwise, complementary evaluations can be represented by the MAE, RMSE, and their relative forms, MAPE and RRMSE [82]. A model was considered excellent when $RMSE_{\overline{G}} < 10\%$, good if the model satisfied $10\% < RMSE_{\overline{G}} < 20\%$, fair if it satisfied $20\% < RMSE_{\overline{G}} < 30\%$, and poor if $RMSE_{\overline{G}} > 30\%$ [69,84–86]. A model achieved better performance when the values of E_{NS} and WI were close to one [8]. However, the Legates and McCabe Index ($\infty \leq E_{LM} \leq 1$), which is the modified version of WI [80], provided greater precision than the classic WI when relatively high values were predicted as a result of squaring of differences [76,78,87].

Table 2

Root-mean square error (*RMSE*; MW), and Legates & McCabes Index (*E_{LM}*) for the weekend, working days, whole week, public holiday and monthly forecast horizons in the validation dataset. The most accurate model is boldfaced.

Model	<i>RMSE</i> (MW)	<i>E_{LM}</i>	<i>RMSE</i> (MW)	<i>E_{LM}</i>	<i>RMSE</i> (MW)	<i>E_{LM}</i>
<i>Weekend forecast horizon</i>			<i>Working days forecast horizon</i>		<i>Monthly forecast horizon</i>	
ICEEMDAN-PSO-SVR	4827.45	0.81	4151.73	0.86	175413.10	0.59
ICEEMDAN-M5 model tree	6994.77	0.73	5088.39	0.84	208829.09	0.56
ICEEMDAN-MARS	5331.29	0.78	3934.86	0.87	184813.41	0.56
PSO-SVR	12939.33	0.47	10897.64	0.63	382025.28	0.08
M5 model tree	15603.05	0.38	11915.72	0.61	492618.18	−0.24
MARS	13118.66	0.46	9565.75	0.66	386776.39	0.01
<i>Whole week forecast horizon</i>			<i>Public holiday forecast horizon</i>			
ICEEMDAN-PSO-SVR	3800.23	0.87	7447.87	0.33		
ICEEMDAN-M5 model tree	6124.74	0.82	10438.52	0.11		
ICEEMDAN-MARS	3931.98	0.86	7793.04	0.24		
PSO-SVR	15489.32	0.53	15454.03	−0.33		
M5 model tree	16048.41	0.44	21105.26	−0.87		
MARS	13340.56	0.50	17838.41	−0.57		

Table 3

The performance of the models in the test period for the short-term measured by Willmott's index (*WI*), Nash–Sutcliffe coefficient (*E_{NS}*), root-mean square error (*RMSE*; MW), and mean absolute error (*MAE*; MW). The most accurate model is boldfaced.

Model	<i>WI</i>	<i>E_{NS}</i>	<i>RMSE</i> (MW)	<i>MAE</i> (MW)	<i>WI</i>	<i>E_{NS}</i>	<i>RMSE</i> (MW)	<i>MAE</i> (MW)
<i>Weekend forecast horizon</i>				<i>Working days forecast horizon</i>				
ICEEMDAN-PSO-SVR	0.95	0.84	5650.53	3867.00	0.98	0.93	3405.50	2336.95
ICEEMDAN-M5 model tree	0.88	0.62	8858.36	5106.96	0.97	0.88	4521.32	3230.30
ICEEMDAN-MARS	0.88	0.67	8198.56	4675.83	0.98	0.93	3506.96	2481.16
PSO-SVR	0.67	0.16	13067.18	7872.27	0.88	0.61	8178.32	5816.27
M5 model tree	0.68	0.02	14144.12	8978.01	0.84	0.44	9821.98	6993.60
MARS	0.66	0.18	12904.16	8403.72	0.88	0.62	8051.44	5760.05
<i>Whole week forecast horizon</i>				<i>Public holiday forecast horizon</i>				
ICEEMDAN-PSO-SVR	0.99	0.95	3619.29	2642.16	0.92	0.76	8292.57	6092.21
ICEEMDAN-M5 model tree	0.96	0.86	6172.75	3828.10	0.87	0.55	11411.28	8208.92
ICEEMDAN-MARS	0.98	0.94	3924.59	2945.40	0.90	0.72	9074.59	7789.78
PSO-SVR	0.88	0.64	9708.61	7273.31	0.62	−0.16	19099.48	15809.16
M5 model tree	0.77	0.20	14541.30	11168.70	0.59	−0.83	23968.12	20400.46
MARS	0.84	0.48	11683.75	9316.54	0.67	0.14	16377.51	12336.55

Table 4

The relative root-mean square error (*RRMSE*%), mean absolute percentage error (*MAPE*%) and Legates & McCabes Index (*E_{LM}*) for the models in the test datasets for the short-term. The most accurate model is boldfaced.

Model	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>E_{LM}</i>	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>E_{LM}</i>
<i>Weekend forecast horizon</i>			<i>Working days forecast horizon</i>			
ICEEMDAN-PSO-SVR	1.34	2.01	0.65	0.77	1.13	0.77
ICEEMDAN-M5 model tree	1.77	3.15	0.53	1.07	1.50	0.68
ICEEMDAN-MARS	1.61	2.92	0.57	0.82	1.16	0.75
PSO-SVR	2.69	4.65	0.28	1.92	2.71	0.42
M5 model tree	3.11	5.03	0.18	2.31	3.26	0.30
MARS	2.90	4.59	0.23	1.90	2.67	0.43
<i>Whole week forecast horizon</i>			<i>Public holiday forecast horizon</i>			
ICEEMDAN-PSO-SVR	0.89	1.22	0.79	2.14	2.97	0.55
ICEEMDAN-M5 model tree	1.29	2.09	0.70	2.89	4.09	0.39
ICEEMDAN-MARS	0.99	1.33	0.77	2.78	3.25	0.42
PSO-SVR	2.46	3.28	0.43	5.55	6.87	−0.13
M5 model tree	3.77	4.92	0.12	7.18	8.62	−0.46
MARS	3.15	3.95	0.27	4.30	5.89	0.12

4. Results and discussion

4.1. Short term forecasting horizon

The hybrid SVR model, combined with PSO and ICEEMDAN, was evaluated using the statistical metrics presented in Section 3.3. The performance of the ICEEMDAN-PSO-SVR model for the short and long-term forecast horizons was compared with ICEEMDAN-MARS and ICEEMDAN-M5 model tree, as well as with the traditional methods of PSO-SVR, MARS and M5 model tree in the testing period. The results of

the comparisons are presented and discussed below.

The predictive capability of the three ICEEMDAN models for the weekend, working days, whole week and public holiday in the testing period is summarized in Table 3. For all short-term forecasting horizons, the ICEEMDAN models yielded better performances (lowest *RMSE*, and *MAE*, as well as largest *WI* and *E_{NS}*) compared to the traditional models indicating that the ICEEMDAN was a robust method, which addressed non-stationary data. This was clearly shown when the highly non-stationary data for the public holiday was used to build and test the models (also see Fig. 1). For example, the *WI* and *RMSE/MAE*

were 48.39% and 130.32%/159.50% greater and lower, respectively, for the ICEEMDAN-PSO-SVR model than for the PSO-SVR model. Comparing the ICEEMDAN-PSO-SVR model’s accuracy statistic ($E_{NS} = 0.76$) with the PSO-SVR model ($E_{NS} = -0.16$) supported this result. Similarly, for the weekend, working days and whole week forecasting

horizons, the advantage of the ICEEMDAN-MARS model and ICEEMDAN-M5 model tree over the classic MARS model and M5 model tree are indicated in Table 3.

Using the weekend dataset, a comparison among the three advanced models indicated that the ICEEMDAN-PSO-SVR ($WI = 0.95$, $E_{NS} = 0.84$,

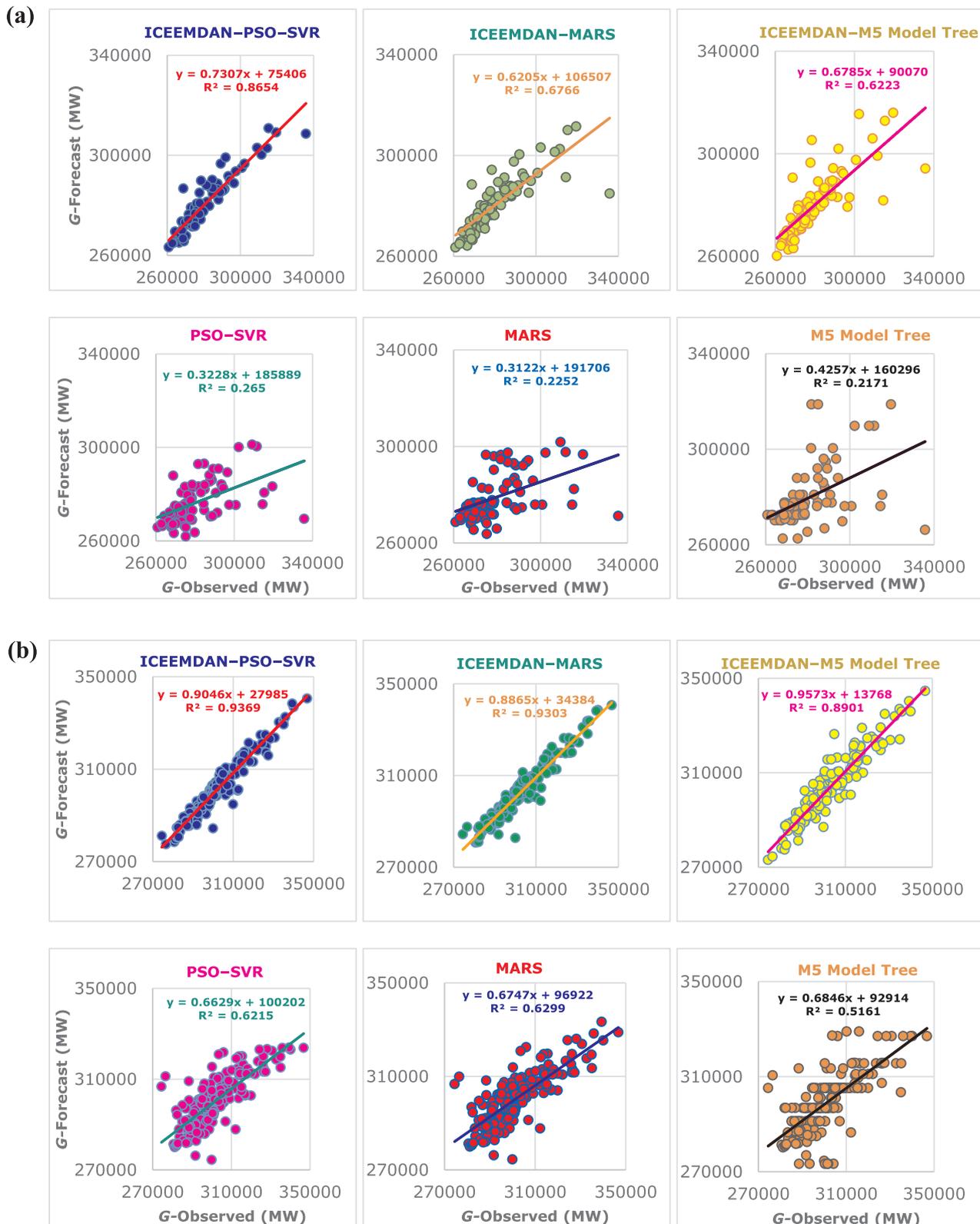


Fig. 5. Scatterplot of the G-forecasted vs. G-observed of electricity demand data in the testing period for the (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons using the six models. The equation of the linear regression line and the coefficient of determination are incorporated.

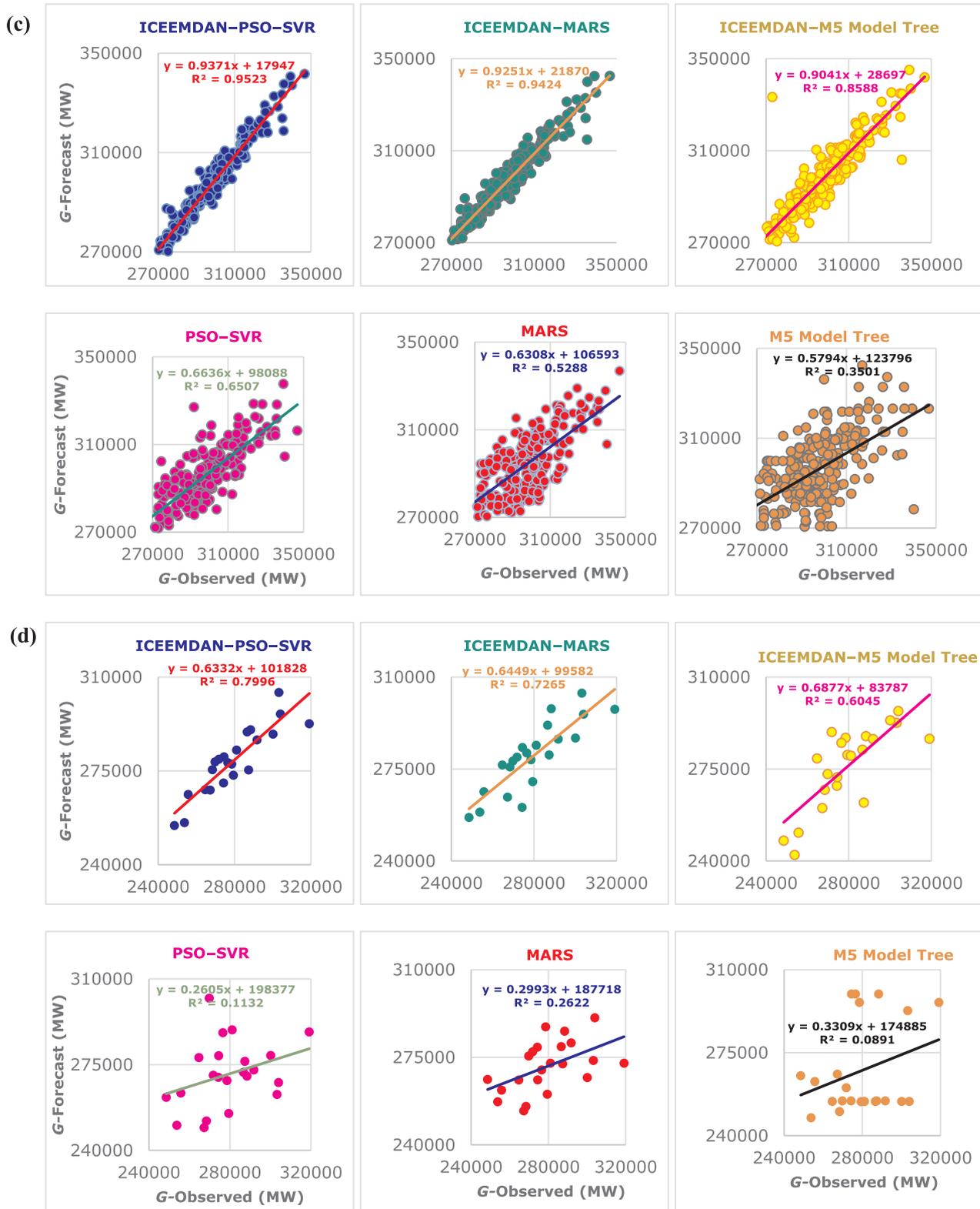


Fig. 5. (continued)

RMSE = 5650.53 MW, and MAE = 3867.00 MW) outperformed both the ICEEMDAN-M5 model tree ($WI = 0.88$, $E_{NS} = 0.62$, $RMSE = 8858.36$ MW, and $MAE = 5106.96$ MW) and the ICEEMDAN-MARS ($WI = 0.88$, $E_{NS} = 0.67$, $RMSE = 8198.56$ MW, and $MAE = 4675.83$ MW). Similar performances were also evident for the working days, whole week and public holiday forecasting horizons (Table 3). Thus, the best accuracy for G forecasting was achieved by the

ICEEMDAN-PSO-SVR.

The best values of MAPE, RRMSE and E_{LM} , which show the ‘goodness-of-fit’ between the model-forecasted and observed data (e.g., [8,69,74,85]), were also yielded by the ICEEMDAN-PSO-SVR model for the four forecasting horizons (Table 4). For the whole week forecast horizon, MAPE, RRMSE and E_{LM} were 0.89%, 1.22% and 0.79, respectively, compared to 1.29%, 2.09% and 0.70 for the ICEEMDAN-M5

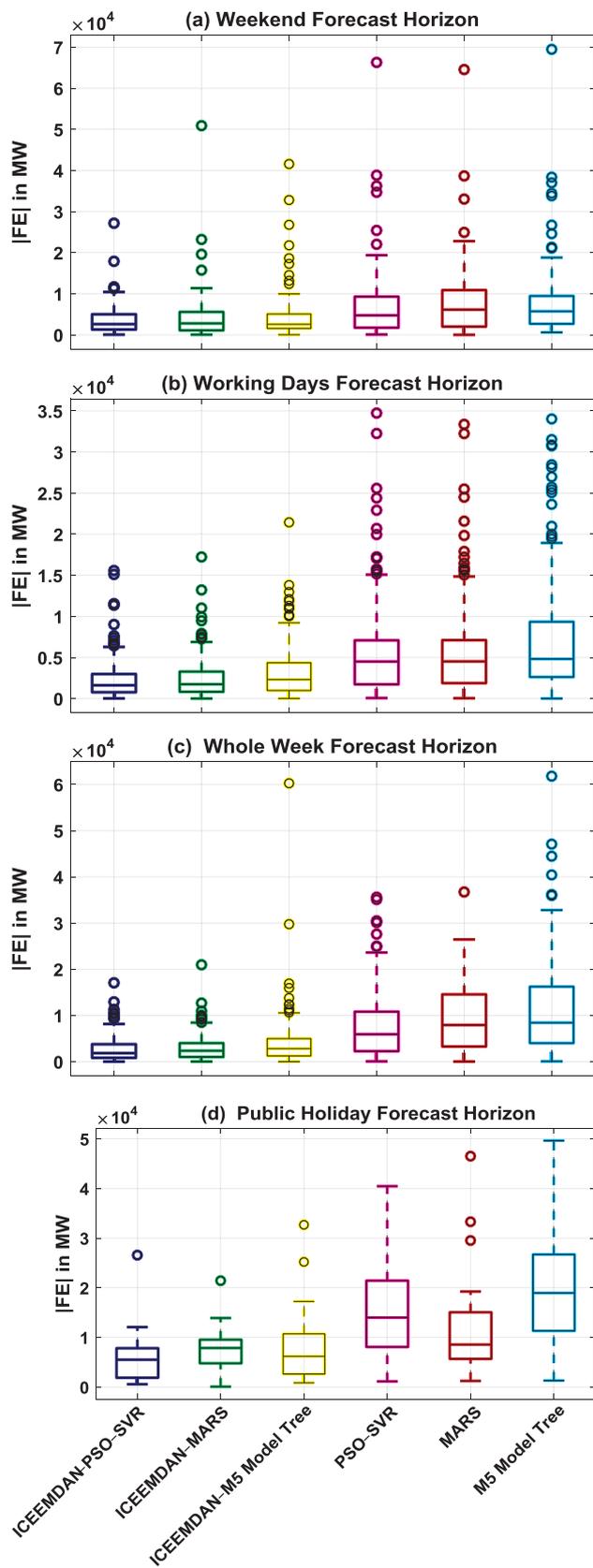


Fig. 6. Boxplots of the absolute forecasted error, $|FE| = |G_{FOR,t} - G_{OBS,t}|$ in the testing period for the (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons using the six models.

model tree and 0.99%, 1.33% and 0.77 for the ICEEMDAN-MARS model. The performance of the classic models showed double values of MAPE and RRMSE and half values of E_{LM} , and were not satisfactory. Table 4 contains further details of metrics for all forecast horizons.

The scatterplots of G_i^{for} vs. G_i^{obs} for the weekend, working days, whole week and public holiday in the testing period are depicted in Fig. 5(a–d). To demonstrate the relationship between G_i^{for} and G_i^{obs} , the least squares regression line, $y = G_i^{for} = a'G_i^{obs} + b'$, where the constant a' and the y-intercept b' were utilized to outline the model's accuracy [8], with the correlation of determination (R^2), was employed. For the weekend forecast horizon (Fig. 5a), the ICEEMDAN-PSO-SVR model ($a' = 0.73$, $b' = 75,406$ and $R^2 = 0.87$) was significantly better than the ICEEMDAN-MARS ($a' = 0.62$, $b' = 106,507$ and $R^2 = 0.68$) and the ICEEMDAN-M5 model tree ($a' = 0.68$, $b' = 90,070$ and $R^2 = 0.62$). According to Fig. 5b and c, the ICEEMDAN-PSO-SVR model ($R^2 = 0.94$ and 0.95) outperformed both the ICEEMDAN-MARS ($R^2 = 0.93$ and 0.94) and the ICEEMDAN-M5 model tree ($R^2 = 0.89$ and 0.86) by a small margin for the working days and whole week datasets, respectively. The proposed model's R^2 was also greater than those for the ICEEMDAN-MARS and ICEEMDAN-M5 model tree scenarios by 9.59% and 33.33%, respectively, for the public holiday data (Fig. 5d). Poor results were generated when the standard models (PSO-SVR, MARS and M5 model tree) were employed to forecast G data.

The absolute value of the forecasted error statistics $|FE| = |G_i^{for} - G_i^{obs}|$ was calculated to design the boxplots, as shown in Fig. 6. The lower and upper lines of the boxplot denoted the first and third quartiles (25th and 75th percentiles), respectively, and the median value (50th percentile) was represented by the central line. Additionally, two horizontal lines were drawn out from the first and third quartiles to the smallest and largest non-outliers, respectively. It was apparent that the non-ICEEMDAN model errors were dramatically larger for all the short-term datasets. In accordance with the previous results (Tables 3 and 4, and Fig. 5), the magnitude of G forecasted errors including the lower, median and upper quartiles were relatively small when the ICEEMDAN-PSO-SVR model was employed compared with the ICEEMDAN-MARS and ICEEMDAN-M5 model tree techniques. Although the maximum point of $|FE|$ values generated by ICEEMDAN-PSO-SVR was larger than ICEEMDAN-MARS in the case of the public holiday forecast horizon (Fig. 6d), the former model was more accurate with respect to the first, second and third quartiles. Overall, a significantly greater accuracy was attained by the ICEEMDAN-PSO-SVR model than the other models.

The Taylor diagram, which determined the angular location to the inverse cosine of the correlation coefficient [88], was plotted in Fig. 7 to show the most close model to the observed data in the testing period. The correlation coefficient (r), on the radial axis, and the standard deviation on the polar axis were used together in this figure to adjust the closest fit model to the predictors. For all the forecasting scenarios, the ICEEMDAN-PSO-SVR, which generated the greatest value of r , yielded the closest forecasted to the observed data. Again, the modelled data, generated by the classic models, were far away from the predictors. Thus, by combining ICEEMDAN with the PSO-SVR model, the distance between the observed and forecasted in the testing datasets was reduced by more than half.

With respect to the percentage of $|FE|$, the empirical cumulative distribution function (ECDF) was plotted at different forecasting abilities (Fig. 8). According to this figure, the ICEEMDAN-PSO-SVR method was slightly better than ICEEMDAN-MARS for weekend, working days and whole week forecasting horizons, and both models were superior to the ICEEMDAN-M5 model tree (Fig. 8(a–c)). Based on the percentage of errors in the begging bracket (0 to $\pm 0.5 \cdot 10^4$ MW) for the public holiday scenario, Fig. 8d clearly confirms that the ICEEMDAN-PSO-SVR was the most responsive model in forecasting G data (0.45%) compared to 0.36% for the ICEEMDAN-M5 model tree and 0.27% for the ICEEMDAN-MARS model. Inferior performances were demonstrated when

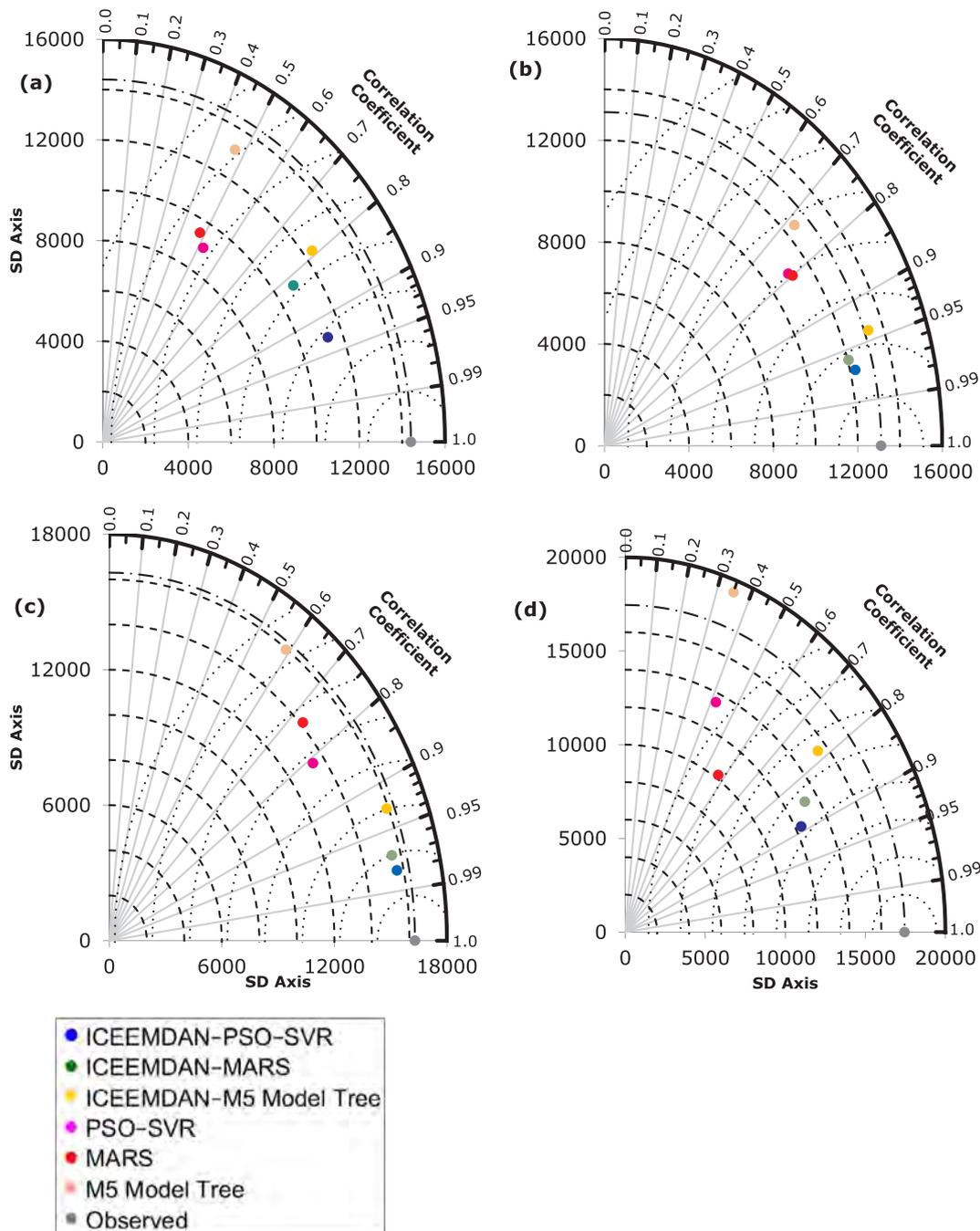


Fig. 7. Taylor diagram showing the correlation coefficient between G -observed and G -forecasted and standard deviation for each model utilizing (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons.

the non-ICEEMDAN mechanisms were utilized. Therefore, the highest performance with the lowest $|FE|$ resulted from the PSO-SVR model based on the current decomposition skill (ICEEMDAN).

4.2. Long-term forecasting horizon

Tables 5 and 6 present the details regarding the model performances for the monthly time period. Based on a comparison between G -observed and G -forecasted, the performance of the main model ($E_{NS} = 0.70$, $RRMSE = 3.18\%$, and $MAPE = 2.18\%$) was more accurate than the ICEEMDAN-M5 model tree ($E_{NS} = 0.65$, $RRMSE = 3.45\%$, and $MAPE = 2.45\%$), and the ICEEMDAN-MARS ($E_{NS} = 0.69$, $RRMSE = 3.26\%$, and $MAPE = 2.26\%$) by a small margin, but it was much better than the classic models.

Figs. 9 and 10 demonstrate the boxplot and Taylor diagram, respectively, while Fig. 11 shows the magnitude of the absolute value of relative errors accumulated over the monthly timescale for the six forecasting criteria in the testing period. Although the forecasting skills generated by the ICEEMDAN-forms were slightly different, the lowest three quartiles (25th, 50th and 75th percentiles) and the top value of r were obtained by the ICEEMDAN-PSO-SVR (Figs. 9 and 10). This performance was confirmed by the low magnitude of relative forecasted errors (Fig. 11a). For example, the relative error in September was smaller by almost 50% for the ICEEMDAN-PSO-SVR model than for any other models. On the other hand, the classic models had large magnitudes of error, such as between November and January for both PSO-SVR and MARS and December to April for the M5 model tree. Overall, the PSO-SVR model provided better performance, including lower error

statistics (Figs. 9 and 11) and higher correlation coefficients (Fig. 10) when it was combined with the ICEEMDAN technique.

Given that the ICEEMDAN-PSO-SVR model provides better results than other new and traditional data driven approaches, this method could provide a very useful and accurate demand forecasting tool (for daily or monthly lead times) that can support the National Electricity Market in Australia (e.g., AEMO). This could improve their electric energy generation, facilitate efficient decision-making in the Electricity Market and infrastructure services, and improve network planning for the short and long-term forecasting horizons. And finally, generating accurate electricity load forecasts using the ICEEMDAN-PSO-SVR model can also be used to forecast electricity prices

that in turn can help both power producers and consumers develop and explore various strategies.

5. Limitations and future work

Although this study was the first to develop and evaluate the ICEEMDAN-PSO-SVR model for G forecasting, several limitations should be addressed in future research studies. The ICEEMDAN technique could be extended into two-layer decomposition scenarios when the variational mode decomposition (VMD) is applied to further decompose IMF1 (and other related high frequency IMF) generated by ICEEMDAN, into a number of variational modes [22,23]. Therefore, the forecast

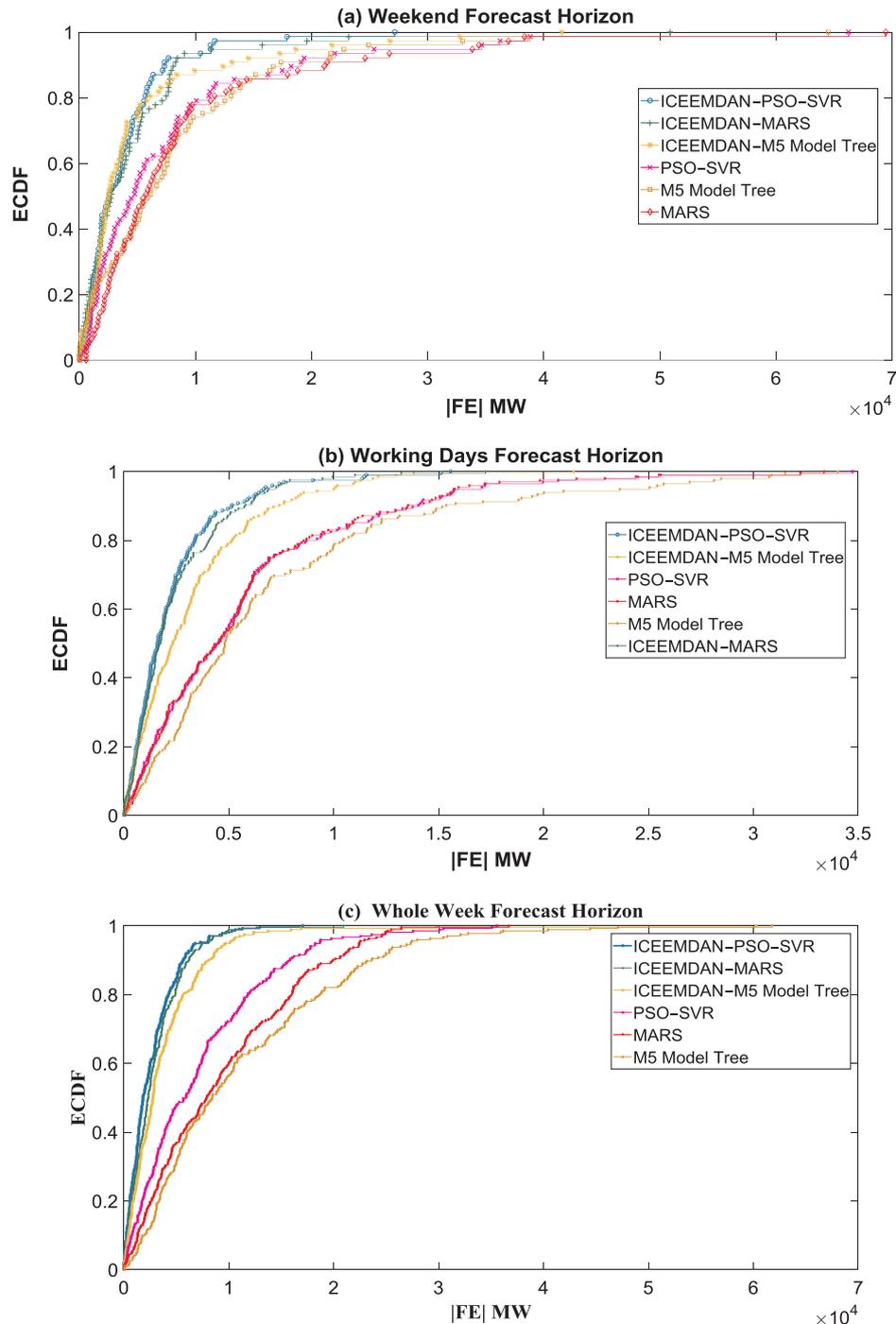


Fig. 8. Empirical cumulative distribution function (ECDF) of the forecast error, $|FE|$ in the testing period for the (a) weekend (b) working days (c) whole week (d) public holiday forecast horizons using the six models.

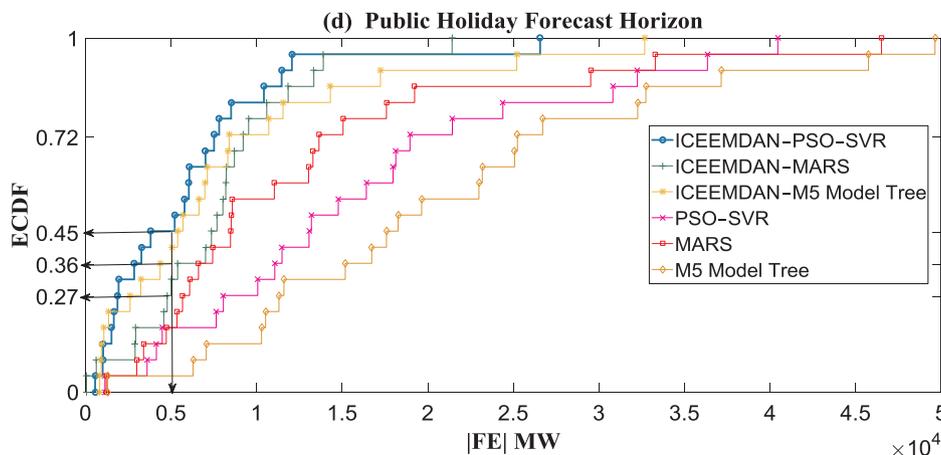


Fig. 8. (continued)

Table 5

The performance of the models in the test period for the long-term (monthly) measured by Willmott’s index (WI), Nash–Sutcliffe coefficient (E_{NS}), root-mean square error ($RMSE$; MW), and mean absolute error (MAE ; MW). The most accurate model is boldfaced.

Model	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)
<i>Monthly forecast horizon</i>				
ICEEMDAN-PSO-SVR	0.91	0.70	282050.71	194002.42
ICEEMDAN-M5 model tree	0.88	0.65	306344.02	217165.21
ICEEMDAN-MARS	0.91	0.69	289159.49	199727.16
PSO-SVR	0.67	0.14	481159.00	394404.91
M5 model tree	0.63	-0.14	551823.54	438279.74
MARS	0.61	0.12	485541.53	394284.68

Table 6

The relative root-mean square error ($RRMSE\%$), mean absolute percentage error ($MAPE\%$) and Legates & McCabe’s Index (E_{LM}) for the models in the test datasets for the long-term (monthly) horizon. The most accurate model is boldfaced.

Model	$MAPE$ (%)	$RRMSE$ (%)	E_{LM}
<i>Monthly forecast horizon</i>			
ICEEMDAN-PSO-SVR	2.18	3.18	0.56
ICEEMDAN-M5 model tree	2.45	3.45	0.50
ICEEMDAN-MARS	2.26	3.26	0.54
PSO-SVR	4.38	5.42	0.10
M5 model tree	4.86	6.21	-0.003
MARS	4.35	5.47	0.10

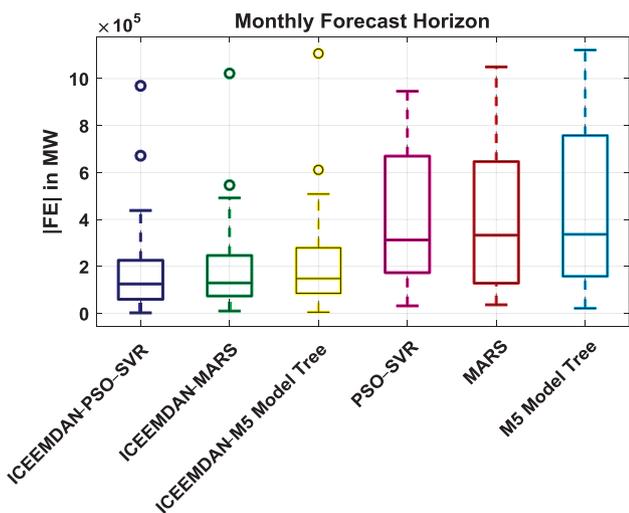


Fig. 9. Boxplots of the absolute forecasted error, $|FE| = |G_{FOR,t} - G_{OBS,t}|$ in the testing period for the monthly forecast horizons using the six models.

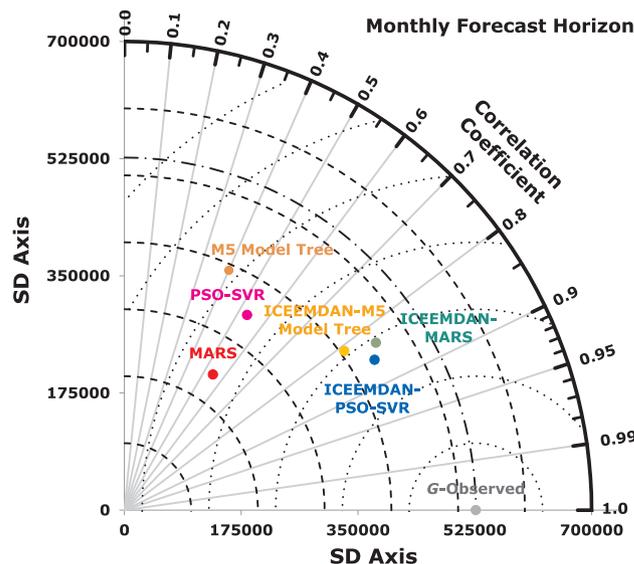


Fig. 10. Taylor diagram showing the correlation coefficient between G -observed and G -forecasted and standard deviation for each model utilizing monthly forecast horizon.

accuracy may improve when ICEEMDAN-VMD is considered. However, Wang et al. [22,23] used the fast EEMD and CEEMD with VMD without considering the ICEEMDAM algorithm. Hence, the combination of ICEEMDAN with VMD has not yet been explored for any type of data forecasting yet. On the other hand, the disadvantage of the ICEEMDAN method is that it produces different numbers of IMFs that lead to different lengths of data when two or several predictors (e.g. climate variables), which can be used to develop a model, are decomposed by the proposed method in this research paper. In addition, despite the PSO demonstrating good performance in selecting the 3D parameters of the SVR model, it would be advantageous to include an improved PSO model [61], multi-swarm PSO (MSPSO) model [89], or develop a genetic algorithm [90] that could help to identify appropriate parameters for the SVR model. As well, as the PSO algorithm, which was used in this study, is slow and takes a longer time to produce the parameters of the SVR model, the above alternative methods could solve this issue. Furthermore, only the historical G data were used to develop the models in this study. Climate data (e.g., temperature, rainfall, humidity and solar radiation), which can also influence G (e.g., [12,91]), could also be incorporated into the G forecasting environment. As a relationship between G data and temperature is clearly shown in winter and summer (not shown here), it would be useful to develop a forecasting model utilizing both the seasonal G and the related climate input data. Finally, the ICEEMDAN-PSO-SVR approach could be

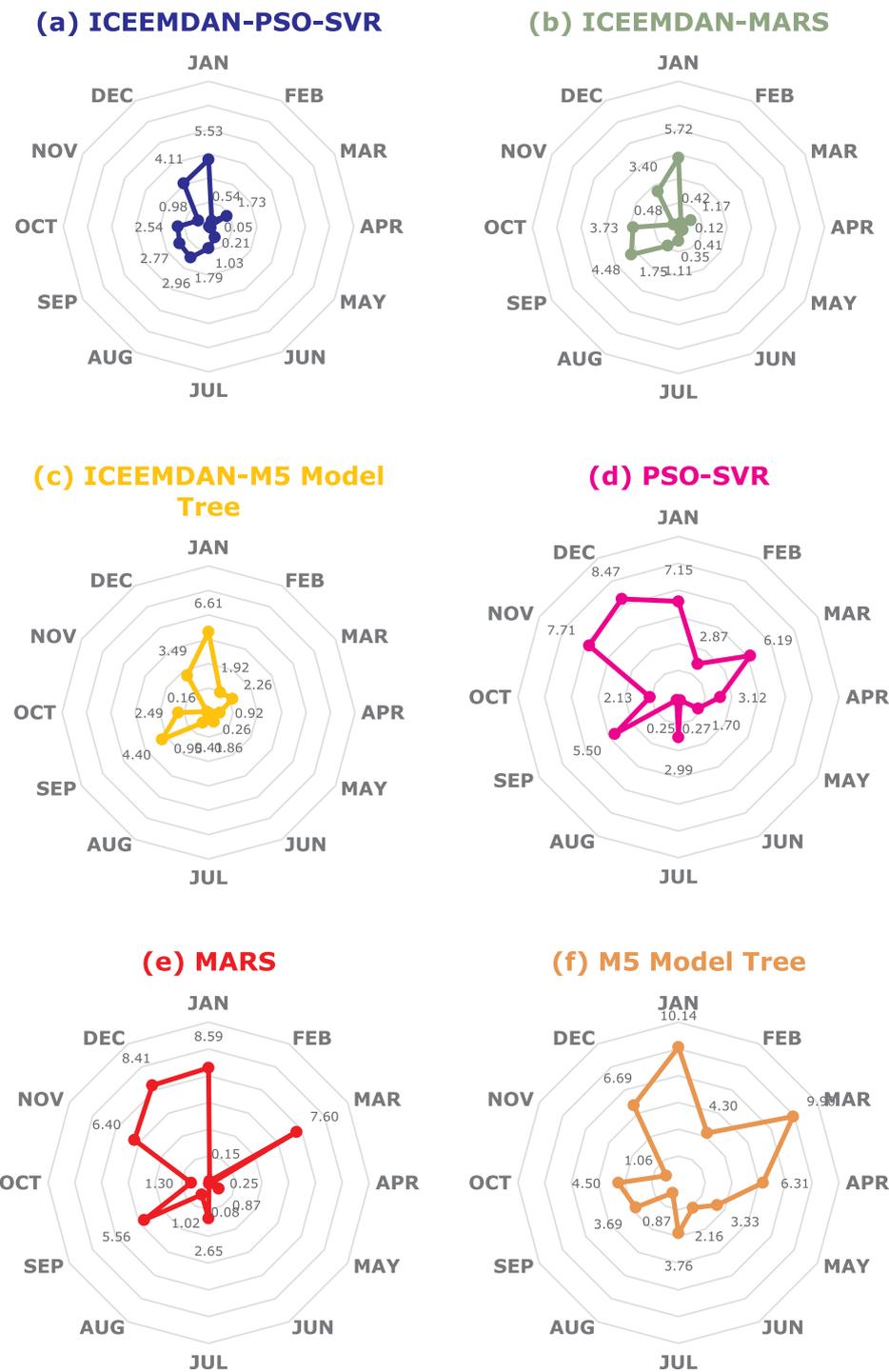


Fig. 11. Relative error analysis over the monthly forecast horizon in the testing period. The six models and the magnitude of relative forecasted errors are indicated on top of each plot and on the radial axis, respectively.

improved with ensemble-based uncertainty testing via a bootstrapping procedure and the Bayesian Model Averaging (BMA) framework [92,93].

6. Concluding remarks

In this paper, a hybrid ICEEMDAN-PSO-SVR model was proposed for short-term (i.e., weekend, working days, whole week, and public holiday), and long-term (i.e., monthly) *G* forecasting horizons in Queensland, Australia where the models used data from the Australian Energy Market Operator (AEMO). The non-stationary time series data

were addressed via the ICEEMDAN technique that was used to decompose the data into IMFs and residual subsets (Table 1 and Fig. 2a). The PSO algorithm was employed to determine the 3D parameters, which are the regulation function (*C*), kernel widths (σ) and loss function (ϵ), of the SVR model.

Five other data-driven benchmarking models, (i.e., ICEEMDAN-MARS, ICEEMDAN-M5 model tree, PSO-SVR, MARS and M5 model tree), were also developed in this research. These models were compared with the proposed ICEEMDAN-PSO-SVR model using several statistical score metrics: mean absolute error (*MAE*), root-mean square error (*RMSE*), relative *MAE* and *RMSE* (%), Willmott's Index (*WI*), the

Nash-Sutcliffe coefficient (E_{NS}), and the Legates and McCabe Index (E_{LM}) computed in the independent test dataset. It was found that the hybrid ICEEMDAN-PSO-SVR model yielded the most accurate forecasting results for all forecasting horizons explored in this study.

Although this study showed the advantages of the ICEEMDAN-PSO-SVR model over other models, some limitations, such as not using a two-layer decomposition by combining the ICEEMDAN with VMD, not optimizing the SVR parameters by MSPSO, and not utilizing seasonal G and climate data, could be addressed in follow-up research studies. The aim of our future work is to explore those mechanisms based on data from the University of Southern Queensland (USQ) to reduce error in energy forecasting and diminish the risk to energy policies due to over or under estimation of G .

Acknowledgments

Foremost, we would like to thank the Australian Energy Market Operator (AEMO) which provided all the required data. We are very grateful to the Ministry of Higher Education and Scientific Research in the Government of Iraq for funding the first author's PhD project. We also thank the University of Southern Queensland for all the facilities, tools and access to different academic services. Our gratitude and appreciation also go to Ms Mela Aryal for proof-reading the paper. Finally we thank both reviewers and the Editor-In-Chief for his constructive comments that have improved the clarity of this paper.

References

- Hu Z, Bao Y, Xiong T. Electricity load forecasting using support vector regression with memetic algorithms. *Sci. World J.* 2013;2013.
- Kaytez F, Taplamacioglu MC, Cam E, Hurdalac F. Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. *Int J Electr Power Energy Syst* 2015;67:431–8.
- Fan S, Chen L. Short-term load forecasting based on an adaptive hybrid method. *IEEE Trans Power Syst* 2006;21:392–401.
- Bunn D, Farmer ED. Comparative models for electrical load forecasting; 1985.
- Haida T, Muto S. Regression based peak load forecasting using a transformation technique. *IEEE Trans Power Syst* 1994;9:1788–94.
- Contreras J, Espinola R, Nogales FJ, Conejo AJ. ARIMA models to predict next-day electricity prices. *IEEE Transactions Power Syst* 2003;18:1014–20.
- Sözen A, Ali Akçayol M. Modelling (using artificial neural-networks) the performance parameters of a solar-driven ejector-absorption cycle. *Appl Energy* 2004;79:309–25.
- Deo RC, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 2016;168:568–93.
- Singh AK, Ibraheem SK, Muazzam M, Chaturvedi D. An overview of electricity demand forecasting techniques. *Network Complex Syst* 2013;3:38–48.
- Etemad-Shahidi A, Mahjoobi J. Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Eng* 2009;36:1175–81.
- Wang L, Kisi O, Zounemat-Kermani M, Zhu Z, Gong W, Niu Z, et al. Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. *Int J Climatol* 2017;37:1141–55.
- Sigauke C, Chikobvu D. Daily peak electricity load forecasting in South Africa using a multivariate non-parametric regression approach. *ORION* 2010;26:97.
- Yu P-S, Chen S-T, Chang I-F. Support vector regression for real-time flood stage forecasting. *J Hydrol* 2006;328:704–16.
- Türkyay BE, Demren D. Electrical load forecasting using support vector machines. 2011 7th International Conference on Electrical and Electronics Engineering (ELECO). IEEE; 2011. p. 1-49–53.
- Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification; 2003.
- Li X, Li C. Improved CEEMDAN and PSO-SVR modeling for near-infrared non-invasive glucose detection. *Comput Math Methods Med* 2016;2016.
- Wu C-H, Tzeng G-H, Goo Y-J, Fang W-C. A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert Syst Appl* 2007;32:397–408.
- Xie W, Li Y, Ma Y. Breast mass classification in digital mammography based on extreme learning machine. *Neurocomputing* 2016;173:930–41.
- Robinson J, Sinton S, Rahmat-Samii Y. Particle swarm, genetic algorithm, and their hybrids: optimization of a profiled corrugated horn antenna. 2002 IEEE Antennas and Propagation Society International Symposium. IEEE; 2002. p. 314–7.
- Deo RC, Samui P, Kim D. Estimation of monthly evaporative loss using relevance vector machine, extreme learning machine and multivariate adaptive regression spline models. *Stoch Environ Res Risk Assess* 2015;1–16.
- Sharda V, Prasher S, Patel R, Ojasvi P, Prakash C. Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data/Performances de régressions par splines multiples et adaptives (MARS) pour la prévision d'écoulement au sein de micro-bassins versants Himalayens d'altitudes intermédiaires avec peu de données. *Hydrol Sci J* 2008;53:1165–75.
- Wang D, Luo H, Grunder O, Lin Y, Guo H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl Energy* 2017;190:390–407.
- Wang D, Wei S, Luo H, Yue C, Grunder O. A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. *Sci Total Environ* 2017;580:719–33.
- Prasad R, Deo RC, Li Y, Maraseni T. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmos Res* 2017;197:42–63.
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences. The Royal Society*; 1998. p. 903–95.
- Wu Z, Huang NE. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Adv Adapt Data Anal* 2009;1:1–41.
- Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. 2011 IEEE international conference on Acoustics, speech and signal processing (ICASSP). IEEE; 2011. p. 4144–7.
- Colominas MA, Schlotthauer G, Torres ME. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed Signal Process Control* 2014;14:19–29.
- Deo RC, Tiwari MK, Adamowski JF, Quilty MJ. Forecasting effective drought index using a wavelet extreme learning machine (W-ELM) model. *Stoch Environ Res Risk Assess* 2016;1–30.
- Deo RC, Wen X, Feng Q. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 2016;168:568–93.
- Wen X, Feng Q, Deo RC, Wu M, Si J. Wavelet analysis-artificial neural network conjunction models for multi-scale monthly groundwater level predicting in an arid inland river basin, northwestern China. *Hydrol Res* 2016;nh2016396.
- Ancil F, Tape DG. An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. *J Environ Eng Sci* 2004;3:S121–8.
- Labat D, Ababou R, Mangin A. Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. *J Hydrol* 2000;238:149–78.
- Nourani V, Komasi M, Mano A. A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resour Manage* 2009;23:2877–94.
- Rathinasamy M, Khosa R, Adamowski J, Partheepan G, Anand J, Narsimlu B. Wavelet-based multiscale performance analysis: an approach to assess and improve hydrological models. *Water Resour Res* 2014;50:9721–37.
- Olvera-Guerrero OA, Prieto-Guerrero A, Espinosa-Paredes G. Decay ratio estimation in BWRs based on the improved complete ensemble empirical mode decomposition with adaptive noise. *Ann Nucl Energy* 2017;102:280–96.
- Li X, Li C. Pretreatment and wavelength selection method for near-infrared spectra signal based on improved CEEMDAN energy entropy and permutation entropy. *Entropy* 2017;19:380.
- Marzband M, Azarnejadian F, Savaghebi M, Guerrero JM. An optimal energy management system for islanded microgrids based on multiperiod artificial bee colony combined with Markov chain. *IEEE Syst J* 2015.
- Marzband M, Ghadimi M, Sumper A, Domínguez-García JL. Experimental validation of a real-time energy management system using multi-period gravitational search algorithm for microgrids in islanded mode. *Appl Energy* 2014;128:164–74.
- Marzband M, Ardesheeri RR, Moafi M, Uppal H. Distributed generation for economic benefit maximization through coalition formation-based game theory concept. *Int Trans Electrical Energy Syst* 2017;27.
- Marzband M, Ghazimirsaeid SS, Uppal H, Fernando T. A real-time evaluation of energy management systems for smart hybrid home Microgrids. *Electr Power Syst Res* 2017;143:624–33.
- Marzband M, Javadi M, Domínguez-García JL, Moghaddam MM. Non-cooperative game theory based energy management systems for energy district in the retail market considering DER uncertainties. *IET Gener Trans Distrib* 2016;10:2999–3009.
- Marzband M, Moghaddam MM, Akorede MF, Khomeyran G. Adaptive load shedding scheme for frequency stability enhancement in microgrids. *Electr Power Syst Res* 2016;140:78–86.
- Marzband M, Sumper A, Domínguez-García JL, Gumara-Ferret R. Experimental validation of a real time energy management system for microgrids in islanded mode using a local day-ahead electricity market and MINLP. *Energy Convers Manage* 2013;76:314–22.
- Marzband M, Sumper A, Ruiz-Álvarez A, Domínguez-García JL, Tomoiagă B. Experimental evaluation of a real time energy management system for stand-alone microgrids in day-ahead markets. *Appl Energy* 2013;106:365–76.
- Marzband M, Yousefnejad E, Sumper A, Domínguez-García JL. Real time experimental implementation of optimum energy management system in standalone microgrid by using multi-layer ant colony optimization. *Int J Electr Power Energy Syst* 2016;75:265–74.
- Marzband M, Javadi M, Pourmousavi SA, Lightbody G. An advanced retail electricity market for active distribution systems and home microgrid interoperability based on game theory. *Electr Power Syst Res* 2018;157:187–99.
- Marzband M, Parhizi N, Adabi J. Optimal energy management for stand-alone

- microgrids based on multi-period imperialist competition algorithm considering uncertainties: experimental validation. *Int Trans Electr Energy Syst* 2016;26:1358–72.
- [49] Marzband M, Parhizi N, Savaghebi M, Guerrero JM. Distributed smart decision-making for a multimicrogrid system based on a hierarchical interactive architecture. *IEEE Trans Energy Convers* 2016;31:637–48.
- [50] Valinejad J, Marzband M, Akorede MF, Barforoshi T, Jovanović M. Generation expansion planning in electricity market considering uncertainty in load demand and presence of strategic GENCOs. *Electr Power Syst Res* 2017;152:92–104.
- [51] Al-Musaylh MS, Deo RC, Adamowski JF, Li Y. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Adv Eng Inf* 2018;35:1–16.
- [52] Che J, Wang J. Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling. *Energy Convers Manage* 2010;51:1911–7.
- [53] Vapnik VN. *Statistical learning theory*. New York: Wiley; 1998.
- [54] Friedman JH. Multivariate adaptive regression splines. *Ann Stat* 1991:1–67.
- [55] Quinlan JR. *Learning with continuous classes*. In: 5th Australian joint conference on artificial intelligence, Singapore; 1992. p. 343–8.
- [56] Rahimikhoob A, Asadi M, Mashal M. A comparison between conventional and M5 model tree methods for converting pan evaporation to reference evapotranspiration for semi-arid region. *Water Resources Manage* 2013;27:4815–26.
- [57] Kenny J. Particle swarm optimization. In: *Proc 1995 IEEE Int Conf Neural Networks*; 1995. p. 1942–8.
- [58] Eberhart R, Kennedy J. A new optimizer using particle swarm theory. *Micro Machine and Human Science, 1995 MHS'95, Proceedings of the Sixth International Symposium on*. IEEE; 1995. p. 39–43.
- [59] Alam MN. *Particle Swarm Optimization: Algorithm and its Codes in MATLAB*. ResearchGate; 2016.
- [60] Hasanipناه M, Shahnazar A, Amnieh HB, Armaghani DJ. Prediction of air-overpressure caused by mine blasting using a new hybrid PSO-SVR model. *Eng Comput* 2017;33:23–31.
- [61] Chen L-F, Su C-T, Chen K-H. An improved particle swarm optimization for feature selection. *Intell Data Anal* 2012;16:167–82.
- [62] Alam MN, Das B, Pant V. A comparative study of metaheuristic optimization approaches for directional overcurrent relays coordination. *Electr Power Syst Res* 2015;128:39–52.
- [63] AEMO. *Aggregated Price and Demand Data – Historical*. Australia: The Australian Energy Market Operator; 2016. p. The Australian Energy Market Operator is responsible for operating Australia's largest gas and electricity markets and power systems.
- [64] Lin H-T, Lin C-J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. submitted to. *Neural Comput* 2003:1–32.
- [65] Chang C-C, Lin C-J. LIBSVM: a library for support vector machines. *ACM Trans Intel Syst Technol (TIST)* 2011;2:27.
- [66] Suykens JAK, De Brabanter J, Lukas L, Vandewalle J. Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing* 2002;48:85–105.
- [67] Maitry R, Bhagwat PP, Bhatnagar A. Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrol Process* 2010;24:917–23.
- [68] Goyal MK, Bharti B, Quilty J, Adamowski J, Pandey A. Modeling of daily pan evaporation in sub tropical climates using ANN, LS-SVR, Fuzzy Logic, and ANFIS. *Expert Syst Appl* 2014;41:5267–76.
- [69] Mohammadi K, Shamshirband S, Anisi MH, Alam KA, Petković D. Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Convers Manage* 2015;91:433–41.
- [70] Hoang N-D, Pham A-D, Cao M-T. A novel time series prediction approach based on a hybridization of least squares support vector regression and swarm intelligence. *Appl Comput Intell Soft Comput* 2014;2014:15.
- [71] Huang C-L, Dun J-F. A distributed PSO-SVM hybrid system with feature selection and parameter optimization. *Appl Soft Comput* 2008;8:1381–91.
- [72] Jekabsons G. *Adaptive regression splines toolbox for matlab/octave*. Version; 2013;1:72.
- [73] Jekabsons G. *M5PrimeLab: M5'regression tree, model tree, and tree ensemble toolbox for Matlab/Octave*; 2016.
- [74] Deo R, Downs N, Parisi A, Adamowski J, Quilty J. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. *Environ Res* 2017;155:141.
- [75] Willmott CJ, Robeson SM, Matsuura K. A refined index of model performance. *Int J Climatol* 2012;32:2088–94.
- [76] Willmott CJ. On the Evaluation of Model Performance in Physical Geography. *Spatial Statistics and Models*. Springer; 1984. p. 443–60.
- [77] Willmott CJ. Some comments on the evaluation of model performance. *Bull Am Meteorol Soc* 1982;63:1309–13.
- [78] Willmott CJ. On the validation of models. *Phys Geogr* 1981;2:184–94.
- [79] Dawson CW, Abraham RJ, See LM. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ Modell Software* 2007;22:1034–52.
- [80] Legates DR, McCabe GJ. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 1999;35:233–41.
- [81] Krause P, Boyle D, Båse F. Comparison of different efficiency criteria for hydrological model assessment. *Adv Geosci* 2005;5:89–97.
- [82] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geoscientific Model Dev* 2014;7:1247–50.
- [83] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I—a discussion of principles. *J Hydrol* 1970;10:282–90.
- [84] Li M-F, Tang X-P, Wu W, Liu H-B. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Convers Manage* 2013;70:139–48.
- [85] Mohammadi K, Shamshirband S, Tong CW, Arif M, Petković D, Ch S. A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers Manage* 2015;92:162–71.
- [86] Heinemann AB, van Oort PA, Fernandes DS, Maia ADHN. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia* 2012;71:572–82.
- [87] Legates DR, McCabe GJ. A refined index of model performance: a rejoinder. *Int J Climatol* 2013;33:1053–6.
- [88] Jolliff JK, Kindle JC, Shulman I, Penta B, Friedrichs MA, Helber R, et al. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *J Mar Syst* 2009;76:64–82.
- [89] Liu Y, Wang G, Chen H, Dong H, Zhu X, Wang S. An improved particle swarm optimization for feature selection. *J Bionic Eng* 2011;8:191–200.
- [90] Shi X, Huang Q, Chang J, Wang Y, Lei J, Zhao J. Optimal parameters of the SVM for temperature prediction. *Proce Int Assoc Hydrol Sci* 2015;368:162–7.
- [91] Hor C-L, Watson SJ, Majithia S. Analyzing the impact of weather variables on monthly electricity demand. *IEEE Trans Power Syst* 2005;20:2078–85.
- [92] Tiwari MK, Chatterjee C. A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. *J Hydroinformatics* 2011;13:500–19.
- [93] Slougher JM, Gneiting T, Raftery AE. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J Am Statist Assoc* 2010;105:25–35.

Chapter 5: Using a Wide Range of Data to Investigate the Impact of Climate and Atmospheric Variables on Electricity Demand Forecasting

5.1 Foreword

This chapter presents an exact copy of the published article in the journal of *Renewable and Sustainable Energy Reviews* (Vol. 113, Pages 109-293).

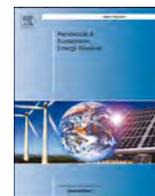
This chapter is designed to address some limitations appeared in the previous work of Chapter 4. The challenges are the influence of exogenous predictor variables and developing ensemble-based uncertainty model to generate the prediction errors. To address these problems, a variety of predictor datasets from climate and atmospheric parameters and a bootstrapping procedure (B) have been used to feed an artificial neural network (ANN) constructing a new forecasting methodology and estimate the forecast uncertainty, respectively. Energex (G) with a wide range of exogenous input datasets for eight stations located in southeast Queensland, Australia are adopted to develop these methods for 6.0 hour and daily G forecasting horizons. A comparison study is then made between the ANN model and other three models of multivariate adaptive regression spline (MARS), multiple linear regression (MLR), and autoregressive integrated moving average (ARIMA). Additionally, a combination of forecasting approach is presented in this chapter by combining the forecasts of the ANN, MARS, and MLR models to build a hybrid ANN model. Finally, the B-hybrid ANN technique is used to construct an ensemble model and evaluate the uncertainty assessments obtained from the forecast values. The novel and main contribution of this chapter is the using of climate and atmospheric predictor variables, as a pertinent driver of energy demand variability, as changes in climatic and atmospheric conditions are likely to influence how people and organisations use electrical energy.

5.2 Research Highlights

- Importance of external datasets (climate and atmospheric Reanalysis) to forecast energy demand.
- Developing a hybrid ANN model by combining the outputs of three models.
- Estimating the forecast uncertainty using a bootstrap method.
- Low and high-dimensional models.

- Energy security studies should explore ANN models trained with climate variables.

5.3 Published Article III



Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland, Australia



Mohanad S. AL-Musaylh^{a,b,*}, Ravinesh C. Deo^{a,*}, Jan F. Adamowski^c, Yan Li^a

^a School of Agricultural, Computational and Environmental Sciences, Institute of Life Sciences and the Environment, University of Southern Queensland, QLD, 4350, Australia

^b Management Technical College, Southern Technical University, Basrah, Iraq

^c Department of Bioresource Engineering, Faculty of Agricultural and Environmental Science, McGill University, Québec, H9X 3V9, Canada

ARTICLE INFO

Keywords:

Predictive model for electricity demand
Climate and ECMWF Reanalysis variables
ANN
MARS
MLR
ARIMA
Hybrid ANN
Bootstrapping
Sustainable energy management systems

ABSTRACT

Reliable models that can forecast energy demand (G) are needed to implement affordable and sustainable energy systems that promote energy security. In particular, accurate G models are required to monitor and forecast local electricity demand. However, G forecasting is a multivariate problem, and thus models must employ robust pattern recognition algorithms that can detect subtle variations in G due to causal factors, such as climate variables. Therefore, this study developed an artificial neural network (ANN) model that used climatic variables for 6-hour (h) and daily G forecasting. The input variables included the six most relevant climate variables from Scientific Information for Land Owners (SILO) and 51 Reanalysis variables obtained from the European Centre for Medium-Range Weather Forecast (ECMWF) models. This information was used to forecast G data obtained from the energy utility (Energex) at 8 stations in southeast Queensland, Australia, by utilizing statistically significant lagged cross-correlations of G with its predictor variables. The developed ANN model was then benchmarked against multivariate adaptive regression spline (MARS), multiple linear regression (MLR), and autoregressive integrated moving average (ARIMA) models using various statistical metrics, such as relative root-mean square error ($RRMSE\%$). Additionally, this study developed a hybrid ANN model by combining the forecasts of the ANN, MARS, and MLR models. The bootstrap (B) technique was also used with the hybrid ANN model, creating the B-hybrid ANN, to estimate the forecast uncertainty. According to both forecast horizons, the results indicated that the ANN model was more accurate than the ARIMA, MARS, and MLR models for G forecasting. Furthermore, the hybrid ANN was the most accurate model developed in this research study. For example, at the best site (Redcliffe), the hybrid ANN model generated an $RRMSE$ of 3.85% and 4.37% for the 6-h and daily horizons, respectively. This study found that an ANN model could be used for accurately forecasting G over multiple horizons in southeast Queensland.

1. Introduction

In 2015 the 2030 Agenda for Sustainable Development was adopted by all United Nations member states, representing their commitment to the strategic implementation of 17 sustainable development goals (SDGs) [1]. In particular, Goal seven aims to improve global energy systems [1] through the achievement of five objectives by 2030. These include (1) provide global access to affordable, reliable, and modern energy, (2) greatly increase the share of renewable energy around the world, (3) double the universal rate of improving energy efficiencies,

(4) improve international collaborations and access to energy studies and skills, and (5) support technologies that supply up-to-date and sustainable energy services [1]. These objectives are important as global energy demand is expected to increase by over 50% before 2030 if worldwide demand growth rates are not curtailed [2].

An overview of projected changes in electricity demand (G) from 2020 to 2030 in the Australian Energy Market identified significant factors that affect the electricity industry [3]. These factors included potential changes in regional markets and the need for better systems to project demand accurately. It is expected that the demand of

* Corresponding authors. School of Agricultural, Computational and Environmental Sciences, Institute of Life Sciences and the Environment, University of Southern Queensland, QLD 4350, Australia.

E-mail addresses: MohanadShakirKhalid.AL-Musaylh@usq.edu.au, mohanadk21@gmail.com (M.S. AL-Musaylh), ravinesh.deo@usq.edu.au (R.C. Deo).

<https://doi.org/10.1016/j.rser.2019.109293>

Received 8 February 2019; Received in revised form 22 July 2019; Accepted 22 July 2019

Available online 26 July 2019

1364-0321/ © 2019 Elsevier Ltd. All rights reserved.

Acronyms			
MW	Megawatt	σ^2	Variance
G	Electricity load demand, MW	logsig	Log sigmoid
MARS	Multivariate adaptive regression splines	ARIMA	Autoregressive integrated moving average
SVR	Support vector regression	ANN	Artificial neural network
RMSE	Root-mean square error	AEMO	Australian Energy Market Operator
MAE	Mean absolute error	purelin	Positive linear
RRMSE	Relative root-mean square error, %	L	Hidden neuron size
MAPE	Mean absolute percentage error, %	PACF	Partial auto-correlation function
WI	Willmott's Index	B	Bootstrapping algorithm
E_{NS}	Nash–Sutcliffe efficiency coefficient	R^2	Coefficient of determination
E_{LM}	Legates and McCabe's Index	BMA	Bayesian model averaging
SILO	Scientific Information for Land Owners	FA	Firefly algorithm
ECMWF	European Centre for Medium Range Weather Forecasts	WT	Wavelet transforms
G^{for}	G forecasted	GCV	Generalized cross-validation
ICEEMDAN	Improved version of empirical mode decomposition with adaptive noise	G_i^{for}	i th forecasted value of G, MW
MLR	Multiple linear regression	G_i^{obs}	i th observed value of G, MW
G^{obs}	G observed	$\frac{G^{for}}{G^{obs}}$	Mean of forecasted values
trainbfg	Levenberg-Marquardt	FE	Absolute forecasted error statistics
trainlm	Broyden-Fletcher-Goldfarb-Shanno	PSO	Particle swarm optimization
tansig	Tangent sigmoid	LH	Log likelihood
p	Autoregressive term in ARIMA	d	Degree of differencing in ARIMA
q	Moving average term in ARIMA	AIC	Akaike information criterion
		SDGs	Sustainable development goals

conventional energy sources in Australia will be reduced over the next decade from increased renewable energy supply, improvements in the energy efficiency of building structures, and the use of solar panels in residential areas [3]. However, the short and long-term stochastic behaviour or subtle variation of energy demand is influenced by several exogenous factors. These factors include changes in solar radiation (*i.e.*, impacting the solar energy industry), wind regimes (*i.e.*, impacting the wind energy industry), and localized air temperature (*i.e.*, driving people to demand more energy to improve their comfort in extremely warm and cold weather). Given the potential influence of these exogenous factors on energy demand, a variety of predictor datasets are needed when developing a model for G monitoring and forecasting. Accurate knowledge of current and forecasted G could aid the design of energy devices that can support nations in achieving Goal 7 of the SDGs.

In regional and national energy grid systems, accurate G forecasts can support the development of energy security strategies. In particular, G forecasts can help in addressing issues of localized energy fluctuation in distribution networks, energy system installations, renewable energy investments, electricity demand planning, and creation of new energy management policies [4]. Currently, in Australia, forecasted energy demand is primarily managed by forecasters working in the Australian Energy Market Operator (AEMO) [5], who utilize semi-parametric-type additive models that explore multiple input variables such as air temperature and calendar effects, as well as demographic and economic datasets associated with energy demands for whole states. Additive models are useful for approximating high-dimensional regression functions and extending generalized linear models to combine statistical learning with interpretability and flexibility. However, these models do not consider demand data for smaller areas, such as substations, which can be significantly affected by climate variables.

Furthermore, these models have limitations due to assumptions of the data distribution, a tendency for overfitting, and a loss of predictability when smoothed variables have values outside of the training data range [6]. However, recent advancements in data analytics have led to improvements in machine learning methods through the black-box approach, which has subsequently become a standard methodology for G forecasting [23]. For example, a number of recent studies have

forecasted G using statistically significant lag combinations of historical demand data along with partial autocorrelation functions (*e.g.*, [7–12]), where different data-driven models for G forecasting were developed with machine learning algorithms.

Specifically, G data for the entire area of Queensland, Australia, were forecasted by Al-Musaylh and Deo [7] using multivariate adaptive regression splines (MARS), support vector regressions (SVR), and autoregressive integrated moving average (ARIMA) models. The authors concluded that the MARS algorithm performed the best for the short-term (0.5 h and 1.0 h) forecast horizons, and SVR performed the best for the daily (24 h) forecast horizon. Another study [8], by the same authors, led to an improvement in forecasting accuracy through the use of a two-phase particle swarm optimized (PSO)-SVR hybrid model, which was integrated with improved empirical mode decomposition with adaptive noise (ICEEMDAN) as a pre-processing algorithm to decompose G data (before running the forecasting model). In that study, multiple forecast horizons, including short-term (*i.e.*, weekends, working days, whole weeks, and public holidays) and long-term (*i.e.*, monthly) horizons, were employed to test the predictive ability of the ICEEMDAN-PSO-SVR hybrid approach. These studies, which specifically focused on Australia, contributed to the utilization of machine learning algorithms to forecast electrical energy demand.

Elsewhere, the performance of artificial neural network (ANN) models were tested using statistically significant combinations for different datasets including energy prediction data for buildings [13,14], electricity price data [15], and energy demand data [16,17]. Although a reliable forecasting accuracy for the energy demand model was obtained using statistically significant lagged inputs of real demand series [7,8], model performance was improved when other variables were incorporated into the model, likely because several interacting elements influence changes in electricity demand. Localized weather-related variables, such as air temperature, rainfall, wind regimes, and cloud cover, are likely to affect electrical energy usage.

A literature review revealed that a limited number of research studies (*e.g.*, [18–20]) have incorporated exogenous predictor variables from climate-based elements to forecast G data. Research undertaken by Mirasgedis et al. [18] focused on a region in Greece where a few meteorological variables, such as air temperature, humidity, wind

speed, and solar radiation were employed for medium-term G^{for} . They concluded that air temperature and humidity could be crucial predictor variables used in an energy demand model. Additionally, in South Africa, Lebotsa et al. [19] used calendar effects, air temperature, and a lagged demand dataset to train a partial linear additive quantile regression model for energy demand forecasting. Other research has indicated a possible causal relationship between climate change and energy demand without examining the feasibility of designing a demand model for G data forecasting [20]. In the context of Australia, where climate-based variables play a crucial role in forecasting solar radiation [21,22], exploratory studies that incorporate these variables into energy demand models are an interesting and novel research endeavour. Studies on multivariate methods for G forecasting can potentially inform engineers, power analysts, and climate-energy policymakers on how to improve energy demand models by considering the role of exogenous climate-based predictors in real-world energy forecasting problems.

Therefore, the current research presents a novel study that aimed to advance a new approach for energy demand forecasting by using extensive sets of climate-based predictors, including both ground-based measurements and atmospheric Reanalysis data obtained from the numerical weather forecasting model, to forecast station-based electricity energy demand. The study region was southeast Queensland, Australia, where ground-based model data were extracted from the Scientific Information for Land Owners (SILO), and numerical weather forecast data from the European Centre for Medium-Range Weather Forecasts (ECMWF). This study utilized a relatively wide range of input features and patterns from climate-based predictor variables that may be related to energy demand data in order to construct a robust model framework. A total of six ground-based variables and 51 ECMWF-based variables

were used. The energy demand data were extracted from the database of the primary energy utility, Energex, a leading power distributor in southeast Queensland.

To achieve this goal, physical and data-driven models were considered. While physically-based models are based on the physical behaviour of model parameters, data-driven models employ machine learning algorithms to detect the relationships between predictors and objective variables using historical datasets [23–25]. Recent studies (e.g., [26,27]), have found that data-driven forecasting models such as ANN led to improvements in forecasting accuracy over physical models such as the Predictive Ocean Atmosphere Model for Australia (POAMA). Five advantages of data-driven models, in comparison to physically-based models, have been identified [26,28,29]. Specifically, data-driven models (1) can be run and the output can be explained with less complexity in the future pattern's description of the climate variables, (2) are easier to assess, (3) require less data and costs, (4) require shorter training and testing periods, and (5) have high specificity. Therefore, applying data-driven techniques to create a G forecasting model can have many advantages as an alternative to physically-based models.

In this paper, an ANN model, a data-driven black-box tool that does not require detailed information about the predictor variables in comparison with physical techniques [23], was developed for G forecasting. The ANN model can predict future values of more than one variable at the same time and handle non-linear datasets [30]. The ANN model has various applications for data forecasting (e.g., [31,32]) when exogenous variables are used, and when it has been used globally, it has often achieved a relatively small margin of predictive errors for different data forecasting horizons (e.g., [16,28,32–34]). Additionally, different studies (e.g., [35–38]) have developed ensemble-based uncertainty

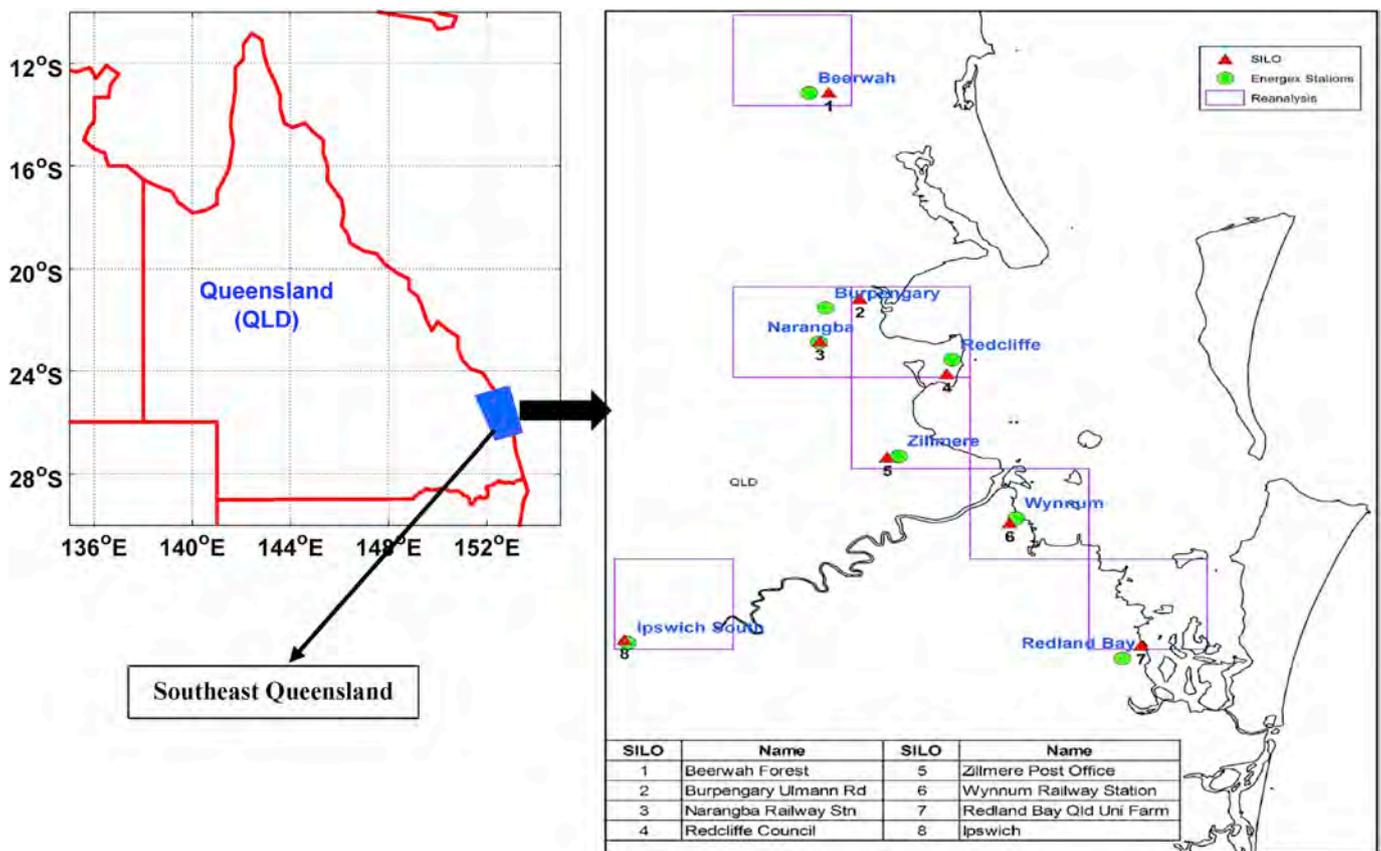


Fig. 1. Map of the present study area showing the locations in southeast Queensland where the energy demand forecasting models were developed and tested using climate-based predictor variables. (a) Energex stations are shown as green circles (b) SILO points are shown as red triangles and (c) ECMWF (ERA-Interim) Reanalysis grids are shown as pink boxes, where each grid contains four points with the mean of these points used as the data point for that grid. Information in Table 1 was used to generate the spatial map. (To view this figure in colour, the reader is referred to the Web version of this article).

Table 1

The longitudes and latitudes for data locations used in this study: (a) Energen (b) SILO (c) ECMWF (ERA-Interim) Reanalysis.

Energen Station			SILO			Reanalysis			
Name	Latitude	Longitude	Name	Latitude	Longitude	Latitude	Longitude	Latitude	Longitude
Beerwah	26.8573	152.9552	Beerwah Forest	26.8564	152.9764	26.7500	26.8750	152.8750	153.0000
Burpengary	27.1544	152.9728	Burpengary Ulmann Rd	27.1414	153.0089	27.1250	27.2500	153.0000	153.1250
Ipswich South	27.61551	152.7641	Ipswich	27.6117	152.7608	27.5000	27.6250	152.7500	152.8750
Narangba	27.2015	152.9655	Narangba Railway Stn	27.2000	152.9667	27.1250	27.2500	152.8750	153.0000
Redcliffe	27.2258	153.1063	Redcliffe Council	27.2450	153.1006	27.1250	27.2500	153.0000	153.1250
Redland Bay	27.6372	153.2857	Redland Bay Qld Uni Farm	27.6192	153.3056	27.5000	27.6250	153.2500	153.3750
Wynnum	27.4437	153.1736	Wynnum Railway Station	27.4500	153.1667	27.3750	27.5000	153.1250	153.2500
Zillmere	27.3583	153.0500	Zillmere Post Office	27.3589	153.0375	27.2500	27.3750	153.0000	153.1250

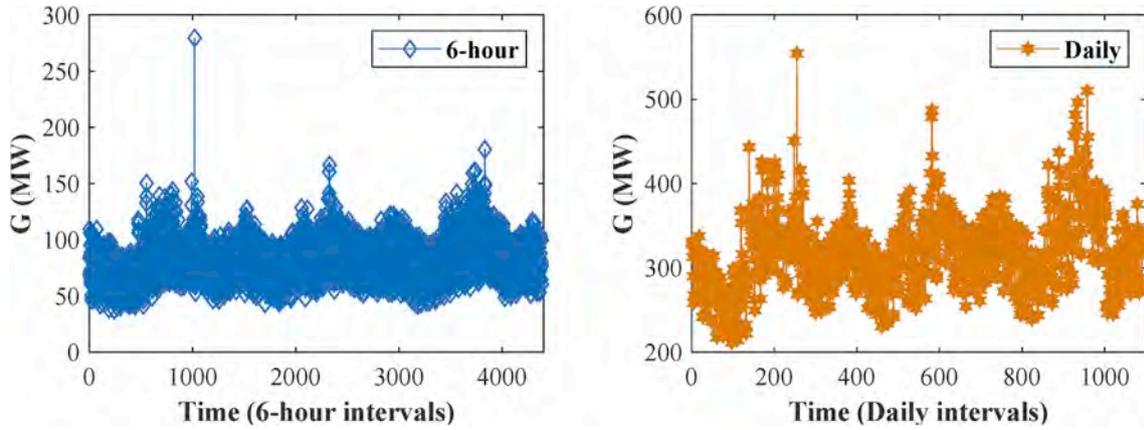


Fig. 2. Time-series of the electricity demand (G, MW) data used for the Beerwah study site for 6-h and daily periods.

Table 2

Descriptive statistics of the electricity demand (G, MW) for the eight Energen stations with 6-h and daily periods.

Forecast horizon	Data Period (dd-mm-yyyy)	Property	Beerwah	Burpengary	Ipswich South	Narangba	Redcliffe	Redland Bay	Wynnum	Zillmere
6-h	01-07-2014 to 30-06-2017	Minimum (MW)	37.40	21.55	38.81	21.27	73.74	18.09	59.40	78.39
		Maximum (MW)	279.52	258.00	347.98	251.66	235.45	359.37	266.43	600.87
		Mean (MW)	79.65	87.93	94.09	100.96	125.72	67.72	109.19	165.51
		Standard deviation (MW)	18.76	30.48	28.08	29.10	27.55	24.55	28.33	39.58
		Daily	01-07-2014 to 29-06-2017	Minimum (MW)	211.54	234.71	258.18	220.63	392.69	203.88
Maximum (MW)	554.27	678.53		696.52	683.55	718.56	557.40	724.38	1106.43	
Mean (MW)	318.59	351.76		376.39	403.85	502.89	270.90	436.76	662.01	
Standard deviation (MW)	44.68	61.77		62.90	78.78	55.90	44.62	59.46	87.58	

forecasted models by using a bootstrap (B) procedure to assess the model's stability based on the prediction intervals of the ANN models. In particular, B is a data-driven technique that can utilize an ensemble framework to help reduce the prediction uncertainties through intensive resampling with replacement [39,40]. In Australia, three studies [23,31,32] have successfully applied the ANN model to forecast precipitation, solar radiation, and streamflow. However, the ability of the ANN model integrated with large datasets to forecast G has not yet been explored everywhere.

Therefore, in this study, the ANN model was evaluated against two alternative multivariate models, MARS and multiple linear regression (MLR), developed as high-dimensional models with several predictors, as well as the low dimensional (simple) model, ARIMA. Various statistical metrics were used to assess the performance of the ANN for G forecasting, in comparison to the other models employed. In addition, a hybrid ANN model was constructed by combining the three models of ANN, MARS, and MLR. For the hybrid model, the prediction interval (PI) was calculated using the B-hybrid ANN model to address uncertainty assessment for G forecasting.

There were five main objectives of this study, (1) to extract datasets from SILO and ECMWF to be used as predictor variables, and datasets

from Energen to be used as the target variable, (2) to develop an ANN model for short-term forecasting of 6-hour (h) and daily horizons, (3) to evaluate the performance of the ANN model in comparison to the MARS, MLR, and ARIMA models, (4) to construct a hybrid ANN model by combining ANN with MARS and MLR, and (5) to estimate the forecast uncertainty using the B-hybrid ANN model. These objectives are achieved in the following sections. Section 2 explains the supporting theory for the different models, while section 3 provides the materials and methods, which include the data, forecast model development, and assessment criteria. Section 4 presents the results and discussion, while section 5 provides possible solutions to address challenges in future work. Finally, section 6 summarizes the research results and contributions.

2. Conceptual framework

In this section, only the ANN, MLR, and B methods are described thoroughly. The details of the MARS, and ARIMA strategies, presented by Friedman [41] and Box and Jenkins [42], are well explained in previous works (e.g., [7,8,31,43–45]), and thus not presented here.

Table 3
Model input variables from SILO and ERA-Interim Reanalysis used for 6-h and daily G^{for}

Source	No.	Variable name	Acronym	Mean (e.g. Beerwah data)	
				6-h	Daily
SILO	1	Maximum temperature	TMax		26.31
	2	Minimum temperature	TMin		15.88
	3	Rainfall	Rain		3.73
	4	Evaporation	Evap		5.15
	5	Solar radiation	Radn		18.28
	6	Vapour pressure	VP		17.84
Reanalysis	7	Vertical integral of mass of atmosphere	p53.162	10186.67	10186.64
	8	Vertical integral of temperature	p54.162	2601493.88	2601506.13
	9	Vertical integral of water vapour	p55.162	25.66	25.66
	10	Vertical integral of ozone	p58.162	0.01	0.01
	11	Vertical integral of kinetic energy	p59.162	2535683.21	2533465.76
	12	Vertical integral of thermal energy	p60.162	2611649009.46	2611660635.11
	13	Vertical integral of potential + internal energy	p61.162	2626949967.16	2626961540.62
	14	Vertical integral of potential + internal + latent energy	p62.162	2691110176.10	2691142610.04
	15	Vertical integral of total energy	p63.162	2693645851.60	2693676068.17
	16	Vertical integral of energy conversion	p64.162	164.85	165.06
	17	Vertical integral of eastward mass flux	p65.162	111538.38	111457.87
	18	Vertical integral of northward mass flux	p66.162	12733.86	12778.81
	19	Vertical integral of eastward kinetic energy flux	p67.162	94633097.10	94501469.70
	20	Vertical integral of northward kinetic energy flux	p68.162	-2924314.67	-2848793.90
	21	Vertical integral of eastward heat flux	p69.162	26271018743.55	26252081657.94
	22	Vertical integral of northward heat flux	p70.162	3401655077.02	3411787299.63
	23	Vertical integral of eastward water vapour flux	p71.162	10.25	10.25
	24	Vertical integral of northward water vapour flux	p72.162	26.61	26.58
	25	Vertical integral of eastward geopotential flux	p73.162	10698893652.73	10690448746.89
	26	Vertical integral of northward geopotential flux	p74.162	653385798.50	658668079.51
	27	Vertical integral of eastward total energy flux	p75.162	37090107708.45	37062579702.24
	28	Vertical integral of northward total energy flux	p76.162	4118857514.61	4134281550.79
	29	Vertical integral of eastward ozone flux	p77.162	-0.02	-0.02
	30	Vertical integral of northward ozone flux	p78.162	0.00	0.00
	31	Vertical integral of divergence of mass flux	p81.162	0.00	0.00
	32	Vertical integral of divergence of kinetic energy flux	p82.162	4.23	4.09
	33	Vertical integral of divergence of thermal energy flux	p83.162	105.85	106.17
	34	Vertical integral of divergence of moisture flux	p84.162	0.00	0.00
	35	Vertical integral of divergence of geopotential flux	p85.162	-178.25	-178.37
	36	Vertical integral of divergence of ozone flux	p87.162	0.00	0.00
	37	Vertical integral of northward cloud liquid water flux	p89.162	-0.02	-0.02
	38	Vertical integral of northward cloud frozen water flux	p91.162	-0.12	-0.12
	39	Vertical integral of mass tendency	p92.162	0.00	0.00
	40	Surface pressure	sp	99897.13	99896.77
	41	Total column water	tcw	25.73	25.74
	42	Total column water vapour	tcwv	25.66	25.66
	43	Soil temperature level 1	stl1	293.26	293.26
	44	Mean sea level pressure	msl	101702.71	101702.33
	45	Total cloud cover	tcc	0.38	0.38
	46	10 m U wind component	u10	-0.68	-0.68
	47	10 m V wind component	v10	0.49	0.49
	48	2 m temperature	t2m	293.53	293.53
	49	2 m dewpoint temperature	d2m	287.97	287.97
	50	Soil temperature level 2	stl2	292.82	292.82
	51	Albedo	al	0.15	0.15
	52	Soil temperature level 3	stl3	292.81	292.82
	53	Low cloud cover	lcc	0.15	0.15
	54	Medium cloud cover	mcc	0.15	0.15
	55	High cloud cover	hcc	0.17	0.17
	56	Total column ozone	tco3	0.01	0.01
	57	Soil temperature level 4	stl4	292.80	292.80

Table 4
Data splitting technique used for model development and testing.

Forecast horizon	Data Period (dd-mm-yyyy)	Number of data points		
		Total	Training (80%)	Testing (20%)
6-h	01-07-2014 to 30-06-2017	4383	3506	877
Daily	01-07-2014 to 29-06-2017	1095	876	219

2.1. Artificial neural network (ANN)

A non-linear regression problem with multiple inputs can be solved by an ANN technique [46] as follows:

$$y(x) = F \left(\sum_{i=1}^L w_i(t) \cdot x_i(t) + b \right) \tag{1}$$

where $x_i(t) = \{x_i\}_{i=1}^N \in R^N$ are the input variables (SILO, ECMWF Reanalysis or partial autocorrelation function (PACF) for G data), and $y(x) = \{y_i\}_{i=1}^N \in R$ is the target variable (G^{for}) in the training period. The values of b , $w_i(t)$, F , and L are the neuronal bias, the weight that

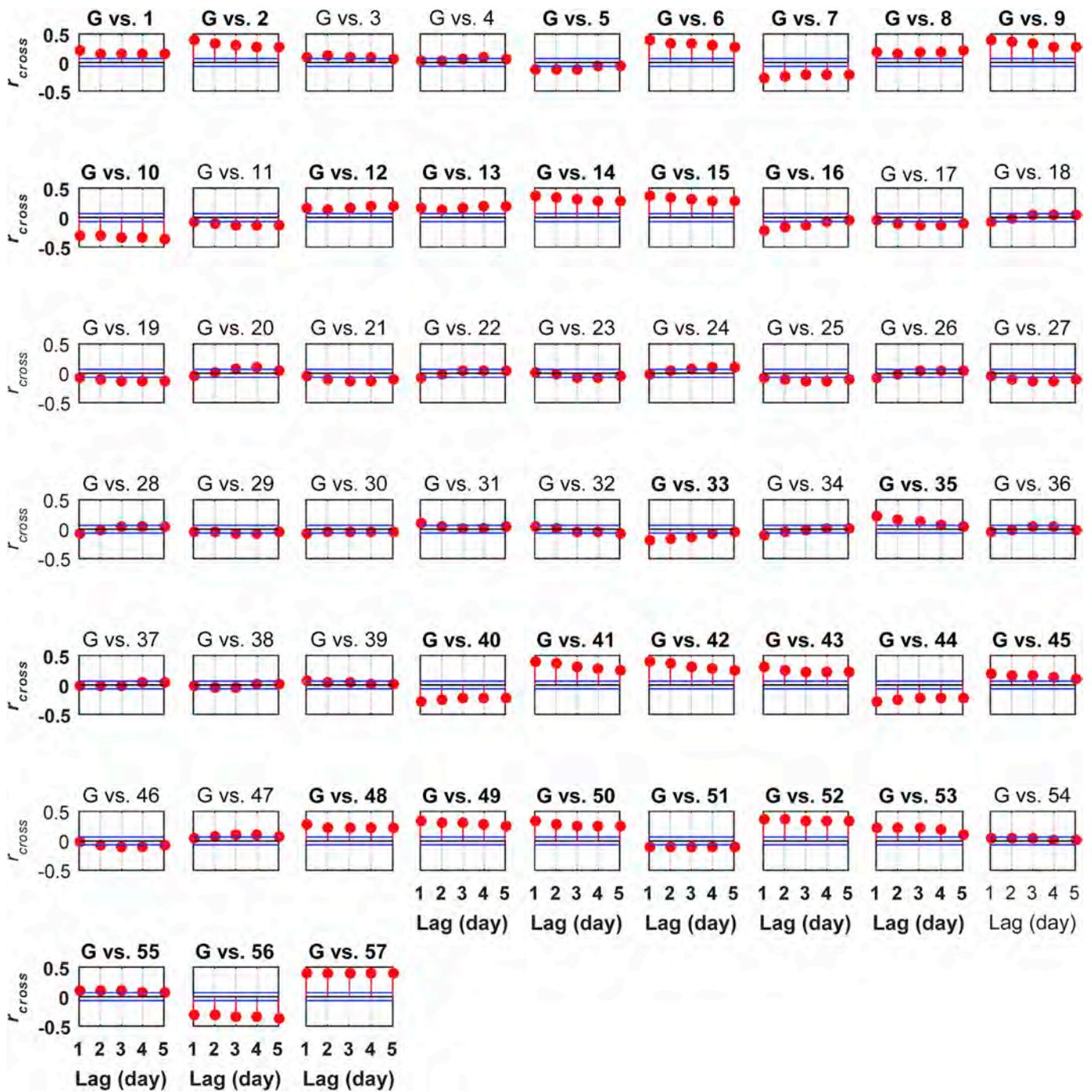


Fig. 3. Cross-correlation coefficients (r_{cross}) performed to investigate the co-variance between daily electricity demand (G) vs. the predictor variables from SILO (variables 1 to 6) and ERA-Interim Reanalysis dataset (variables 7 to 57) for the Beerwah station in the model training period. The 95% significance boundaries are shown in blue, indicating the variables selected to calculate statistically significant lagged 1 inputs matrix for the models. The selected variables are shown in boldface. Details of the variables are shown in Table 3. (To view this figure in colour, the reader is referred to the Web version of this article.)

connects the neuron in the input layer, the hidden transfer function, and the hidden neuron size determined iteratively, respectively.

The ANN algorithm is a black-box model that needs an iterative identification process to identify the training algorithm [31,32]. Therefore, two training MATLAB algorithms, Levenberg-Marquardt (LM) (*trainbfg*) and Broyden-Fletcher-GoldfarbShanno (BFGS) (*trainlm*) [47–49], were used to minimize the mean square error [50].

2.2. Multiple linear regression (MLR)

The MLR model, a statistical procedure that shows the relationship

between the inputs (SILO, ECMWF Reanalysis, and PACF G data) and the target (G) [31], was used in to evaluate the performance of the ANN model. The regression equation of the MLR can be expressed as [51,52]:

$$y(x) = c + a_1x_1 + a_2x_2 + \dots + a_kx_k \tag{2}$$

where x and y are defined in Eq. (1). The values of k , c , and a are the number of the predictor variables, the y -intercept, and the multiple regression coefficient for each regression variable, respectively [31,33,53].

Table 5

Optimum model development for the ANN, MARS, and ARIMA models, showing the models' parameters and predictor datasets for both forecast horizons in the training period.

Station	No. input variables (*)	ANN				MARS			ARIMA (*)							
		Lag 1-SILO & Reanalysis	G-PACF lags	Hidden transfer function	Output transfer function	Training algorithm	Hidden neuron size (L)	Optimum number of spline function	Variable importance		p	d	q	σ^2	LH	AIC
									Lowest	Highest						
6-h forecast horizon																
Beerwah	21	5 (1-5)	logsig	tansig	trainlm	10	49	p92.162	G-lag 4	5	1	3	0.0015	6357.94	-12697.89	
Burpengary	24	3 (2, 4, 5)	tansig	tansig	trainlm	8	47	u10	G-lag 4	5	1	2	0.0023	5697.31	-11378.62	
Ipswich South	24	4 (1, 4-6)	logsig	tansig	trainlm	5	29	d2m	G-lag 4	5	1	3	0.0015	6456.23	-12894.46	
Narangba	24	14 (1-14)	tansig	logsig	trainbfg	37	33	G-lag 11	G-lag 1	2	1	2	0.011	2895.47	-5780.95	
Redcliffe	33	34 (1-34)	tansig	purelin	trainlm	6	49	G-lag 13	G-lag 4	2	1	2	0.0097	3113.22	-6216.45	
Redland Bay	24	6 (1-6)	tansig	purelin	trainlm	5	49	p81.162	G-lag 4	4	1	3	0.0009	7261.34	-14506.68	
Wynnum	20	9 (1-9)	tansig	purelin	trainlm	10	17	G-lag 3	G-lag 4	4	1	3	0.0022	5770.37	-11524.73	
Zillmere	26	6 (1-6)	logsig	tansig	trainlm	4	17	G-lag 2	G-lag 4	4	1	3	0.0011	6992.26	-13968.53	
Daily forecast horizon																
Beerwah	30	3 (1-3)	tansig	logsig	trainlm	2	27	stl3	G-lag 1	5	1	2	0.0057	1021.87	-2027.74	
Burpengary	35	1	logsig	purelin	trainbfg	20	11	d2m	G-lag 1	2	1	1	0.0053	1047.28	-2086.56	
Ipswich South	31	3 (1-3)	tansig	purelin	trainbfg	12	33	tcw	G-lag 1	2	1	1	0.0064	965.01	-1922.01	
Narangba	24	3 (1-3)	tansig	purelin	trainbfg	5	23	TMax	G-lag 1	2	1	2	0.012	689.22	-1368.44	
Redcliffe	37	8 (1-8)	tansig	purelin	trainbfg	2	41	stl3	G-lag 7	3	1	3	0.0073	906.40	-1798.80	
Redland Bay	27	3 (1-3)	tansig	purelin	trainlm	1	17	G-lag 2	G-lag 1	3	0	1	0.0047	1105.41	-2198.82	
Wynnum	32	3 (1-3)	tansig	logsig	trainbfg	3	31	p55.162	G-lag 1	1	0	3	0.005	1075.79	-2139.58	
Zillmere	38	3 (1-3)	tansig	purelin	trainbfg	1	17	stl4	G-lag 1	2	1	3	0.0068	940.32	-1868.65	

* Note that for the ANN and MARS models, the SILO and ERA-Interim Reanalysis data with the G-PACF lags were used for daily forecast horizon, whereas only ERA-Interim Reanalysis data with the G-PACF lags were used for 6-h forecast horizons due to the availability of data. However, only the G data with no lags (single input) were used for the ARIMA model for both forecasting horizons.

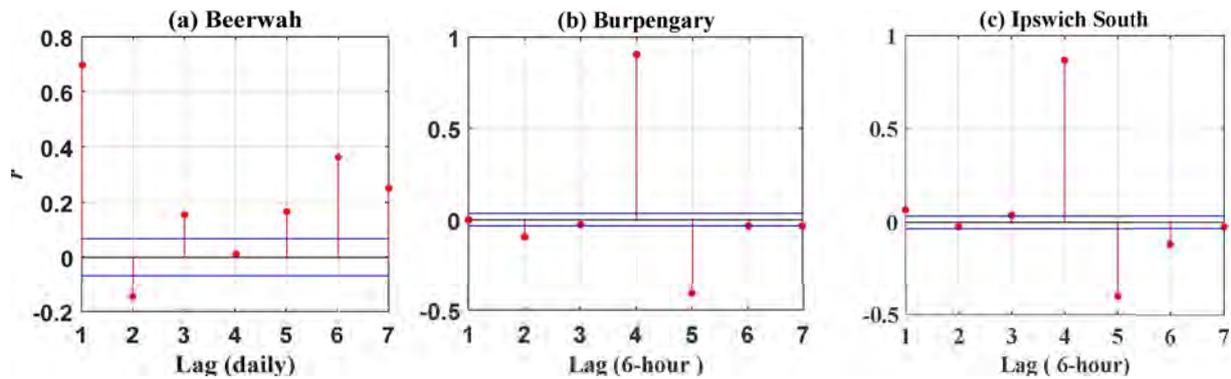


Fig. 4. Statistically significant lags of the electricity demand data (G) analysed in the training phase to construct the model's input variables. (a) The first three lags were selected for daily Beerwah data (b) Lags 2, 4 and 5 were selected for 6-h Burpengary data (c) Lags 1, 4, 5 and 6 were selected for 6-h Ipswich South data.

2.3. Bootstrapping procedures (ensemble approach)

The bootstrap is a data-driven technique that aims to reduce uncertainties through intensive resampling with replacement [39,40]. Suppose that T_n is a random sample of size n where $T_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ is extracted from a population with an unknown probability distribution F . Also, $t_i(x_i, y_i)$ is a realization drawn independently and identically distributed (iid) from F , and consists of an input vector x_i and the corresponding target variable y_i . For T_n , the empirical distribution function is defined as \hat{F} with a probability of $1/n$ for each $t_i(x_i, y_i)$, where $i = 1, 2, \dots, n$. The bootstrap sample of size, n , can be denoted by T_n^* , taken from iid with replacement from \hat{F} , and written as $T^1, T^2, \dots, T^S, \dots, T^S$, where S is the total number of bootstrap samples. According to Efron [39], the value of S is usually between 50 and 200 samples. A model can be developed for each T^s using all n observations in this study. The result will be considered as $f_{\text{model}}(x_i, w_s)$, where w_s is the weight vector for the developed model based on the

bootstrap sample, T^s . In the training period, the model performance will be evaluated using a set A from the observations, $t_i(x_i, y_i)$, that are not involved in the bootstrap samples. The estimation of the generalization error (\hat{E}_0) is used to calculate the model performance, as shown below [38,54]:

$$\hat{E}_0 = \frac{\sum_{s=1}^S \sum_{i \in A} (y_i - f_{\text{model}}(x_i, w_s))^2}{\sum_{s=1}^S |A|} \tag{3}$$

Jia and Culver [38] show that $\hat{y}(x)$ and $\sigma^2(x)$ are the average and the variance of the S bootstrapped estimates for every developed model, respectively:

$$\hat{y}(x) = \frac{1}{S} \sum_{s=1}^S f_{\text{model}}(x_i, w_s) \tag{4}$$

$$\sigma^2(x) = \frac{\sum_{s=1}^S (\hat{y}(x) - f_{\text{model}}(x_i, w_s))^2}{S - 1} \tag{5}$$

Table 6

Optimum model development for hybrid ANN and B-hybrid ANN showing the models' parameters and predictor datasets for both forecast horizons in the training period.

Station	No. input variables	Hidden neuron size (L)	Hybrid ANN Hidden transfer function	Output transfer function	Training algorithm	B-hybrid ANN Hidden transfer function	Output transfer function	Training algorithm
6-h forecast horizon								
Beerwah	3	7	logsig	tansig	trainlm	logsig	tansig	trainbfg
Burpengary	3	10	tansig	purelin	trainlm	tansig	purelin	trainbfg
Ipswich South	3	4	logsig	tansig	trainbfg	logsig	tansig	trainbfg
Narangba	3	2	tansig	purelin	trainbfg	tansig	purelin	trainbfg
Redcliffe	3	8	tansig	tansig	trainlm	tansig	tansig	trainbfg
Redland Bay	3	4	logsig	purelin	trainbfg	logsig	purelin	trainbfg
Wynnum	3	3	tansig	tansig	trainlm	tansig	tansig	trainlm
Zillmere	3	5	tansig	tansig	trainlm	tansig	logsig	trainlm
Daily forecast horizon								
Beerwah	3	2	tansig	tansig	trainlm	tansig	tansig	trainlm
Burpengary	3	8	logsig	tansig	trainbfg	logsig	tansig	trainbfg
Ipswich South	3	4	tansig	purelin	trainlm	logsig	tansig	trainbfg
Narangba	3	4	tansig	purelin	trainbfg	logsig	tansig	trainlm
Redcliffe	3	8	logsig	logsig	trainbfg	logsig	logsig	trainbfg
Redland Bay	3	9	logsig	tansig	trainbfg	tansig	logsig	trainlm
Wynnum	3	3	logsig	purelin	trainlm	tansig	logsig	trainbfg
Zillmere	3	4	tansig	tansig	trainbfg	tansig	tansig	trainbfg

Table 7

Root-mean square error (RMSE, MW), for all the station datasets for the 6-h and daily forecast horizons in the training period using the ANN, MARS, MLR, hybrid ANN, and ARIMA models.

Station	6-h					Daily				
	ANN	MARS	MLR	Hybrid ANN	ARIMA	ANN	MARS	MLR	Hybrid ANN	ARIMA
Beerwah	5.88	7.76	9.05	5.83	9.52	24.38	25.23	26.62	23.86	25.65
Burpengary	6.49	9.37	11.45	6.31	11.22	21.89	28.51	28.94	21.53	32.37
Ipswich South	8.36	9.81	11.42	8.32	11.81	27.00	30.22	33.78	26.51	35.03
Narangba	9.08	13.18	13.96	9.04	24.31	40.71	50.15	51.57	39.81	50.74
Redcliffe	3.48	5.66	6.22	3.39	15.95	22.45	21.12	22.72	20.42	27.69
Redland Bay	6.89	8.14	9.85	6.88	10.35	21.76	20.69	22.49	19.78	24.10
Wynnum	4.80	8.66	8.90	4.78	9.60	26.20	25.31	27.26	24.67	29.81
Zillmere	11.54	14.11	16.06	11.35	17.19	48.17	46.16	48.05	44.71	51.08

Since the method contains repeated applications, the frequency (K) for the prediction interval (PI) that would contain the true value is calculated as $K = (1 - \alpha) * 100\%$. With $\alpha = 0.05$, $K = (1 - 0.05) * 100\% = 95\%$ represents the prediction bounds [35]. Efron and Tibshirani [40] stated that PI can be estimated using the equation below:

$$PI = [UB, LB] = [\hat{y}(x) + t_{n-p}^{\alpha/2} \sigma(x), \hat{y}(x) - t_{n-p}^{\alpha/2} \sigma(x)] \tag{6}$$

where UB and LB are the upper and lower bands, respectively, n is the total number of discharge observations, p is the total number of model parameters, and $t_{n-p}^{\alpha/2}$ is the $\alpha/2$ percentile for the Student t distribution with $n - p$ degrees of freedom.

3. Materials and methods

3.1. Electricity demand data (G)

In this study, G data were requested from Energex [55] for 01/07/2004 to 30/06/2017. The G data in megawatts (MW) were recorded every 30 min, using Brisbane (the capital city of Queensland) time, from various stations in southeast Queensland, which covered more than 200 suburbs in the cities of Brisbane, Gold Coast, Sunshine Coast, Logan, Ipswich, Redlands, and Moreton Bay. For some stations, there were data limitations, such as missing points, zeros, negative values, and absent G values for previous or subsequent years. Furthermore, at some stations, there were difficulties in matching G data with corresponding SILO and ECMWF Reanalysis datasets. However, the study examined all the stations in the different periods to obtain the best and most accurate target (G data) by avoiding all the data subject to the issues stated above. Consequently, data from eight Energex stations, which included Beerwah, Burpengary, Ipswich South, Narangba,

Redcliffe, Redland Bay, Wynnum, and Zillmere were used from 01/07/2014 to 30/06/2017 for the 6-h forecast horizon, and 01/07/2014 to 29/06/2017 for the daily forecast horizon. The locations of the Energex stations are shown in Fig. 1. The longitudes and latitudes of each station are indicated in Table 1.

The half-hour Energex data phases were converted to 6-h data phases by calculating the sum of every 12 values starting from 01/07/2014 at 4:30 a.m. until 30/06/2017 at 10:00 p.m. As a result, four points, constructed at 10:00 a.m. (from 4:30 a.m. to 10:00 a.m.), 4:00 p.m. (from 10:30 a.m. to 4:00 p.m.), 10:00 p.m. (from 4:30 p.m. to 10:00 p.m.), and 4:00 a.m. (from 10:30 p.m. to 4:00 a.m.), were obtained for each day. These four points were summed to get the daily G values. These data intervals were chosen so that the Energex G data were temporally aligned with the SILO and ECMWF Reanalysis datasets. Fig. 2 shows plots of the G data at the Beerwah site for both daily and 6-h horizons. Table 2 provides descriptive statistics for both forecast horizons at all sites.

3.2. SILO data

Historical climate datasets were extracted from SILO [56] for the period and locations corresponding to the G data retrieved from the eight Energex stations. The SILO station names, longitudes, and latitudes are shown in Table 1 and Fig. 1 in section 3.1. Table 3 shows the details of the SILO data, including the mean values for each variable, at the Beerwah site. The six SILO variables, which included the maximum and minimum air temperature (TMax and TMin), rainfall (Rain), evaporation (Evap), solar radiation (Radn), and vapour pressure (VP) time-series, were recorded every 24 h (daily) and, thus, only used to feed the daily forecasting models.

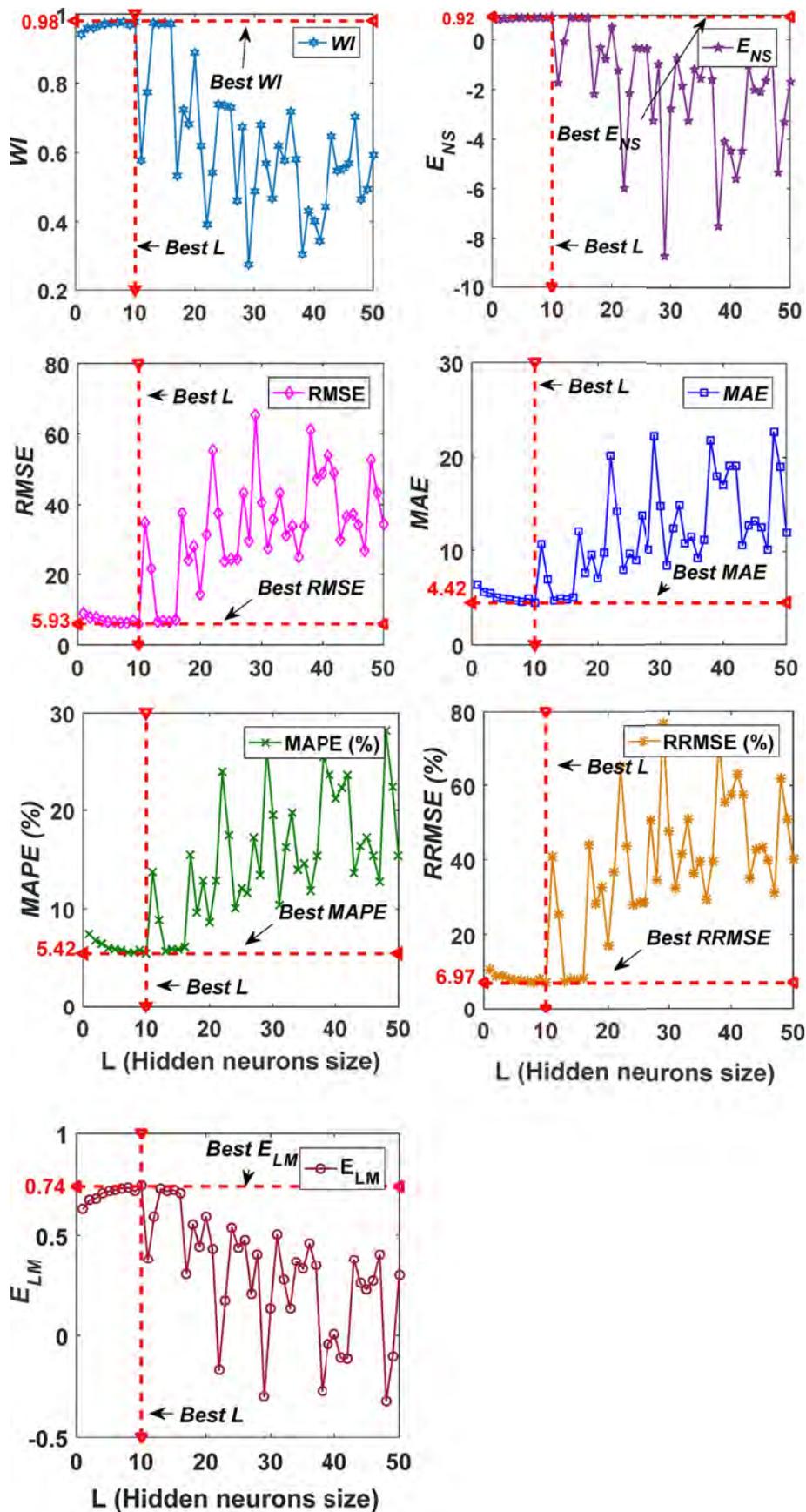


Fig. 5. Selecting the best ANN model in the test period for 6-h Beerwah forecast horizons based on the statistical indicators (Eqs. (7)–(13)). Only the 50 ANN models with the best parameters, indicated in Table 5, are shown.

Table 8

Optimum model performance in the test period for the 6-h forecast horizon as shown by Willmott's index (*WI*), Nash–Sutcliffe efficiency coefficient (*E_{NS}*), root-mean square error (*RMSE*, *MW*), and mean absolute error (*MAE*, *MW*).

Station	ANN				MARS				MLR				ARIMA			
	<i>WI</i>	<i>E_{NS}</i>	<i>RMSE</i> (MW)	<i>MAE</i> (MW)	<i>WI</i>	<i>E_{NS}</i>	<i>RMSE</i> (MW)	<i>MAE</i> (MW)	<i>WI</i>	<i>E_{NS}</i>	<i>RMSE</i> (MW)	<i>MAE</i> (MW)	<i>WI</i>	<i>E_{NS}</i>	<i>RMSE</i> (MW)	<i>MAE</i> (MW)
Beerwah	0.98	0.92	5.93	4.42	0.96	0.88	7.34	5.31	0.95	0.83	8.63	6.39	0.31	-0.06	21.49	16.75
Burpengary	0.98	0.93	8.83	6.27	0.97	0.90	10.90	7.70	0.96	0.85	13.39	9.03	0.39	-0.09	35.58	24.95
Ipswich South	0.98	0.94	7.85	5.89	0.98	0.92	9.10	6.68	0.97	0.89	10.80	7.86	0.35	-0.04	33.22	24.007
Narangba	0.97	0.89	11.10	7.99	0.95	0.81	14.56	10.83	0.93	0.77	16.06	12.07	0.20	-0.02	33.68	27.56
Redcliffe	0.99	0.97	5.03	3.44	0.99	0.96	5.68	4.04	0.99	0.95	6.40	4.58	0.14	0.01	29.34	24.02
Redland Bay	0.98	0.93	6.98	4.88	0.97	0.90	8.69	6.02	0.96	0.87	9.68	6.73	0.41	-0.13	28.99	20.26
Wynnum	0.99	0.95	7.06	5.06	0.98	0.91	9.43	6.70	0.97	0.90	10.09	7.31	0.41	-0.08	32.82	24.97
Zillmere	0.97	0.90	13.35	9.41	0.97	0.88	14.66	10.12	0.96	0.87	15.58	11.02	0.45	-0.19	47.13	35.21

Table 9

The mean absolute percentage error (*MAPE*, %), relative root-mean square error (*RRMSE*, %), and Legates & McCabe's Index (*E_{LM}*) for the optimum models for the 6-h forecast horizon for the testing datasets.

Station	ANN			MARS			MLR			ARIMA		
	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>m</i>	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>E_{LM}</i>	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>E_{LM}</i>	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>E_{LM}</i>
Beerwah	5.42	6.97	0.74	6.37	8.62	0.69	7.57	10.13	0.63	18.93	25.24	0.02
Burpengary	7.44	9.37	0.77	9.21	11.56	0.71	10.20	14.21	0.66	25.16	37.75	0.07
Ipswich South	6.24	7.92	0.77	6.90	9.19	0.74	8.15	10.90	0.69	22.81	33.55	0.07
Narangba	8.06	10.26	0.71	10.93	13.47	0.60	12.13	14.86	0.56	31.07	31.16	-0.008
Redcliffe	2.63	3.88	0.86	3.12	4.39	0.83	3.54	4.94	0.81	19.40	22.65	0.01
Redland Bay	7.06	9.43	0.78	8.84	11.72	0.72	9.96	13.06	0.69	25.62	39.13	0.07
Wynnum	4.45	6.07	0.81	5.86	8.09	0.74	6.41	8.67	0.72	20.29	28.18	0.04
Zillmere	5.31	7.41	0.74	5.65	8.14	0.72	6.14	8.65	0.69	17.76	26.16	0.01

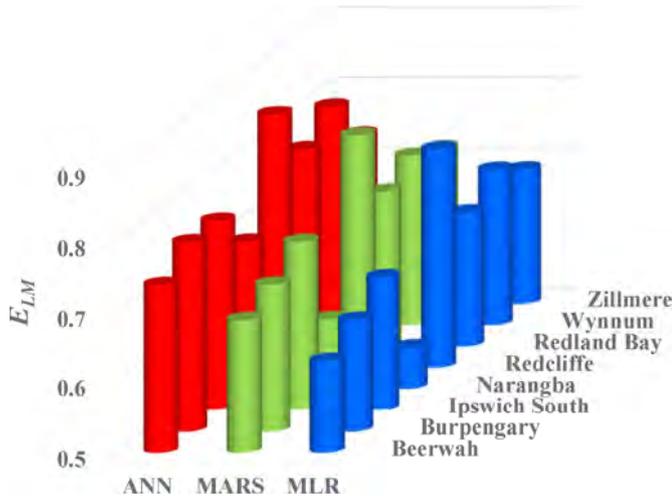


Fig. 6. Bar graphs of the Legates & McCabe's Index (*E_{LM}*) showing the optimal 6-h models for the eight tested stations in southeast Queensland.

3.3. ECMWF (ERA-Interim) Reanalysis data

A variety of factors, such as longitude, latitude, step, time, date, grid, and variable selection, were carefully considered before downloading 'Interim' Reanalysis data, which are available from 1979 until the current date. The data were generated from atmospheric models and observational data based on the European Centre for Medium-Range Weather Forecasts (ECMWF) and global numerical weather prediction models [57]. The data were recorded every 6 h at 00, 06, 12, and 18 UTS time, which was equivalent to 10:00 a.m., 4:00 p.m., 10:00 p.m., and 4:00 a.m. Brisbane time, respectively. In this study, 51 predictor variables were extracted for the same dates and stations as the *G* and SILO datasets. These variables were used for the 6-h forecasts, as well as for the daily forecasts (averaging every four data points).

Table 3 shows all the Reanalysis variables, with their abbreviations and averages, using the Beerwah site data as an example. However, as ECMWF data was in a grid format, the locations of the ECMWF (ERA-Interim) Reanalysis data was not fully aligned with the *G* and SILO data. Therefore, the ECMWF data for four datasets corresponding with the grid corners of the tiles covering each *G* and SILO station were aggregated and averaged into one dataset based the two longitudes and two latitudes shown in Table 1. The ECMWF (ERA-Interim) grids are shown as the pink boxes in Fig. 1.

3.4. Forecast model development

For the ANN, MARS, and MLR models, the historical SILO and ERA-Interim Reanalysis data were used to forecast daily *G* data, while only the Reanalysis data were used for the 6-h forecasting. The data were split into training (80%) and testing (20%) datasets (Table 4) [22]. Three steps were used to build the models in the training period:

1. The higher frequency data from SILO, ECMWF Reanalysis, and *G* datasets were normalized between zero and one using Equation (7) to avoid large numeric ranges from the values of the predictor variables [58].

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{7}$$

2. The best input variables were selected by calculating the cross-correlation between the target (*G*) and the inputs (SILO and ERA-Interim Reanalysis). The variables that showed statistically significant relationships with the *G* data, for a 95% prediction interval in lag 1, were used as an input matrix for the models. For example, the daily forecast horizon using the Beerwah datasets is shown in Fig. 3. Table 5 presents the number of selected variables for all the stations for the 6-h and daily forecast horizons.
3. The statistically significant lags were calculated relating the

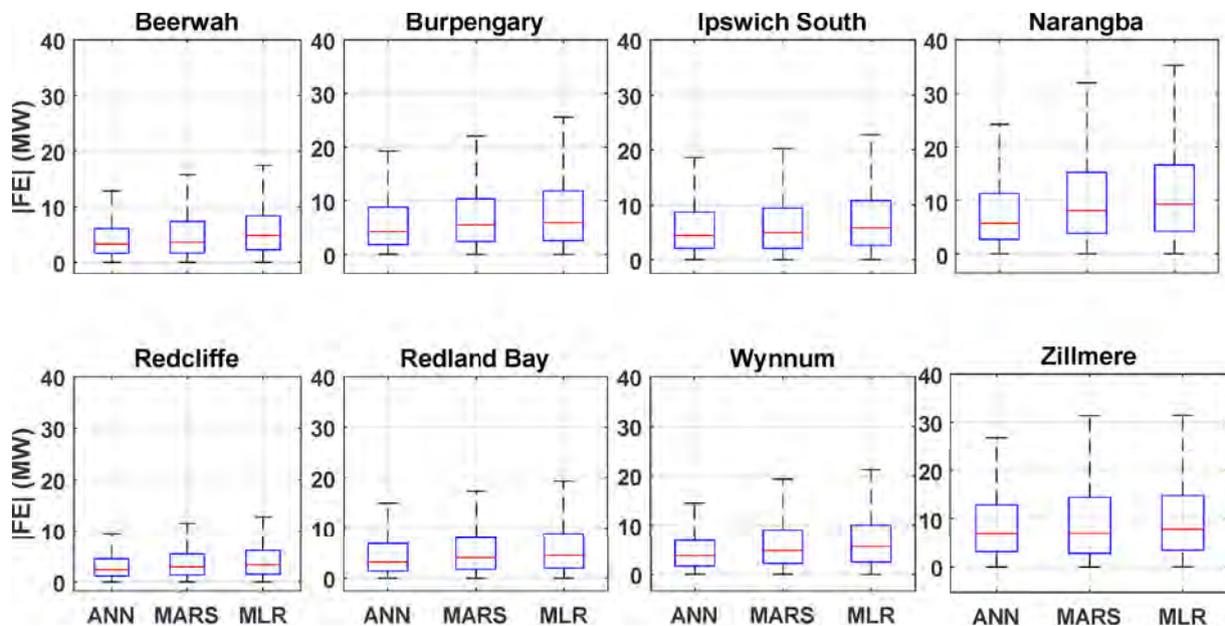


Fig. 7. Boxplots of the 6-h absolute forecasting error, $|FE| = |G_t^{for} - G_t^{obs}|$, in the testing period for the eight stations.

historical G data to the model's input variables through the use of the partial autocorrelation function (PACF). Fig. 4a uses the Beerwah datasets to show the PACF for the daily forecast horizon with the first three statistically significant lags marked. The same technique was applied to the data of the other stations for both forecast horizons. However, for the Burpengary and Ipswich South areas, the 6-h G data showed no or low significance with the first lag, respectively. This helped to identify other highly statistically significant lags, i.e., the second, fourth and fifth lags for Burpengary (Fig. 4b) and the fourth, fifth, and sixth lags, plus the first one, for Ipswich South (Fig. 4c). Table 5 lists the statistically significant lags for both the 6-h and daily forecast horizons using datasets at all the sites.

The main model (ANN) and the comparative models (MARS and MLR) were developed using MATLAB running on an Intel i7 processor at 3.60 GHz. For the ANN algorithm, previous studies (e.g., [31,32]) have indicated that the selection of the training algorithm is an important part of model development. The hidden transfer (F) and output layer ($y(x)$) functions, as well as the optimal hidden neuron size (L), must also be determined. Therefore this study tested two hidden transfer functions (*tansig* and *logsig*) and three output layer functions (*tansig*, *purelin* and *logsig*), as well as L values from 1 to 50. These functions are available in the MATLAB programming environment and are defined below [31,32] in Eqs. (8)–(10). The ANN parameters, which include the training algorithm, hidden transfer function, output layer function, and L were tested one by one, resulting in 600 ANN models with 6-h and daily forecast horizons for each dataset. The optimal ANN model parameters for G forecasting were selected during the testing phase based on the statistical indicators shown in Eqs. (11)–(17). Table 5 describes the structures used for the ANN in the training period.

$$F(x) = \frac{2}{1 + \exp(-2x)} - 1 \text{ Tangent Sigmoid (tan sig)} \tag{8}$$

$$F(x) = \frac{1}{1 + \exp(-x)} \text{ Log Sigmoid (log sig)} \tag{9}$$

$$F(x) = x \text{ Positive Linear (purelin)} \tag{10}$$

The MARS and MLR models were constructed based on the piecewise cubic and linear regression functions, respectively, to evaluate the accuracy of the ANN model [7,8,43,44]. However, in this study, 50

MARS models were developed for each forecast horizon dataset based on the number of the model spline functions (Table 5). The best MARS model was selected based on the statistical metrics (Eqs. (11)–(17)) analysed in the testing period. Additional work, which was not presented previously, showed the importance of estimating the input variables when training the MARS model. This was calculated using the scaled formula of the “square root of the Generalized Cross-Validation (GCV) of the model with all basis functions minus square root of the GCV score of the corresponding full model”, where a value of zero meant that the variable was not used during the training period, while a value close to 100 or zero indicated the variables with the highest and lowest relative importance, respectively [59,60]. Table 5 shows these two variables for the eight stations for both forecasting horizons. The MLR technique was built with the y -intercept utilizing the same input variables shown in Table 5. A single MLR forecast model was constructed in the training phase for each dataset and forecast horizon.

To show the ANN's ability to forecast G as a high-dimensional model, a low-dimensional ARIMA model was built using an R package for automatic time series forecasting [61,62]. In the training phase, the ARIMA model was constructed as a single input model using only G data with no lags, based on the model's three parameters: the autoregressive (p), differencing (d), and moving average (q). The model was evaluated using the Akaike information criterion (AIC), log likelihood (LH), and variance (σ^2) [63]. Table 5 shows the model development for the 6-h and daily forecasting horizons.

To improve the ability of the ANN model to forecast G , a hybrid ANN model was created by combining the forecasted values of the high-dimensional ANN, MARS, and MLR models. The outputs of the three models were used as inputs to feed the hybrid ANN model. After testing this technique, an accurate groundwater risk map [64] and a forecast of soil moisture [65] were created. Additionally, the B method, a powerful pre-processing forecasting tool, was applied in this research paper to investigate the uncertainty in G forecasting using an ensemble of the hybrid ANN model. The B-hybrid ANN model was then developed using 200 bootstrap samples of the raw datasets in the training period. These 200 constructed hybrid ANN models were used to build the 95% prediction bands from the testing datasets using Eq. (6). Table 6 shows the optimum parameters for the hybrid ANN and B-hybrid ANN models in the training period for both forecasting horizons.

The performance of the models in the training period is summarized in Table 7 based on the root-mean square error (RMSE, Eq. (11)) using

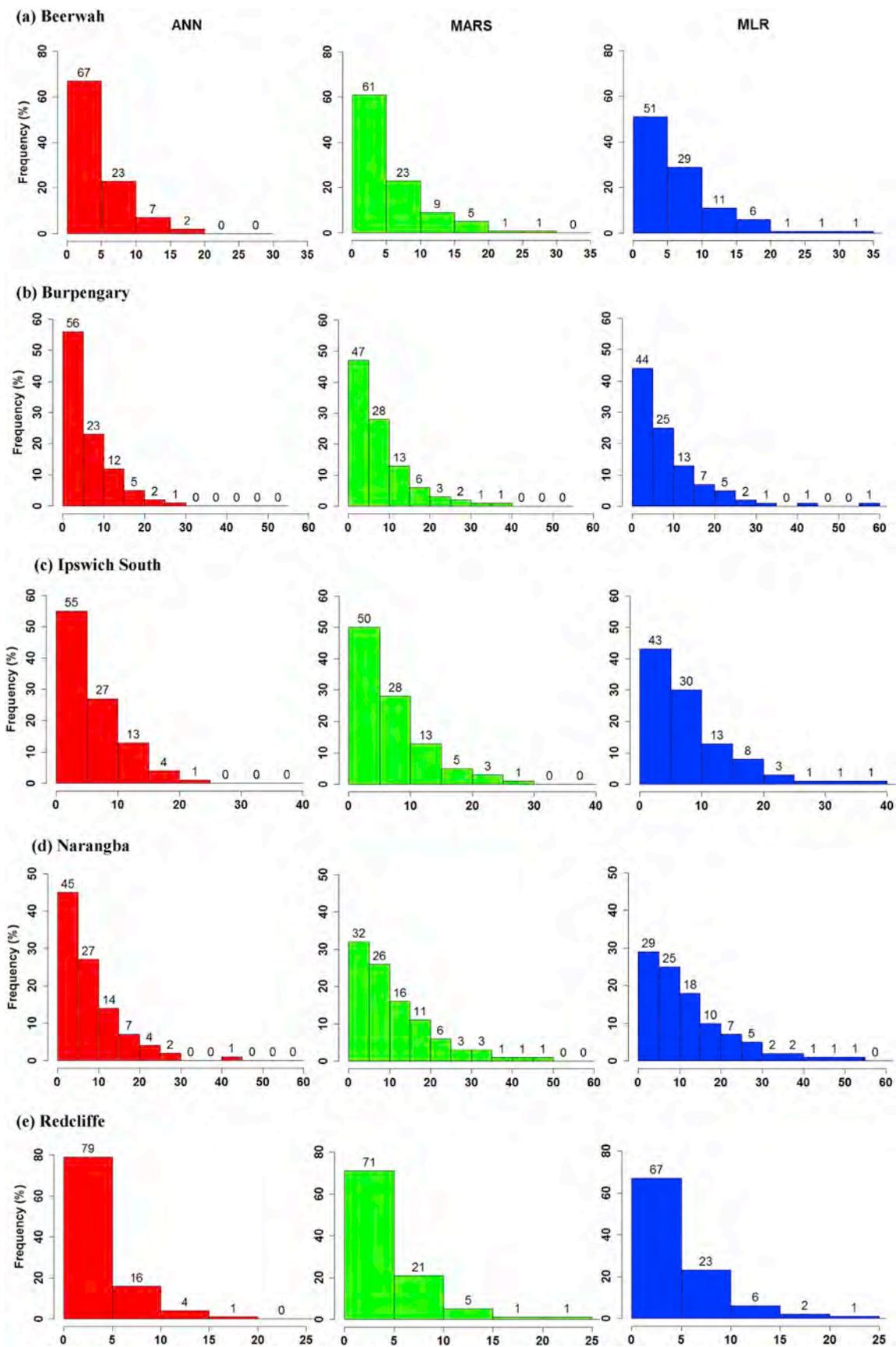


Fig. 8. The percentage frequency distribution of the 6-h forecasted error, $|FE|$, for the eight stations in the testing phase.

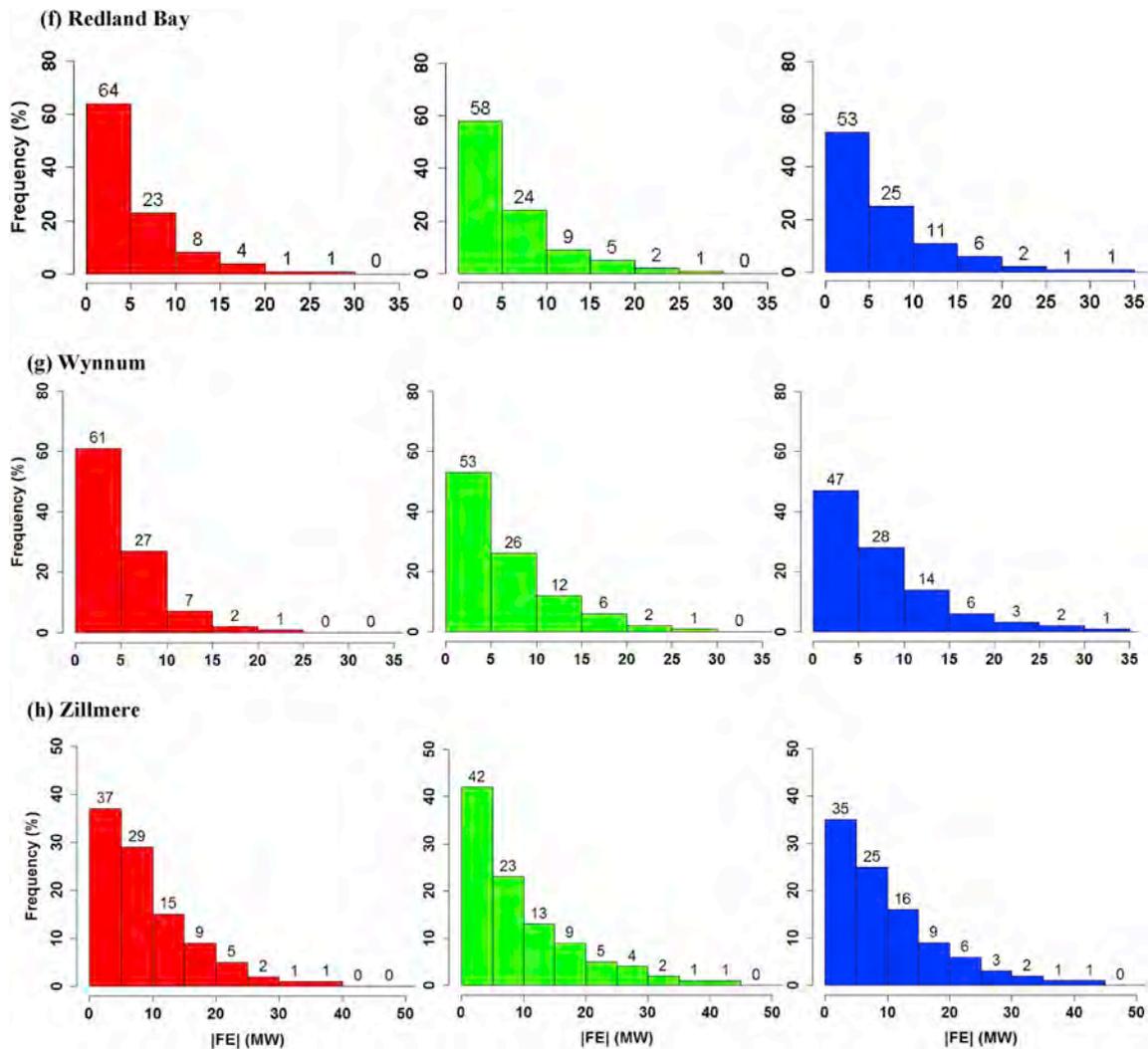


Fig. 8. (continued)

Table 10

Model performance in the test period for the daily forecast horizon by Willmott's index (WI), Nash–Sutcliffe efficiency coefficient (E_{NS}), root-mean square error ($RMSE$, MW), and mean absolute error (MAE , MW).

Station	ANN				MARS				MLR				ARIMA			
	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)
Beerwah	0.89	0.68	28.47	21.20	0.88	0.67	29.20	22.33	0.87	0.65	29.78	22.90	0.44	-0.57	63.19	48.46
Burpengary	0.95	0.81	33.45	24.58	0.93	0.79	35.36	25.26	0.93	0.78	36.49	27.07	0.43	-0.54	96.27	68.19
Ipswich South	0.94	0.80	36.32	27.39	0.93	0.79	37.27	27.49	0.93	0.78	38.24	29.48	0.41	-0.35	94.34	68.79
Narangba	0.89	0.68	51.04	37.68	0.86	0.60	57.38	46.02	0.84	0.58	58.80	48.03	0.44	-0.67	116.70	97.86
Redcliffe	0.96	0.85	25.80	19.59	0.96	0.84	26.02	19.93	0.95	0.83	27.34	20.81	0.35	-0.13	69.90	56.79
Redland Bay	0.92	0.75	28.17	19.50	0.92	0.75	28.29	19.14	0.93	0.78	26.43	19.25	0.41	-0.40	67.21	47.77
Wynnum	0.94	0.80	34.34	25.50	0.92	0.77	36.23	27.04	0.93	0.78	35.98	27.11	0.39	-0.29	86.30	66.22
Zillmere	0.88	0.65	55.83	42.45	0.88	0.63	57.08	44.26	0.88	0.65	55.95	43.88	0.43	-0.97	132.17	103.17

the datasets for all the stations for both forecast horizons. For the 6-h forecasts, the hybrid ANN model yielded the highest accuracy with the lowest $RMSE$ for all datasets compared with the ANN, MARS, MLR, and ARIMA models. Similarly, for the daily forecasting, the hybrid ANN model outperformed the ANN, MARS, MLR, and ARIMA models for all stations. Table 7 presents the details of the models' evaluation in the training phase.

3.5. Model prediction quality

Seven statistical metrics, including Legates and McCabe's Index (E_{LM}), Willmott's Index (WI), and Nash–Sutcliffe efficiency coefficient (E_{NS}), as well as mean absolute error (MAE) and root-mean square error ($RMSE$) together with their relative error values ($MAPE$, % and $RRMSE$, %) were employed to test the performance of the models for the 6-h and daily G forecast horizons in the testing period. The equations for the statistical indicators are listed below [66–75]:

Table 11

The mean absolute percentage error (*MAPE*, %), relative root-mean square error (*RRMSE*, %), and Legates & McCabe's Index (*E_{LM}*) for the optimum models with daily forecast horizon in the test datasets.

Station	ANN			MARS			MLR			ARIMA		
	<i>MAPE</i> (%)	<i>RRMSE</i> (%)	<i>E_{LM}</i>									
Beerwah	6.31	8.36	0.47	6.53	8.57	0.44	6.77	8.74	0.43	13.15	18.55	-0.20
Burpengary	6.38	8.87	0.60	6.51	9.38	0.59	7.04	9.68	0.56	15.83	25.53	-0.11
Ipswich South	6.79	9.17	0.57	6.87	9.41	0.57	7.37	9.65	0.54	15.60	23.82	-0.07
Narangba	9.01	11.80	0.47	11.13	13.26	0.35	11.67	13.59	0.32	21.40	26.98	-0.38
Redcliffe	3.88	4.98	0.65	3.89	5.02	0.65	4.12	5.28	0.63	10.53	13.49	-0.01
Redland Bay	6.19	9.51	0.57	5.99	9.55	0.58	6.20	8.92	0.58	14.31	22.69	-0.05
Wynnum	5.50	7.37	0.59	5.76	7.77	0.57	5.86	7.72	0.57	13.25	18.52	-0.05
Zillmere	5.83	7.75	0.40	6.09	7.92	0.38	6.04	7.77	0.38	13.32	18.35	-0.45

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2} \tag{11}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |G_i^{for} - G_i^{obs}| \tag{12}$$

$$RRMSE = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}}{\overline{G^{obs}}} \tag{13}$$

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^{i=n} \left| \frac{G_i^{for} - G_i^{obs}}{G_i^{obs}} \right| \tag{14}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (|G_i^{for} - \overline{G^{obs}}| + |G_i^{obs} - \overline{G^{obs}}|)^2} \right], \text{ and } 0 \leq WI \leq 1 \tag{15}$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (G_i^{obs} - \overline{G^{obs}})^2} \right], \text{ and } \infty \leq E_{NS} \leq 1 \tag{16}$$

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^{i=n} |G_i^{obs} - G_i^{for}|}{\sum_{i=1}^{i=n} |G_i^{obs} - \overline{G^{obs}}|} \right], \text{ and } (\infty \leq E_{LM} \leq 1) \tag{17}$$

where G_i^{for} , G_i^{obs} , $\overline{G^{for}}$ and $\overline{G^{obs}}$ are the forecasted, observed, mean of forecasted, and mean of observed values of *G*, respectively, and *n* is the total number of G_i^{for} (or G_i^{obs}) values in the testing period.

Since a single indicator would not be able to show all shortcomings of the models used, multiple statistical criteria were used to assess each model's performance [74]. The best performing model had values of *RMSE* and *MAE* that were closest to zero and values of *WI*, *E_{LM}*, and *E_{NS}* that were closest to one. For comparative analysis, *MAPE* (%) and *RRMSE* (%) were also used to describe the models' behaviour over a range of statistically different hydrological flows [32]. Specifically, a model's performance was considered excellent with *RRMSE* < 10%, good with 10% < *RRMSE* < 20%, fair with 20% < *RRMSE* < 30%, and poor with *RRMSE* > 30% [66,76–78].

4. Results and discussion

4.1. 6-h forecast horizon

The models were assessed using the testing dataset based on the seven statistical criteria shown in Eqs. (11)–(17). The best model parameters were selected according to the lowest *RMSE*, *MAE*, *RRMSE*, and *MAPE* values, as well as the highest *WI*, *E_{NS}*, and *E_{LM}* values. The best overall model was also selected using these criteria. For the ANN, a total 600 models were created for each station's dataset with two types of training algorithms and hidden transfer functions, as well as three types of output layer functions and *L* between 1 and 50. For the 6-h Beerwah forecast horizon, the best ANN parameters were found to be

trainlm, *logsig*, *tansig*, and 10 for the training algorithm, hidden transfer function, output layer function, and *L*, respectively. Table 5 shows these factors in detail for all models, stations, and forecast horizons, while Fig. 5 illustrates the best ANN model with *L* equal to 10. Similarly, Eqs. (11)–(17) helped to select the best number of spline functions for the MARS model, whereas only one MLR model and one ARIMA model were developed in this study.

According to the results, the low-dimensional ARIMA model had the lowest accuracy (Tables 8 and 9), in contrast to the high-dimensional ANN, MARS, and MLR models, which likely performed better because of their use of the extensive data from the SILO and ERA-Interim Re-analysis datasets. When comparing the three high-dimensional models developed in this study, the ANN model outperformed the MARS and MLR models for all eight datasets (Table 8). For example, for the Beerwah dataset, *WI* = 0.98, *E_{NS}* = 0.92, *RMSE* = 5.93 MW, and *MAE* = 4.42 MW (for ANN) compared with *WI* = 0.96, *E_{NS}* = 0.88, *RMSE* = 7.34 MW, and *MAE* = 5.31 MW (for MARS) and *WI* = 0.95, *E_{NS}* = 0.83, *RMSE* = 8.63 MW, and *MAE* = 6.93 MW (for MLR). Similar results were obtained for all other stations indicating that the ANN model had the highest accuracy when forecasting *G*.

The performance of the ANN model was further evaluated using three other metrics, namely *MAPE*, *RRMSE*, and *E_{LM}* (Table 9). These statistical indicators were extracted from *MAE*, *RMSE*, and *WI* to calculate the 'goodness-of-fit' between the G_i^{for} and G_i^{obs} values (e.g., [7,8]). For all the forecasting datasets, the ANN model outperformed the MARS and MLR models. For the Beerwah dataset, the ANN model had an *MAPE* = 5.42%, *RRMSE* = 6.97%, and *E_{LM}* = 0.74, while the MARS model had an *MAPE* = 6.37%, *RRMSE* = 8.62%, and *E_{LM}* = 0.69 and the MLR model had an *MAPE* = 7.57%, *RRMSE* = 10.13% and *E_{LM}* = 0.63. Table 9 shows these results for all stations, while Fig. 6 shows bar graphs of *E_{LM}* metric for the three forecasting models and all stations.

Boxplots and histograms were plotted for the absolute forecasted error statistics, $|FE| = |G_i^{for} - G_i^{obs}|$, for the 6-h forecasts of the ANN, MARS, and MLR models to examine the forecasting datasets. The boxplots (Fig. 7) display the lower (first line), median (second line), and upper (third line) quartiles of the $|FE|$. Based on those quartiles, larger error values were evident for the MARS and MLR models, in comparison to the ANN. The histograms (Fig. 8) show the percentage of error frequency distributions for the $|FE|$ for all datasets. For the first seven stations, the ANN model had the least error, which was indicated by the highest percentage of error frequency in the smallest (0 to ± 5 MW) bracket for the 6-h G^{for} . Although the MARS model achieved better results than the ANN model (i.e., 42% vs. 37% of errors in the 0 to ± 5 MW bracket, respectively) for the Zillmere station forecasting dataset, the ANN model outperformed the MARS and MLR models in the 0 to ± 10 MW bracket (i.e., 66% vs. 65% vs. 60%, respectively). Generally, the lower prediction error for the ANN model forecasts demonstrated that this model had better accuracy for G^{for} than the MARS and MLR models.

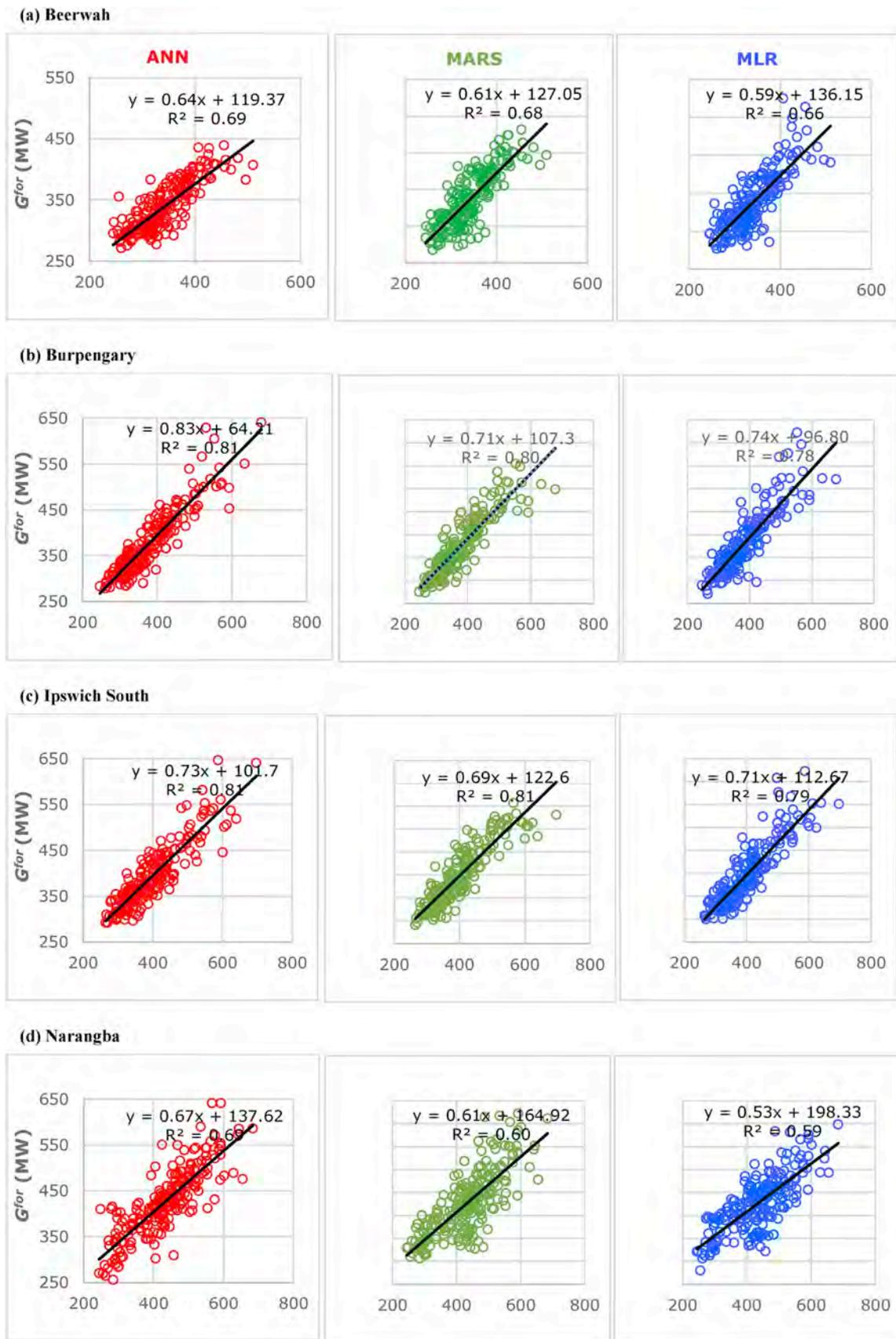


Fig. 9. Scatterplot of the daily forecasted electricity demand (G^{for}) vs. observed electricity demand (G^{obs}) in the testing phase for the eight stations.

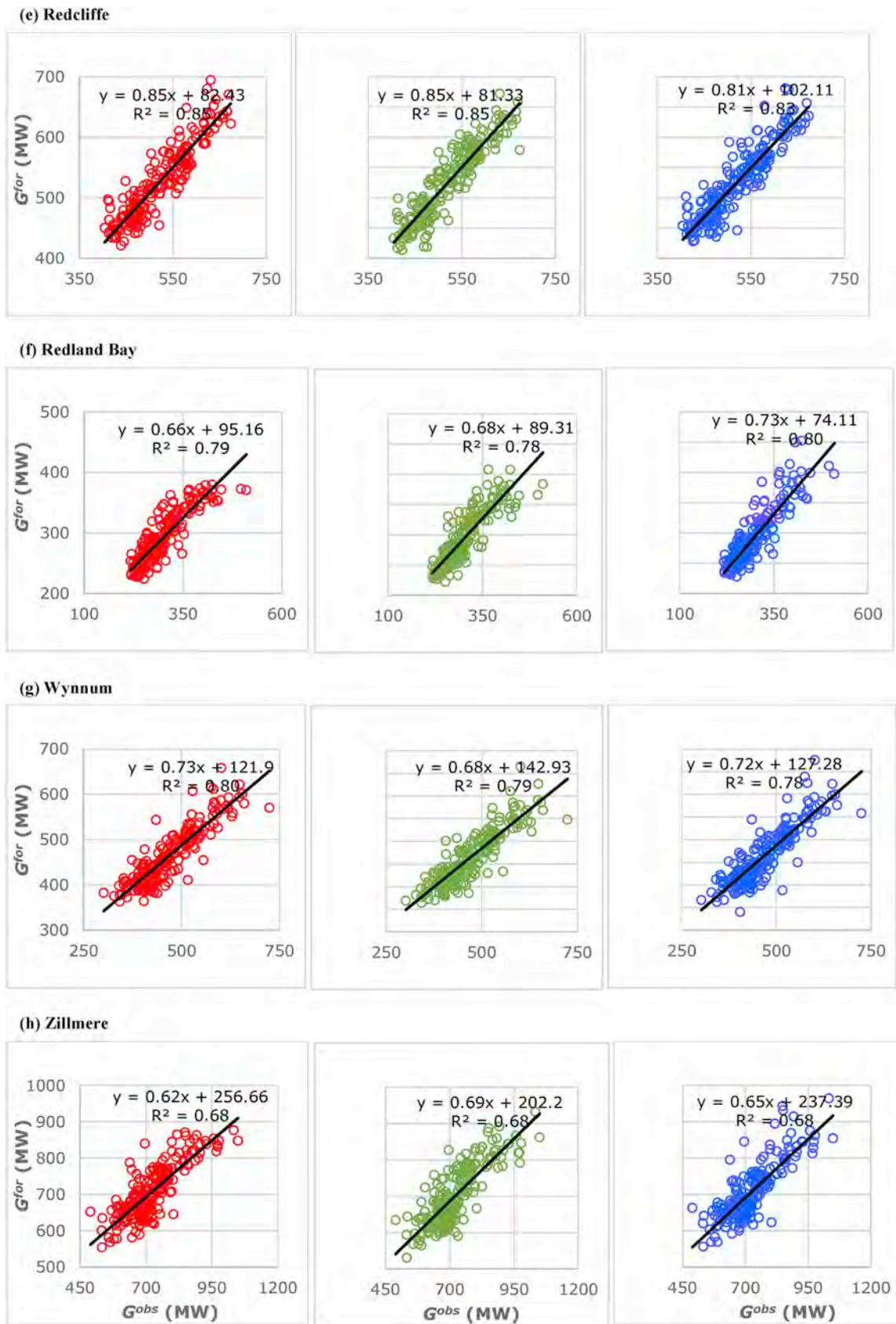


Fig. 9. (continued)

4.2. Daily forecast horizon

The ability to forecast G over a longer forecasting horizon was

evaluated for the ANN, MARS, and MLR models, using a daily time step. Again, the ARIMA model underperformed compared to the other models for daily G forecasting. Tables 10 and 11 summarize the

Table 12

Optimum hybrid ANN model performance in the test period for both forecast horizons by Willmott's index (WI), Nash–Sutcliffe efficiency coefficient (E_{NS}), root-mean square error ($RMSE$, MW), mean absolute error (MAE , MW), mean absolute percentage error ($MAPE$, %), relative root-mean square error ($RRMSE$, %), and Legates & McCabe's Index (E_{LM}).

Station	6-h						daily							
	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)	$MAPE$ (%)	$RRMSE$ (%)	E_{LM}	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)	$MAPE$ (%)	$RRMSE$ (%)	E_{LM}
Beerwah	0.98	0.92	5.84	4.29	5.29	6.86	0.75	0.90	0.69	28.05	20.69	6.12	8.23	0.49
Burpengary	0.98	0.94	8.65	6.14	7.34	9.18	0.77	0.95	0.83	32.20	23.85	6.20	8.54	0.61
Ipswich South	0.99	0.95	7.63	5.76	6.13	7.71	0.78	0.94	0.81	35.18	25.90	6.45	8.88	0.60
Narangba	0.97	0.89	10.96	7.85	7.94	10.14	0.71	0.90	0.70	49.44	37.03	8.93	11.43	0.48
Redcliffe	0.99	0.97	4.99	3.39	2.57	3.85	0.86	0.96	0.86	24.48	18.46	3.63	4.73	0.67
Redland Bay	0.98	0.94	6.85	4.80	6.95	9.25	0.78	0.94	0.79	26.04	18.05	5.78	8.79	0.60
Wynnum	0.99	0.95	7.03	5.03	4.40	6.04	0.81	0.93	0.80	34.33	25.25	5.39	7.37	0.60
Zillmere	0.98	0.91	13.04	9.21	5.22	7.24	0.74	0.90	0.68	53.09	41.48	5.75	7.37	0.42

performance of each model. For most of the stations' datasets, the ANN model only achieved the highest forecasting accuracy by a small margin, in comparison to the MARS and MLR models. However, for the Narangba station dataset the ANN model significantly outperformed the other models, with $WI = 0.89$, $E_{NS} = 0.68$, $RMSE = 51.04$ MW, $MAE = 37.68$ MW, $MAPE = 9.01\%$, $RRMSE = 11.80\%$, and $E_{LM} = 0.47$, in comparison to the MARS model with $WI = 0.86$, $E_{NS} = 0.60$, $RMSE = 57.38$ MW, $MAE = 46.02$ MW, $MAPE = 11.13\%$, $RRMSE = 13.26\%$, and $E_{LM} = 0.35$, and the MLR model with $WI = 0.84$, $E_{NS} = 0.58$, $RMSE = 58.80$ MW, $MAE = 48.03$ MW, $MAPE = 11.67\%$, $RRMSE = 13.59$, and $E_{LM} = 0.32$. Overall, considering the seven statistical indicators used in this study, better G^{for} was yielded by the ANN model when compared to the MARS and MLR models.

Fig. 9 shows scatterplots with the least squares regression line, $[y(G_i^{for}) = aG_i^{obs} + b]$, and correlation of determination (R^2) used to evaluate the relationship between G_i^{for} and G_i^{obs} for all the G^{for} station datasets. For six stations' datasets (Beerwah, Burpengary, Ipswich South, Narangba, Redcliffe, and Wynnum), based on the values of a , b , and R^2 , the ANN model obtained better results than the MARS and MLR models, especially for forecasting data at the Narangba station. The outcomes were $a = 0.67$, $b = 137.62$, and $R^2 = 0.69$ (for ANN), $a = 0.61$, $b = 164.92$, and $R^2 = 0.60$ (for MARS), and $a = 0.53$, $b = 198.33$, and $R^2 = 0.59$ (for MLR). However, for Zillmere and Redland Bay stations, MARS and MLR outperformed the ANN by a small margin.

4.3. Hybrid ANN and uncertainty assessment using a B-hybrid ANN

To further enhance the accuracy of the ANN model, a hybrid ANN model was developed using the forecast values of the ANN, MARS, and MLR models as inputs. Table 12 shows the performance of the hybrid ANN model for the 6-h and daily forecasting horizons during the testing period. The daily forecasts at the Redland Bay site showed the most significant improvement of hybrid ANN ($WI = 0.94$, $E_{NS} = 0.79$, $RMSE = 26.04$ MW, $MAE = 18.05$ MW, $MAPE = 5.78\%$, $RRMSE = 8.79\%$, and $E_{LM} = 0.60$) in comparison to the standard ANN ($WI = 0.92$, $E_{NS} = 0.75$, $RMSE = 28.17$ MW, $MAE = 19.50$ MW, $MAPE = 6.19\%$, $RRMSE = 9.51\%$, and $E_{LM} = 0.57$). Consequently, G was found to be most accurately forecasted using the hybrid technique, which could support the National Electricity Market in Queensland, Australia, and potentially other areas as well.

To estimate the forecasting bands for the 6-h and daily horizons, ensemble-based uncertainty assessments via the bootstrapping procedure (B) was used in this paper in combination with the hybrid ANN model. The B-hybrid ANN model was developed using 200 B samples with replacement. The 95% prediction bands were built using Eq. (6) based on the averages and variances of the 200 forecasted values in the test datasets. As shown in Fig. 10, the lower and upper forecasted bands are close to the observed values. This demonstrates the ability and

reliability of the hybrid ANN model to forecast G by addressing some of the uncertainty associated with forecasted values.

5. Further discussion, limitations and opportunity for future work

Developing a globally trained model for electricity demand forecasting was identified as an important component for supporting policy development related to energy technology by the United Nations Development Program on sustainable energy. Recently, national electricity markets globally have reported high forecasting errors due to the use of old techniques to predict G , which has encouraged researchers to develop more accurate forecasting strategies. Since G usage is expected to increase (e.g., [2,79,80]), there is a need to develop new methods to accurately forecast G to support the transition to sustainable energy resources, as stated in Goal 7 of the UN SDGs. The aim of this study was to support the National Electricity Market in Australia, and in other areas, by constructing a highly accurate G forecasting model with an ANN algorithm using a wide range of datasets extracted from three sources: SILO and ECMWF (ERA-Interim) for the input variables, and Energex for the target variable (G). The forecasting performance of the developed ANN model was compared with the MARS, MLR, and ARIMA models based on several statistical indicators using data from southeast Queensland, Australia. Furthermore, a hybrid ANN model was designed and compared with the standard ANN model, while the PI was calculated using the B method integrated with the hybrid model to evaluate the uncertainty assessments obtained from the forecast values. This study found that the hybrid ANN model performed well and generally outperformed the ANN, MARS, MLR, and ARIMA models in forecasting G .

This study was the first to use extensive datasets from SILO and ECMWF (ERA-Interim) Reanalysis to develop an accurate ANN model for G forecasting. However, some challenges should be addressed in future works. Since the SILO and ERA-Interim Reanalysis data were not available for every single Energex station in the study area, future research should consider using satellite data. For example, satellite data can be extracted from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) and utilized as an alternative predictor dataset to develop subsequent G forecasting models. Hourly variables over a $0.05^\circ \times 0.05^\circ$ grid resolution can be generated from this source with two primary sensors (Terra and Aqua MODIS) [31,81–83]. Satellite data have successfully been used to forecast solar radiation in the past [31,57,84], which resulted in high-quality ANN and self-adaptive differential evolutionary extreme learning machines (SaDE-ELM) models. Therefore, further research could investigate creating G forecasting models by combining datasets from alternate sources, such as SILO, Reanalysis and/or satellite data.

Another challenge with the current methodology is that non-climatic factors were not considered in this study. Since energy demand is a multivariate problem, social and population variables may need to be taken into account to further improve the accuracy of G forecasts.

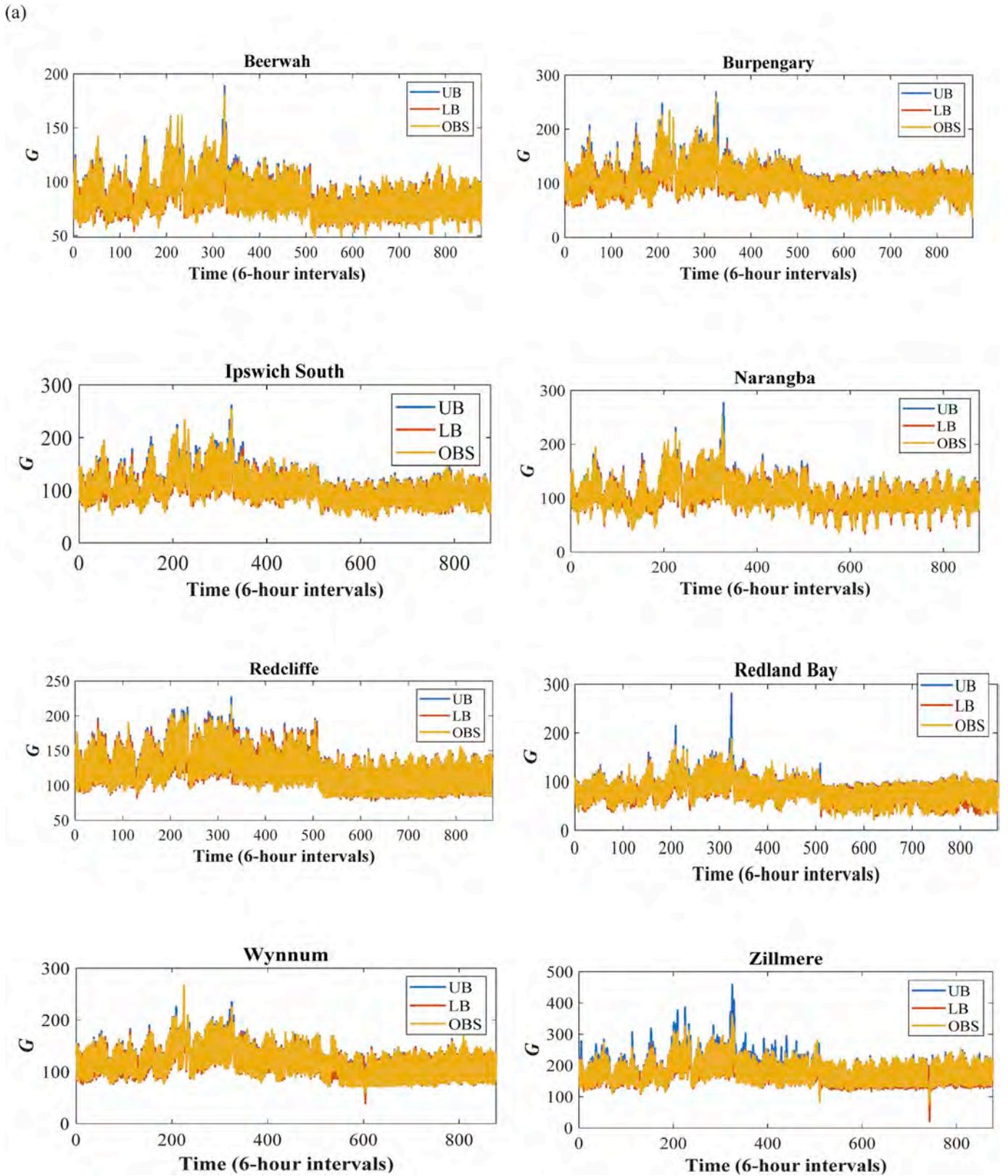


Fig. 10. Electricity demand (G) observed with 95% prediction bands forecasted using a B-hybrid-ANN for all the eight stations for the (a) 6-h and (b) daily forecasting horizons.

However, this may be addressed in an upcoming work on deep learning or long-short term memory network methods (e.g., [85]) using real-life energy predictor variables, for instance, population, that can be extracted from the Australian Bureau of Statistics [86].

Different pre-processing techniques could also be used to improve the forecasting performance of the ANN model. Firstly, a suitable feature selection method, such as iterative input selection (IIS) [87], grouping genetic algorithm (GGA) [88], or coral reef optimisation

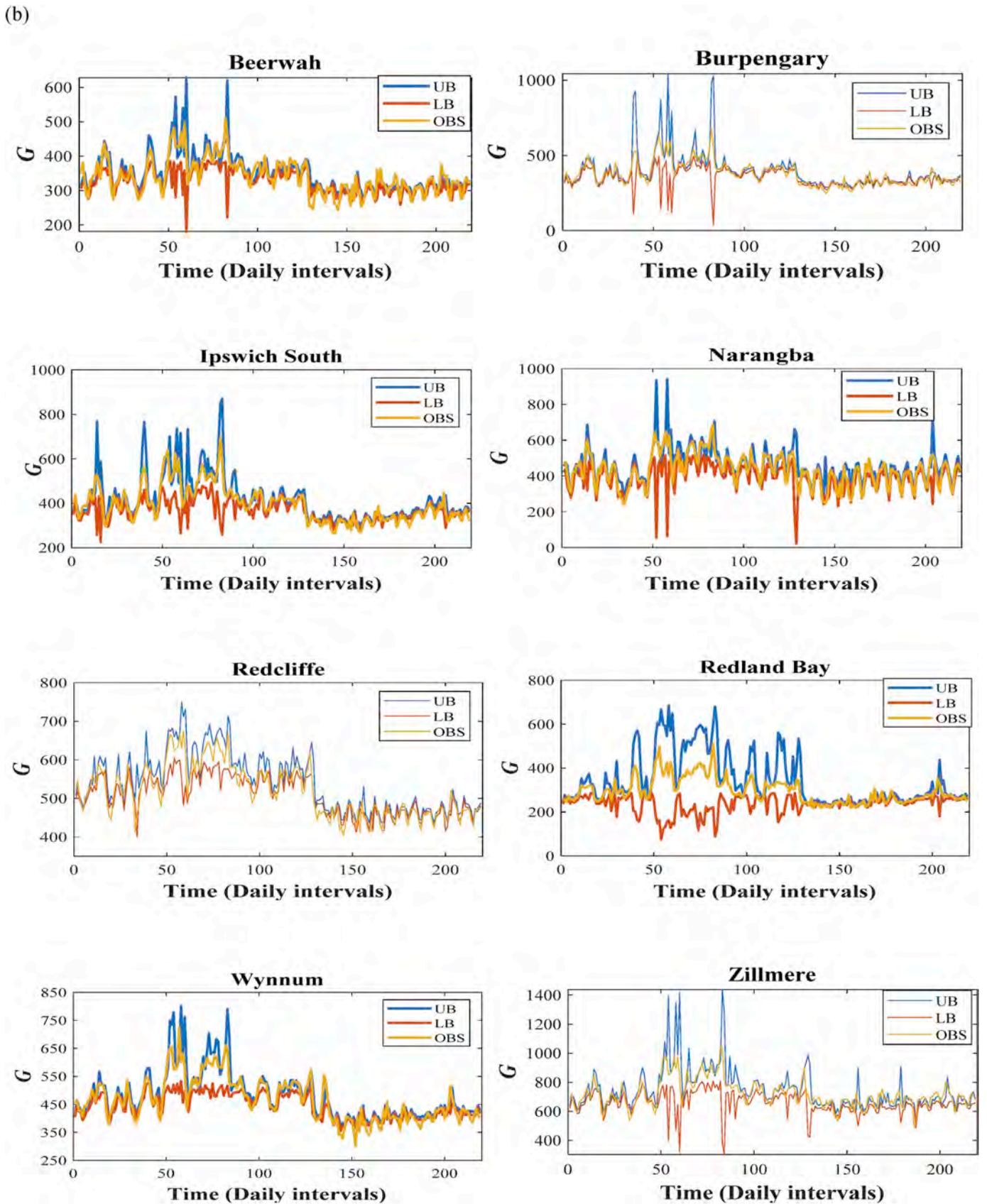


Fig. 10. (continued)

(CRO) [89,90] could be applied with a deep learning strategy or long-short term memory network to select the input variables that have significant influence on the model. This would help reduce both the size

of the predictor datasets used to forecast electricity demand and model complexity. Secondly, wavelet transformations (WT) (e.g., [22,32,35,91]) could be used to address non-stationary data, which

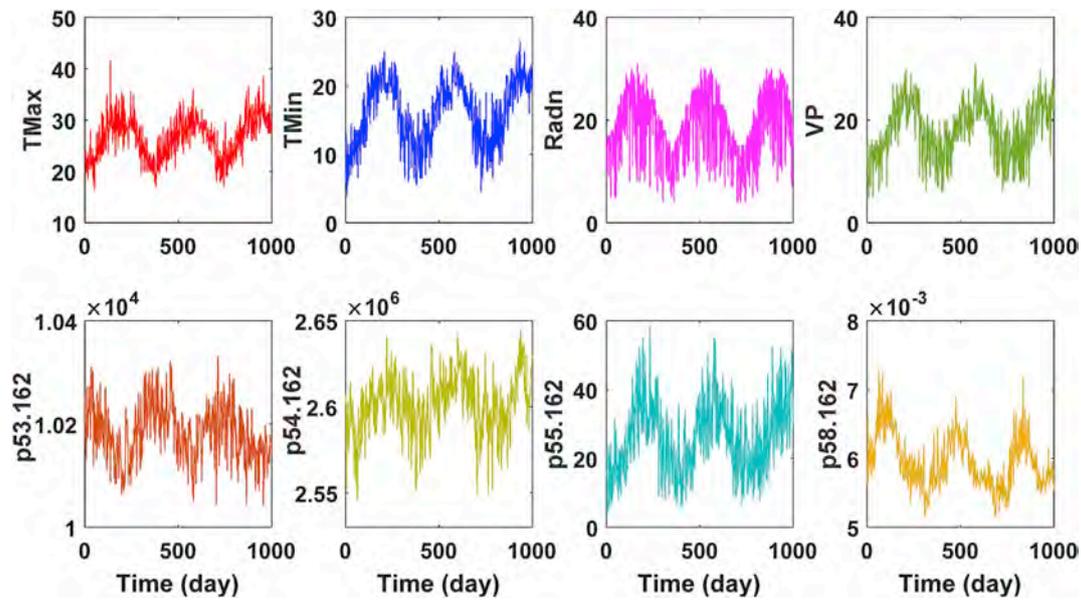


Fig. 11. Non-stationary daily time-series of the eight variables from SILO and ECMWF (ERA-Interim) Reanalysis fields.

appeared in the current SILO and ERA-Interim Reanalysis inputs variables (e.g., Fig. 11), by decomposing the data into high and low-pass filters [22]. However, Quilty and Adamowski [92] stated that WT have been incorrectly used in several research studies to forecast real-world datasets, as they created errors in the inputs of a forecasting model during wavelet decomposition, which also should be addressed in future research. Furthermore, a bootstrapping algorithm (B) or a Bayesian model averaging (BMA) technique could be used with WT and one of the feature selection techniques, such as IIS, to further improve G forecasting accuracy (e.g., [32,37,93]). Additionally, the G forecasting accuracy could be enhanced by using the Firefly algorithm (FA), which has been tested with backpropagation (BP) [94] to forecast electricity prices, to optimize the parameters of the ANN model. Hence, based on the methods suggested above, a hybrid model could be constructed as IIS-WT-B-FA-ANN or IIS-WT-BMA-FA-ANN to forecast G data.

6. Concluding remarks

In this paper, the normal and hybrid ANN models were introduced for 6-h and daily G forecasting horizons using datasets for southeast Queensland, Australia, extracted from Energen, SILO, and ECMWF. Three other models (MARS, MLR, and ARIMA) were also employed in this work to allow for a comparative evaluation of the forecasting accuracy of the ANN and hybrid ANN models. The performance accuracy of the models was evaluated using seven statistical metrics as presented in section 3.5. The results concluded that the Hybrid ANN model achieved the best forecasting accuracy in comparison to other models developed in this study.

In spite of the advantages of the ANN model for G forecasting, the ANN forecasting performance could be improved in future studies by integrating an ANN with different techniques, such as IIS for feature selection, WT-B or WT-BMA for data decomposition and uncertainty assessment, and FA for the ANN optimization problem in order to construct a hybrid model of IIS-WT-B-FA-ANN or IIS-WT-BMA-FA-ANN. Follow-up research studies could also explore the use of shrinkage methods, such as Lasso and elastic-net, including Lasso via hierarchical interactions with the ANN model to potentially further improve forecasting accuracy.

In summary, the ANN model reliably achieved an acceptable G forecasting accuracy and could be used to support the National Electricity Markets, such as Energen. In particular, this study found that the ANN model generally outperformed the MARS, MLR, and AIRMA

models. Furthermore, the hybrid ANN model outperformed the standard ANN. However, the limitations described above should be carefully addressed in future research works.

Acknowledgments

We would like to thank Energen, Scientific Information for Land Owners (SILO) and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the required data. We are also grateful to the Ministry of Higher Education and Scientific Research in the Government of Iraq for funding the first author's Ph.D. project. Furthermore, thank you to the University of Southern Queensland for providing access to different academic services. Special thanks to Dr. Barbara Harmes (Language Centre, Open Access College, University of Southern Queensland, Australia), Ms. Mela Aryal (Ph.D. student, University of Southern Queensland, Australia), and Ms. Laura Gilbert (Ph.D. student in Natural Resource Sciences, McGill University, Canada) for providing help in proof-reading this paper. We also would like to thank Dr. Mukesh Tiwari (Department of Irrigation & Drainage Engineering, College of Agricultural Engineering and Technology, Anand Agricultural University, Gujarat, India) for his advice on the bootstrap (B) technique. Finally, we thank both reviewers and the Editor-In-Chief for their comments that have improved this paper.

References

- [1] Assembly UNG. Resolution adopted by the general assembly on 25 september 2015. Washington: United Nations; 2015.
- [2] Smith M, et al. Energy transformed: sustainable energy solutions for climate change mitigation. The Natural Edge Project, CSIRO, and Griffith University; 2007.
- [3] Brinsmead TS, et al. Australian electricity market analysis report to 2020 and 2030 2014. CSIRO Report No. EP141067.
- [4] Pai P-F, Hong W-C. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electr Power Syst Res* 2005;74(3):417–25.
- [5] AEMO. Aggregated Price and Demand Data - Historical. The Australian Energy Market Operator is responsible for operating Australia's largest gas and electricity markets and power systems]. 2018 <http://www.aemo.com.au/>.
- [6] Schimek, M.G. and B.A. Turlach, *Additive and generalized additive models: a survey*. 1998, discussion papers, interdisciplinary research project vol. 373: Quantification.
- [7] Al-Musaylh MS, et al. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Adv Eng Inf* 2018;35:1–16.
- [8] Al-Musaylh MS, et al. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. *Appl Energy* 2018;217:422–39.

- [9] Taylor JW. Short-term electricity demand forecasting using double seasonal exponential smoothing. *J Oper Res Soc* 2003;54(8):799–805.
- [10] Taylor JW. Triple seasonal methods for short-term electricity demand forecasting. *Eur J Oper Res* 2010;204(1):139–52.
- [11] Pang Y, et al. Hierarchical electricity demand forecasting by exploring the electricity consumption patterns. *ICPRAM*. 2018, 576–581.
- [12] Fiot J-B, Dinuzzo F. Electricity demand forecasting by multi-task learning. *IEEE Trans on Smart Grid* 2018;9(2):544–51.
- [13] Yang J, Rivard H, Zmeureanu R. On-line building energy prediction using adaptive artificial neural networks. *Energy Build* 2005;37(12):1250–9.
- [14] Yang J, Rivard H, Zmeureanu R. Building energy prediction with adaptive artificial neural networks. Ninth international IBPSA conference. Canada: Montréal; 2005. August.
- [15] Szkuta B, Sanabria LA, Dillon TS. Electricity price short-term forecasting using artificial neural networks. *IEEE Trans Power Syst* 1999;14(3):851–7.
- [16] Kandanand K. Forecasting electricity demand in Thailand with an artificial neural network approach. *Energies* 2011;4(8):1246–57.
- [17] Ghiassi M, Zimbra DK, Saidane H. Medium term system load forecasting with a dynamic artificial neural network model. *Electr Power Syst Res* 2006;76(5):302–16.
- [18] Mirasgedis S, et al. Models for mid-term electricity demand forecasting incorporating weather influences. *Energy* 2006;31(2–3):208–27.
- [19] Lebotsa ME, et al. Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. *Appl Energy* 2018;222:104–18.
- [20] Mohanad SA-M, Ravinesh CD, Yan L. Particle swarm optimized-support vector regression hybrid model for daily horizon electricity demand forecasting using climate dataset. *E3S web of conferences*. EDP Sciences; 2018.
- [21] Deo RC, Sahin M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew Sustain Energy Rev* 2017;72:828–48.
- [22] Deo RC, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 2016;168:568–93.
- [23] Deo RC, Sahin M. Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydro-meteorological parameters and climate indices in eastern Australia. *Atmos Res* 2015;161:65–81.
- [24] Acharya N, et al. Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. *Clim Dyn* 2014;43(5–6):1303–10.
- [25] Şahin M. Modelling of air temperature using remote sensing and artificial neural network in Turkey. *Adv Space Res* 2012;50(7):973–85.
- [26] Abbot J, Marohasy J. Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks. *Atmos Res* 2014;138:166–78.
- [27] Hudson D, et al. Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Q J R Meteorol Soc* 2011;137(656):673–89.
- [28] Abbot J, Marohasy J. Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Adv Atmos Sci* 2012;29(4):717–30.
- [29] Ortiz-García E, Salcedo-Sanz S, Casanova-Mateo C. Accurate precipitation prediction with support vector classifiers: a study including novel predictive variables and observational data. *Atmos Res* 2014;139:128–36.
- [30] Hamzaçebi C. Forecasting of Turkey's net electricity energy consumption on sectoral bases. *Energy Policy* 2007;35(3):2009–16.
- [31] Deo RC, Şahin M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew Sustain Energy Rev* 2017;72:828–48.
- [32] Prasad R, et al. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmos Res* 2017;197:42–63.
- [33] Şahin M, Kaya Y, Uyar M. Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data. *Adv Space Res* 2013;51(5):891–904.
- [34] Nastos P, et al. Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece. *Atmos Res* 2014;144:141–50.
- [35] Tiwari MK, Adamowski J. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resour Res* 2013;49(10):6486–507.
- [36] Tiwari MK, Chatterjee C. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J Hydrol* 2010;394(3):458–70.
- [37] Tiwari MK, Chatterjee C. A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. *J Hydroinf* 2011;13(3):500–19.
- [38] Jia Y, Culver TB. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *J Hydrol* 2006;331(3):580–90.
- [39] Efron B. Bootstrap methods: another look at the jackknife. *Breakthroughs in statistics*. Springer; 1992. p. 569–93.
- [40] Efron B, Tibshirani RJ. An introduction to the bootstrap. *CRC press*; 1994.
- [41] Friedman JH. Multivariate adaptive regression splines. *The annals of statistics*; 1991. p. 1–67.
- [42] Box GE, Jenkins GM. *Time series analysis: forecasting and control*. revised ed Holden-Day; 1976.
- [43] Deo RC, Kisi O, Singh VP. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmos Res* 2017;184:149–75.
- [44] Deo R, et al. Very short-term reactive forecasting of the solar ultraviolet index using an extreme learning machine integrated with the solar zenith angle. *Environ Res* 2017;155:141.
- [45] Erdogdu E. Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey. *Energy Policy* 2007;35(2):1129–46.
- [46] McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943;5(4):115–33.
- [47] Huang H-Y. Unified approach to quadratically convergent algorithms for function minimization. *J Optim Theory Appl* 1970;5(6):405–23.
- [48] Dennis Jr. JE, Schnabel RB. *Numerical methods for unconstrained optimization and nonlinear equations* vol. 16. Siam; 1996.
- [49] Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 1963;11(2):431–41.
- [50] HariKumar R, Vasanthi N, Balasubramani M. Performance analysis of artificial neural networks and statistical methods in classification of oral and breast cancer stages. *Int J Soft Comput Eng* 2009;ume. 2.
- [51] Draper NR, Smith H. *Applied regression analysis* vol. 326. John Wiley & Sons; 2014.
- [52] Montgomery DC, Peck EA, Vining GG. *Introduction to linear regression analysis* vol. 821. John Wiley & Sons; 2012.
- [53] Civelekoglu G, et al. Prediction of bromate formation using multi-linear regression and artificial neural networks. *Ozone: Sci Eng* 2007;29(5):353–62.
- [54] Twomey JM, Smith AE. Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. *IEEE Trans Syst Man Cybern C Appl Rev* 1998;28(3):417–30.
- [55] Energex. Zone substation load data request form. <https://www.energex.com.au/about-us/contact-us/forms/general-forms/zone-substation-load-data-request-form>; 2018.
- [56] Jeffrey SJ, et al. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ Model Softw* 2001;16(4):309–30.
- [57] Ghimire S, et al. Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities. *Remote Sens Environ* 2018;212:176–98.
- [58] Hsu C-W, Chang C-C, Lin C-J. A practical guide to support vector classification. 2003.
- [59] Friedman JH. Estimating functions of mixed ordinal and categorical variables using adaptive splines. *Stanford Univ CA Lab for Computational Statistics*; 1991.
- [60] Jekabsons G. Adaptive regression splines toolbox for matlab/octave. Version 2013;1:72.
- [61] Hyndman R, Khandakar Y. Automatic time series forecasting: the forecast package for R 7. 2008. 2007 <https://www.jstatsoft.org/article/view/v027i03>.
- [62] Wang X, Smith K, Hyndman R. Characteristic-based clustering for time series data. *Data Min Knowl Discov* 2006;13(3):335–64.
- [63] Hu S. Akaike information criterion. *Center for Research in Scientific Computation*; 2007.
- [64] Barzegar R, et al. Mapping groundwater contamination risk of multiple aquifers using multi-model ensemble of machine learning algorithms. *Sci Total Environ* 2018;621:697–712.
- [65] Prasad R, et al. Ensemble committee-based data intelligent approach for generating soil moisture forecasts with multivariate hydro-meteorological predictors. *Soil Tillage Res* 2018;181:63–81.
- [66] Mohammadi K, et al. Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Convers Manag* 2015;91:433–41.
- [67] Willmott CJ, Robeson SM, Matsuura K. A refined index of model performance. *Int J Climatol* 2012;32(13):2088–94.
- [68] Willmott CJ. On the evaluation of model performance in physical geography. *Spatial statistics and models*. Springer; 1984. p. 443–60.
- [69] Willmott CJ. Some comments on the evaluation of model performance. *Bull Am Meteorol Soc* 1982;63(11):1309–13.
- [70] Willmott CJ. On the validation of models. *Phys Geogr* 1981;2(2):184–94.
- [71] Dawson CW, Abraham RJ, See LM. HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ Model Softw* 2007;22(7):1034–52.
- [72] Legates DR, McCabe GJ. Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 1999;35(1):233–41.
- [73] Krause P, Boyle D, Båse F. Comparison of different efficiency criteria for hydrological model assessment. *Adv Geosci* 2005;5:89–97.
- [74] Chai T, Draxler RR. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci Model Dev (GMD)* 2014;7(3):1247–50.
- [75] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models part I—a discussion of principles. *J Hydrol* 1970;10(3):282–90.
- [76] Li M-F, et al. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Convers Manag* 2013;70:139–48.
- [77] Mohammadi K, et al. A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers Manag* 2015;92:162–71.
- [78] Heinemann AB, et al. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia* 2012;71(4):572–82.
- [79] Campbell A. Price and income elasticities of electricity demand: evidence from Jamaica. *Energy Econ* 2018;69:19–32.
- [80] Akay D, Atak M. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. *Energy* 2007;32(9):1670–5.
- [81] Wan Z. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD). Santa Barbara: Institute for Computational Earth System Science; 1999. p. 75.
- [82] Wan Z, et al. Quality assessment and validation of the MODIS global land surface

- temperature. *Int J Remote Sens* 2004;25(1):261–74.
- [83] MODIS. Moderate-resolution imaging spectroradiometer). https://modis.gsfc.nasa.gov/about/media/modis_brochure.pdf; 2018. *MODIS*.
- [84] Deo RC, et al. Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: a new approach. *Renew Sustain Energy Rev* 2019;104:235–61.
- [85] Ghimire S, et al. Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction. *Energies* 2019;12(12):2407.
- [86] Statistics ABO. Census. 2018<http://www.abs.gov.au/>.
- [87] Galelli S, Castelletti A. Tree-based iterative input variable selection for hydrological modeling. *Water Resour Res* 2013;49(7):4295–310.
- [88] Cornejo-Bueno L, et al. Significant wave height and energy flux prediction for marine energy applications: a grouping genetic algorithm–Extreme Learning Machine approach. *Renew Energy* 2016;97:380–9.
- [89] Salcedo-Sanz S, et al. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–Extreme learning machine approach. *Energy Convers Manag* 2014;87:10–8.
- [90] Salcedo-Sanz S, et al. Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization–Extreme Learning Machine approach. *Sol Energy* 2014;105:91–8.
- [91] Ghimire S, et al. Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. *Renew Sustain Energy Rev* 2019;113:109247.
- [92] Quilty J, Adamowski J. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *J Hydrol* 563, 2018, 336-353.
- [93] Slougher JM, Gneiting T, Raftery AE. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J Am Stat Assoc* 2010;105(489):25–35.
- [94] Wang D, et al. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl Energy* 2017;190:390–407.

Chapter 6: Using a Suitable Wavelet Transformation Tool to Achieve Better Electricity Demand Forecasting

6.1 Foreword

This chapter presents an exact copy of the published article in the journal of *Energies* (Vol. 13, Page 2307).

As outlined in Chapter 5 that many current energy-forecasting studies have incorrectly applied wavelet transformation (WT) in which falsely results have been generated, this chapter addresses this issue by applying a suitable and correct approach of maximum overlap discrete wavelet transformation (MODWT) to improve the forecast accuracy. To achieve this accuracy, the partial autocorrelation function (PACF) is applied firstly on the daily electricity demand data obtained for three regional campuses of the University of Southern Queensland, Australia to select the significant input variables. Secondly, the MODWT technique is then used with several filters and levels of decomposition to decompose each variable into wavelet and scaling coefficients before running the model. Finally, the online sequential extreme learning machine (OS-ELM) is developed with a variety of hidden neuron sizes constructing the MODWT-PACF-OS-ELM (MPOE) model. A large number of models developed in this chapter and the accuracy of the MPOE model is compared with its traditional model of PACF-OS-ELM (POE) developed without wavelet to show the advantages of the MODWT procedure. The main contribution of this chapter is to demonstrate the ability of the correct method of MODWT to enhance the accuracy of electricity demand (G) forecasting.

6.2 Research Highlights

- Applying a partial autocorrelation function (PACF) to select the model significant inputs.
- Using a suitable wavelet transformation (WT) algorithm to address data non-stationary issue.
- An online sequential-extreme learning machines (OS-ELM) model integrated with a maximum overlap discrete wavelet transform (MODWT) algorithm.

- The MODWT-PACF-OS-ELM (MPOE) model was evaluated against the PACF-OS-ELM (POE) model.
- The MODWT-PACF-OS-ELM (MPOE) model outperformed the benchmark model at daily forecasting horizon in all three sites of university campuses.
- Energy security studies and national electricity markets should explore the capability of the MODWT-PACF-OS-ELM (MPOE) model for demand forecasting.

6.3 Published Article IV

Article

Electrical Energy Demand Forecasting Model Development and Evaluation with Maximum Overlap Discrete Wavelet Transform-Online Sequential Extreme Learning Machines Algorithms

Mohanad S. Al-Musaylh ^{1,2} , Ravinesh C. Deo ^{1,*}  and Yan Li ¹ 

¹ School of Sciences, Institute of Life Sciences and the Environment, University of Southern Queensland, Toowoomba, QLD 4350, Australia; MohanadShakirKhalid.Al-Musaylh@usq.edu.au or mohanadk21@gmail.com (M.S.A.-M.); Yan.Li@usq.edu.au (Y.L.)

² Management Technical College, Southern Technical University, Basrah 61001, Iraq

* Correspondence: ravinesh.deo@usq.edu.au

Received: 7 March 2020; Accepted: 2 May 2020; Published: 6 May 2020



Abstract: To support regional electricity markets, accurate and reliable energy demand (G) forecast models are vital stratagems for stakeholders in this sector. An online sequential extreme learning machine (OS-ELM) model integrated with a maximum overlap discrete wavelet transform (MODWT) algorithm was developed using daily G data obtained from three regional campuses (i.e., Toowoomba, Ipswich, and Springfield) at the University of Southern Queensland, Australia. In training the objective and benchmark models, the partial autocorrelation function (PACF) was first employed to select the most significant lagged input variables that captured historical fluctuations in the G time-series data. To address the challenges of non-stationarities associated with the model development datasets, a MODWT technique was adopted to decompose the potential model inputs into their wavelet and scaling coefficients before executing the OS-ELM model. The MODWT-PACF-OS-ELM (MPOE) performance was tested and compared with the non-wavelet equivalent based on the PACF-OS-ELM (POE) model using a range of statistical metrics, including, but not limited to, the mean absolute percentage error ($MAPE\%$). For all of the three datasets, a significantly greater accuracy was achieved with the MPOE model relative to the POE model resulting in an $MAPE = 4.31\%$ vs. $MAPE = 11.31\%$, respectively, for the case of the Toowoomba dataset, and a similarly high performance for the other two campuses. Therefore, considering the high efficacy of the proposed methodology, the study claims that the OS-ELM model performance can be improved quite significantly by integrating the model with the MODWT algorithm.

Keywords: energy security; time-series forecasting; predictive model for electricity demand; OS-ELM; wavelet transformation; MODWT; sustainable energy management systems

1. Introduction

To promote the application of appropriate strategic measures and provide accurate scheduling of electrical power in energy security platforms, a forecasting model that can reliably and precisely forecast the electricity energy demand (G), is required. Arising from a shift in consumer energy usage, the large fluctuations evident in G data records leave traditional machine learning models, for example, artificial neural network (ANN) [1], multivariate adaptive regression spline (MARS) [1–3], support vector regression (SVR) [2,3], M5 model tree [3], online sequential extreme learning machine (OS-ELM) [4], and multiple linear regression (MLR) [1] needing to improve their capability to accurately forecast the G data. To achieve this task, a data pre-processing method, to be implemented before running

the model, is required to address this issue if such data are unsteady, stochastic, or chaotic, as found with real life variables.

Wavelet transformation (WT) algorithm, a popular data pre-processing technique that has been widely adopted in the field of energy forecasting (e.g., [5–10]), has been largely explored to decompose the model input datasets through high and low-pass filters. By applying WT, a more coherent structure of the complex time-series can be supplied and fed to a machine learning model to significantly improve the forecast accuracy. Additionally, energy modelers can potentially address issues of non-stationary input data using the WT algorithm, thereby, assisting the model to be more responsive to the input variables' stochastic behaviors [5]. Wavelet transformation can also provide the relevant information regarding the time-series decomposition process, including the provision of patterns of energy usage within the time and frequency domains, thereby increasing a forecasting model's capacity to capture such valuable information at different levels of resolution [7,11]. Because of the detailed information that is produced by WT to convert the data from time domain to frequency domain, a machine learning model can work more intelligently to forecast electricity demand data. The fact that several recent studies have applied the WT algorithm to improve global forecasting accuracies in a range of parameters in several fields, for example, rainfall [11], price of electricity [7,8,10], solar radiation [5,6,9], synthetic hydrological time-series [12], flood levels [13,14], water demand [15], electricity demand [16] and streamflow [17], demonstrates the ability of the WT technique to significantly enhance forecasting accuracy.

Although several research studies (e.g., [5–8,17,18]) have implemented WT for different data forecasting purposes, a recent study in hydrological and water resources forecasting has shown that these studies may have incorrectly applied WT in the data decomposition step. In doing so, they have generated models that should not be employed for real-world forecasting problems because their accuracy is potentially falsely represented [19]. This issue can arise because: (i) future data is drawn upon when the WT uses some data from the testing period to calculate the wavelet and scaling coefficients for the training data, (ii) decomposition levels and wavelet filters are incorrectly selected, and (iii) the training/validation/testing data are split up in an inappropriate manner [19]. It is important to note that these three problems have not been addressed by the current studies in the field of energy forecasting when they apply WT. While some other studies [5,6,17,20,21] have tried to address these issues by applying different forms of WT multiresolution analysis (MRA), for example, discrete wavelet transform (DWT)-MRA or maximal overlap discrete wavelet transform (MODWT)-MRA separately to the training, validation, and testing of data, these approaches require the full time-series to calculate the detail and approximation coefficients, leaving some of the previously mentioned issues unresolved [19]. Therefore, again these studies have failed to apply WT in real-world forecasting problem. Consequently, as only one study has correctly applied MODWT without added MRA to forecast hydrological and water resources, additional studies are needed to explore the impact of MODWT and address the drawbacks cited above when this approach has been used on energy forecasting sector.

In the present study, an OS-ELM model, a fast, reliable and accurate machine learning tool that can offer a better generalization performance than other algorithms in a range of forecasting applications (e.g., regression, classification or time-series) and can learn data one by one as do basic extreme learning machine (ELM) algorithms [22,23], was coupled for the first time with the correct MODWT technique in an effort to forecast G data. Moreover, OS-ELM input and output parameters as well as weights were randomly and analytically selected, respectively [22]. First, the study selected the significant model input variables using a partial autocorrelation function (PACF) to construct a PACF-OS-ELM (POE) model. The novel MODWT-PACF-OS-ELM (MPOE) model was thus built and compared with the standalone (non-wavelet) POE model to investigate the influence of WT on G forecasting when the MODWT transform was applied separately to each input variable to generate the wavelet and scaling coefficients used in feeding the OS-ELM portion of the model. Data from the three regional campuses (Toowoomba, Ipswich, and Springfield) of the University of Southern Queensland (USQ), Australia, were drawn upon to develop and evaluate the accuracy of these techniques in forecasting daily G .

To outline how these goals were achieved, this paper is organized as follows. The theory of the OS-ELM and MODWT algorithms are presented in Section 2. Section 3 describes the study area, data, and methods, while model evaluation criteria, results, and discussions are shown in Section 4. Finally, the study limitations showing future work opportunities, and conclusions are summarized in Sections 5 and 6, respectively.

2. Theoretical Background

2.1. Online Sequential Extreme Learning Machine Model (OS-ELM)

In this paper, the machine learning data intelligent ELM-based model architecture was employed to design a single-layer feed-forward neural network (SLFN), expressed as [4,24]:

$$y_k = \sum_{i=1}^{i=M} \rho_i f(w_i \cdot x_k + c_i) \quad (1)$$

where $k = 1, 2, \dots; N$, M is the hidden nodes of N inputs (lagged variables generated from partial autocorrelation function (PACF) for G data or the decomposition of the lagged data resulting from MODWT) $x_k(t) = \{x_k\}_{k=1}^N \in R^N$, and output (G forecasted) $y_k(x) = \{y_k\}_{k=1}^N \in R$ in the training period). A high-level flowchart is presented in Figure 1 to show how these input variables were used to feed the OS-ELM model. $f(\cdot)$ is the activation function, $c_i \in \Gamma$ denotes the threshold of the i th hidden node, the weight vectors that connect the i th hidden node with the input and output nodes are $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$ and $\rho_i = [\rho_{i1}, \rho_{i2}, \dots, \rho_{im}]^T$, respectively, and the term $w_i \cdot x_k$ refers to the inner product of w_i and x_k .

According to [24], Equation (1) can be simplified to the form below:

$$H\rho = Y \quad (2)$$

where $H = \begin{bmatrix} f(w_1 \cdot x_1 + c_1) & \cdots & f(w_M \cdot x_1 + c_M) \\ \vdots & & \\ f(w_1 \cdot x_N + c_1) & \cdots & f(w_M \cdot x_N + c_M) \end{bmatrix}_{N \times M}$ is the hidden layer output matrix of the neural network, $\rho = \begin{bmatrix} \rho_1^T \\ \vdots \\ \rho_M^T \end{bmatrix}_{M \times m}$ and $Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times m}$.

The following output weight is calculated by applying the least square solution of the linear systems as follows:

$$\rho = H^*Y \quad (3)$$

where H^* denotes the inverse matrix of H .

With the classical ELM model, all N samples of data are used during the learning process, making the model relatively time consuming [4]; however, the OS-ELM model, such as the one developed in the present study, addresses this issue: data are only used once within the two learning stages of initialization and sequential learning [4]. The hidden layer output matrix is designed in the initialization step by allocating the input node (w_i) and the threshold (c_i) to a small piece of initial training data, while the second step of sequential learning is then launched on a one-by-one basis to stop reusing training data [4,22,23,25].

$$V_{j,i} = \sum_{l=0}^{L_j-1} g_{j,l} X_{i-1 \bmod N} \quad (5)$$

where X is an input time-series vector with N values; $j = 1, 2, \dots, J$, where J is the decomposition level at the time i ; $h_{j,l}$ and $g_{j,l}$ are the j th level wavelet ($W_{j,i}$) and scaling ($V_{j,i}$) filters of MODWT, respectively, and L_j denotes the width of the j th level filters.

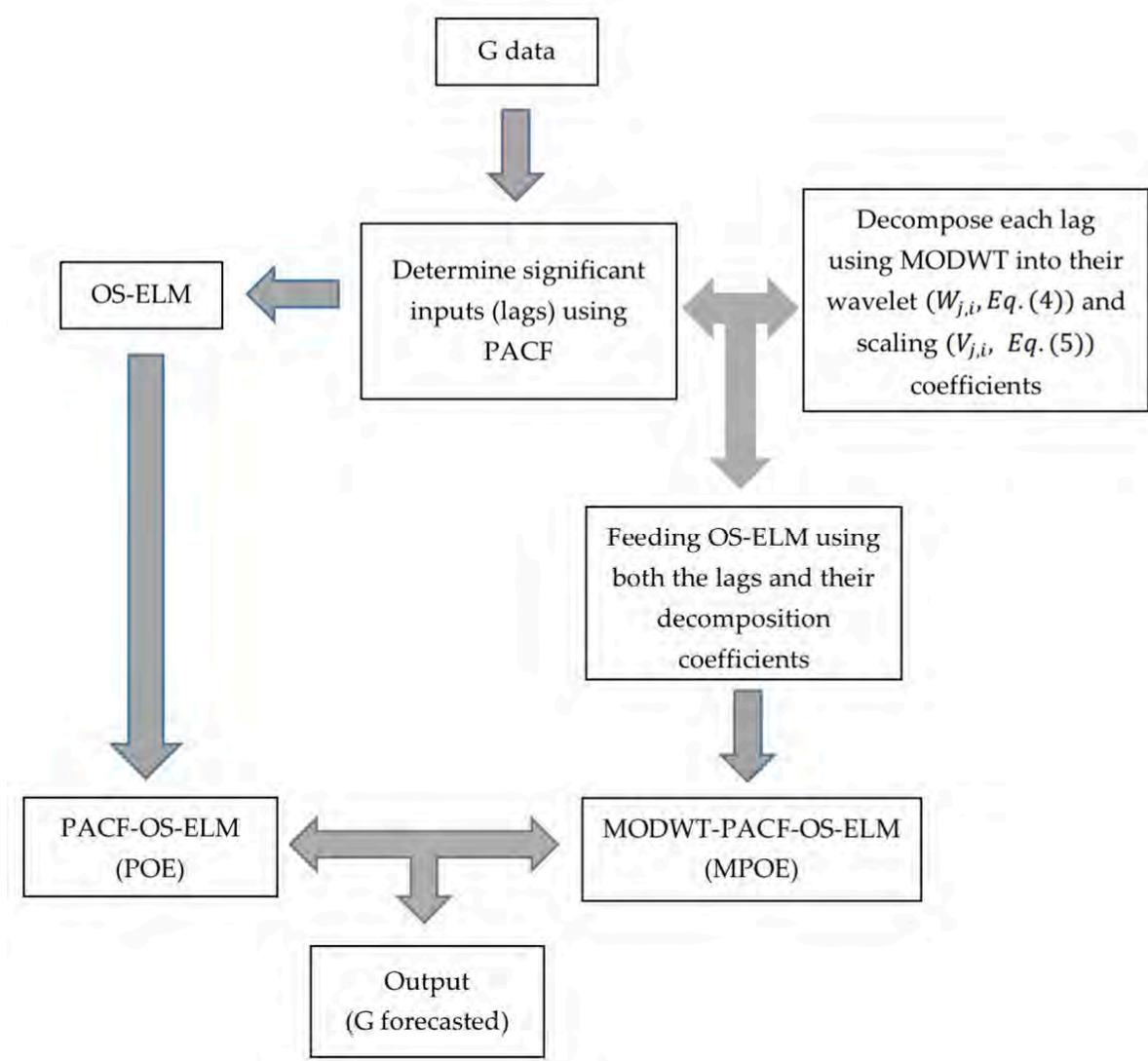


Figure 1: Flowchart showing the model development stages.

2.2. Maximum Overlap Discrete Wavelet Transform (MODWT)

Serving as a pre-processing method, the MODWT algorithm [26] was implemented before running the model to address non-stationarity issues in time-series datasets by decomposing the input data into high and low-pass filters, resulting in MODWT wavelet and scaling coefficients, respectively (Figure 1). Basically, those components are defined as follows [6,19,26]:

$$W_{j,i} = \sum_{l=0}^{L_j-1} h_{j,l} X_{i-1 \bmod N} \quad (4)$$

$$V_{j,i} = \sum_{l=0}^{L_j-1} g_{j,l} X_{i-1 \bmod N} \quad (5)$$

where X is an input time-series vector with N values; $j = 1, 2, \dots, J$, where J is the decomposition level at the time i ; $h_{j,l}$ and $g_{j,l}$ are the j th level wavelet ($W_{j,i}$) and scaling ($V_{j,i}$) filters of MODWT, respectively, and L_j denotes the width of the j th level filters.

3. Data and Methods in the Training Period

3. Data and Methods in the Training Period

3.1. Study Area and Data

3.1. Study Area and Data

In the present case study, the ability of the MPOE model to forecast daily electricity demand (G) was tested using electricity use data from three regional university campuses (Toowoomba, Ipswich, and Springfield) of the University of Southern Queensland (USQ), Australia. The historical data were provided by the university campus services for the periods of 1 January 2013 to 31 December 2014 for the main feed of Toowoomba, 1 September 2015 to 31 August 2016 for the main feed and Building A block of Ipswich and Springfield campuses, respectively. Data were recorded every 15-min (96 times per day) in kilowatts (kW), with a total of 70,080 values including 60 zeros for Toowoomba, 35,136 points each for Ipswich and Springfield including 30 zeros and non-zeros, respectively. Zeros were filled in by taking the average values for the points at the same time of day, across the previous month. Daily data were then obtained by summing each set of the 96 values, resulting in 730 points (days) for Toowoomba and 366 points (days) each for Ipswich and Springfield. Descriptive statistics for the daily time-series datasets are given in Table 1, while plots of the series datasets for the three university campuses are shown in Figure 2 to demonstrate the electricity demand values recorded for each day. The current G data clearly showed large fluctuations in G values, resulting in the need to implement wavelet transformation through MODWT to address non-stationary issues.

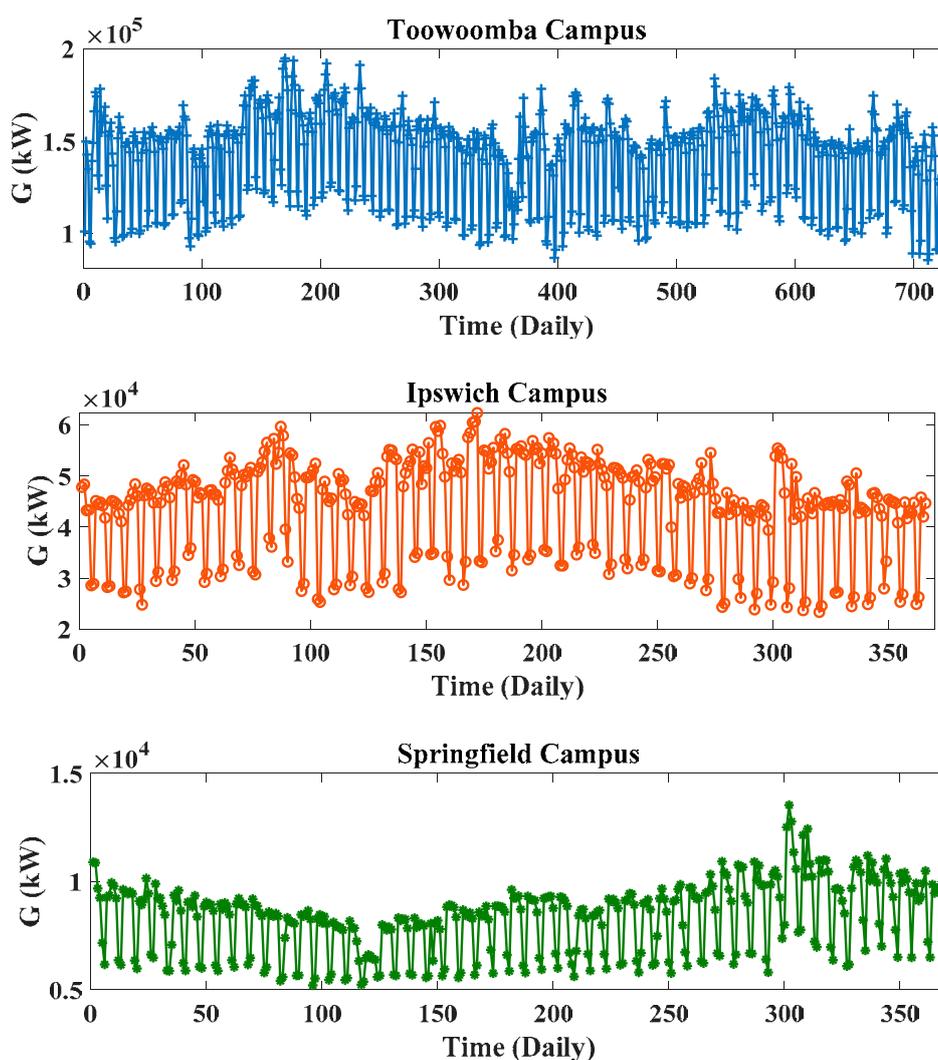


Figure 2. Daily electricity demand (G , kW) time series for the three study sites of Toowoomba, Ipswich and Springfield campuses.

Table 1. Data splitting and descriptive statistics for the G data from the University of Southern Queensland campuses' three stations.

Station	Data Period (dd-mm-yyyy)	Original 15-Mins Data		No. Daily Data Points			Descriptive Statistics for the Whole Daily Datasets			
		Total	No. Zeros	Total	Training (70%)	Validation (15%)	Testing (15%)	Minimum (kW)	Maximum (kW)	Mean (kW)
Toowoomba campus (Main feed)	01-01-2013 to 31-12-2014	70,080	60	730	512	109	109	81,579.70	195,037	141,328
Ipswich campus (Main feed)	01-09-2015 to 31-08-2016	35,136	30	366	256	55	55	23,336	62,378.06	43,716.96
Springfield campus (A Block)		35,136	0							

3.2. Forecast Model Development and Validation

In this study, the proposed MPOE model and its traditional non-WT equivalent POE were developed under the MATLAB environment running on an Intel i7 processor at 3.60 GHz. The (original (non-wavelet) dataset with its statistically significant lagged variables, identified using the partial autocorrelation function (PACF) operating in a 95% confidence interval, was used as an input to develop the classical POE model. Figure 3 illustrates the number of those lags used to build the model where the first two significant lags were selected using the data from all three study sites.

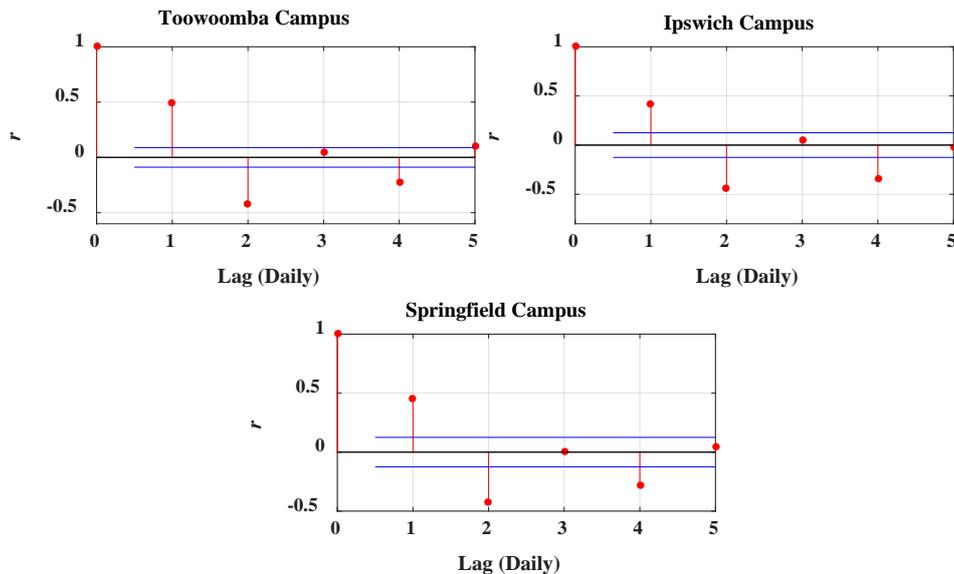


Figure 3. The model input variables constructed from the statistically significant lags at a 95% confidence interval from the original daily G data sets in the training period for the three study sites based on correlation coefficient (r) of predictors (lags) using the partial autocorrelation function (PACF).

On the other hand, to construct the MPOE model, wavelet transformation through MODWT was employed on the individual PACF lagged variables and the wavelet outputs (wavelet and scaling coefficients) were used along with the PACF lagged components as model inputs. The critical task in achieving a robust model with wavelet transformation is to identify the type of wavelet scaling (achieving a robust model with wavelet transformation is to identify the type of wavelet scaling and the level of decomposition can be confirmed in the literature [6,27], a trial-and-error method was employed in the present case. Defining a total of 30 wavelet filters, four different widely tested wavelet families (e.g., [5,6,12,17,19]) were used: Daubechies ($db_i, i = 1, 2, \dots, 10$), where db_1 is the same as the Haar wavelet (haar), Fejer-Korovkin ($fk_i, i = 4, 6, 8, 14, 18, 22$), Coiflets ($coif_i, i = 1, 2, \dots, 5$), and Symlets ($sym_i, i = 2, 3, \dots, 10$). The maximum level of decomposition (J) was computed using Equation (6) [6,17,28]:

$$J = \text{int}(\log_2 N) \quad (6)$$

where N is the number of daily data points in this work, and $\text{int}()$ is the function that returns the nearest integer. For example, for the Toowoomba campus data, a value of $J = 9$ was computed, so all possible levels of decomposition ($J = 1, 2, \dots, 9$) were tested. More details about the MODWT filter and the decomposition level are shown in Table 2. Figure 4 shows the two MODWT wavelet coefficients (WC1 and WC2) and the MODWT scaling coefficient (SC) using lag 1 data from the Toowoomba campus with the best wavelet filter (fk_8) and decomposition level (2). The results of MODWT (wavelet and scaling coefficients) with db_2 and db_3 were found to be the same as those with sym_2 and sym_3 , for all study datasets, respectively.

While forecasting models must go through training, validation, and testing datasets, there is no single agreed-upon scenario for data splitting [2,3,5]. Accordingly, these data were divided into

Table 2. Optimum model performance and parameters in the training and validation phases based on correlation coefficient (r) and root-mean square error ($RMSE$; kW), for the three stations with the daily forecast horizon. The models in **boldface** are the optimal (best performing) models.

Station	Model	No. Hidden Neurons	No. Wavelet/Scaling Filter	No. Wavelet Level	No. Models Developed	Best Model						
						Training		Validation		Wavelet/Scaling Filter	Wavelet Level	Hidden Neuron Size
						r	$RMSE$ (kW)	r	$RMSE$ (kW)			
Toowoomba campus (Main feed)	POE	100	Non-wavelet model		100	0.70	17715.42	0.74	16284.72	Non-wavelet model	9	
	MPOE	100	30	9	27,000	0.96	7260.42	0.94	8026.74	<i>fk₈</i>	2	90
Ipswich campus (Main feed)	POE	100	Non-wavelet model		100	0.68	6944.43	0.67	7543.30	Non-wavelet model	10	
	MPOE	100	30	8	24,000	0.97	2476.19	0.90	4279.08	<i>db₂/sym₂</i>	3	54
Springfield campus (A Block)	POE	100	Non-wavelet model		100	0.65	1036.28	0.61	1641.08	Non-wavelet model	4	
	MPOE	100	30	8	24,000	0.95	441.64	0.89	1164.65	<i>fk₁₄</i>	5	76

1 and 0, respectively. For all the three sites, MPOE models outperformed POE models. For example, for the Toowoomba campus, the MPOE training/validation model accuracy statistics were $r = 0.96/0.94$ and $RMSE = 7260.42/8026.74$ kW with fk_8 , 2 and 90 as the best wavelet filter, decomposition level, and hidden neuron size, respectively. Comparatively, the POE training/validation model accuracy was poorer: $r = 0.70/0.74$ and $RMSE = 17,715.42/16,284.72$ kW for the best hidden neuron size of 9.

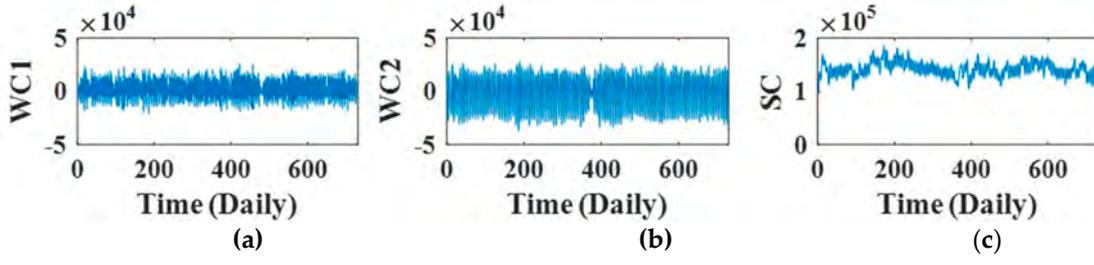


Figure 4 Maximum overlap discrete wavelet transform (MODWT) coefficients constructed using lag of the Toowoomba campus data with the optimum wavelet (scaling) filter and decomposition level. (a) WCI, (b) WC2, and (c) SC represent the wavelet and scaling coefficients, respectively.

While forecasting models must go through training, validation, and testing datasets, there is no single agreed-upon scenario for data splitting [2,3,5]. Accordingly, these data were divided into 70:15:15 for training: validation: testing (Table 1). Data normalization, a very common practice in machine learning, was applied using Equation (7) to scale values down to a range of (0 1), thereby avoiding large numbers in the predictor values of datasets [29]. De-normalization was then applied on predicted data to scale those data back to their original range before models were evaluated.

$$x_{normalized} = \frac{x - x_{minimum}}{x_{maximum} - x_{minimum}} \quad (7)$$

The MATLAB-based OS-ELM function [23], was used to build the present OS-ELM models in this paper. The most important step in developing an OS-ELM model is the selection of the activation function ($f(\cdot)$) and the hidden neuron size (M ; Equation (1)). The radial basis function (RBF) was employed as the activation function in developing the present model, while values of hidden neuron size from 1 to 100 were tested, resulting in 100 POE models for each of the three stations. Additionally, many MPOE models were developed as a result of the numbers of hidden neuron size (M), wavelet filters and decomposition levels (J); for example $100 (M) \times 30$ (wavelet filters) $\times 9 (J) = 27000$ models for the Toowoomba site alone. Table 2 summarizes the details of model development including those factors tested in the training period.

The model accuracy statistics of Pearson correlation coefficient (r) and root-mean square error ($RMSE$; kW) were used to assess the performance of the POE and MPOE models in the training and validation periods, and thereby to identify the best wavelet and model parameters (Table 2).

$$r = \frac{\sum_{i=1}^{i=n} \left[\left(G_i^{obs} - \overline{G^{obs}} \right) \left(G_i^{for} - \overline{G^{for}} \right) \right]}{\sqrt{\sum_{i=1}^{i=n} \left(G_i^{obs} - \overline{G^{obs}} \right)^2} \cdot \sqrt{\sum_{i=1}^{i=n} \left(G_i^{for} - \overline{G^{for}} \right)^2}} \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} \left(G_i^{for} - G_i^{obs} \right)^2} \quad (9)$$

where n is the total number of values of forecast (or observed) G , G_i^{for} and G_i^{obs} are the i th forecasted and observed values, while $\overline{G^{for}}$ and $\overline{G^{obs}}$ are the means of the forecasted and observed values, respectively, in the training period.

The statistical metrics of r and $RMSE$ indicate the greatest model accuracy when they approach 1 and 0, respectively. For all the three sites, MPOE models outperformed POE models. For example, for the Toowoomba campus, the MPOE training/validation model accuracy statistics were $r = 0.96/0.94$ and $RMSE = 7260.42/8026.74$ kW with fk_8 , 2 and 90 as the best wavelet filter, decomposition level, and hidden neuron size, respectively. Comparatively, the POE training/validation model accuracy was poorer: $r = 0.70/0.74$ and $RMSE = 17,715.42/16,284.72$ kW for the best hidden neuron size of 9.

4. Model Evaluation and Results in the Testing Period

4.1. Model Prediction Quality

As the quality of model forecasts of G data cannot be established by a single statistical metric for the testing phase [30], additional measures, besides the $RMSE$ (Equation (9)), were used [30–39]. These included the Mean absolute error (MAE), relative root-mean square error ($RRMSE\%$), and relative mean absolute error ($MAPE\%$),

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |G_i^{for} - G_i^{obs}| \quad (10)$$

$$RRMSE = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}}{\overline{G^{obs}}} \quad (11)$$

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^{i=n} \left| \frac{G_i^{for} - G_i^{obs}}{G_i^{obs}} \right| \quad (12)$$

This MAE shows the model approaching perfection as its value approaches 0. The $RRMSE$ and $MAPE$, both best when approaching 0, present an assessment of model accuracy relative to the range and mean of the forecasted parameter, when a clear evaluation cannot be provided by the $RMSE$ or MAE alone [17]. Model performance is considered to be excellent when $RRMSE < 10\%$, good if $10\% < RRMSE < 20\%$, fair if $20\% < RRMSE < 30\%$, and poor if $RRMSE > 30\%$ [9,31,40,41].

Further model accuracy indexes include the Willmott's Index (WI), Nash–Sutcliffe model efficiency coefficient (E_{NS}) and Legates and McCabe's Index (LM) below where values closest to 1 indicate the best performance.

$$WI = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (|G_i^{for} - \overline{G^{obs}}| + |G_i^{obs} - \overline{G^{obs}}|)^2} \right], \text{ and } 0 \leq WI \leq 1 \quad (13)$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (G_i^{obs} - \overline{G^{obs}})^2} \right], \text{ and } \infty \leq E_{NS} \leq 1 \quad (14)$$

$$LM = 1 - \left[\frac{\sum_{i=1}^{i=n} |G_i^{obs} - G_i^{for}|}{\sum_{i=1}^{i=n} |G_i^{obs} - \overline{G^{obs}}|} \right], \text{ and } (\infty \leq E_{LM} \leq 1) \quad (15)$$

Two statistical tests were also used to show that the MPOE model performs better than the POE model. Those were Wilcoxon Signed-Rank test [42–44] and T test [2]. They were performed with 0.05 significance level and 2-tailed hypothesis.

4.2. Results and Discussion

Using a plethora of model accuracy statistics (i.e., Equations (9)–(15)), the capability and accuracy of the MPOE model in forecasting daily electricity demand (G) was evaluated and compared to that of a traditional POE model, both drawing on testing datasets obtained from three USQ campuses (Toowoomba, Ipswich, and Springfield). While several models were developed and evaluated in this study, only results from optimum models, selected from these several trained models, are shown in Table 3. For all three stations' datasets, the MPOE model showed close to 50% lower values of $RMSE$,

MAE, RRMSE and MAPE and near 50% greater values of E_{NS} and LM than those of the POE model. For example, for the Toowoomba campus dataset, the MPOE model ($MAPE = 4.31\%$, $LM = 0.74$) clearly outperformed the POE model ($MAPE = 11.31\%$, $LM = 0.39$). Moreover, the MPOE model yielded better WI values (0.98, 0.98, and 0.95) than the POE model (0.76, 0.75, and 0.67) for the Toowoomba, Ipswich, and Springfield study areas, respectively. This comparison (Table 3) demonstrated the MPOE model to have yielded a better performance than the non-wavelet POE model.

Table 3. Optimum model performance in the testing phase for daily forecast horizon based on Willmott's Index (WI), Nash–Sutcliffe model efficiency coefficient (E_{NS}), root-mean square error ($RMSE$; kW), mean absolute error (MAE ; kW), mean absolute percentage error ($MAPE\%$), relative root-mean square error ($RRMSE\%$), as well as Legates and McCabes Index (LM) for the three stations. The models in **boldface** are the optimal (best performing) models.

Station	Model	WI	E_{NS}	$RMSE$ (kW)	MAE (kW)	$MAPE$ (%)	$RRMSE$ (%)	LM
Toowoomba campus (Main feed)	POE	0.76	0.42	18,030.44	12,812.32	11.31	13.58	0.39
	MPOE	0.98	0.91	7267.62	5400.99	4.31	5.47	0.74
Ipswich campus (Main feed)	POE	0.75	0.23	7564.84	4860.76	16.29	19.29	0.36
	MPOE	0.98	0.93	2337.80	1980.87	5.46	5.96	0.74
Springfield campus (A Block)	POE	0.67	-0.10	1612.92	1142.36	12.49	17.43	0.11
	MPOE	0.95	0.80	692.78	540.30	5.84	7.49	0.58

Additionally, to ensure the superiority of the proposed approach and support the results introduced in Table 3, Wilcoxon Signed-Rank test and T test have been presented in Table 4 for the forecasted error statistic $|FE| = |G_i^{for} - G_i^{obs}|$ generated by the MPOE model against the $|FE|$ generated by the POE model. With 0.05 significance level and 2-tailed hypothesis, significant results were shown in both tests (p value < 0.05). These results clearly indicate that the MOPE model receives the significance than the POE model.

Table 4. Wilcoxon Signed-Rank test and T test results for the $|FE|$ of the MOPE model vs. the $|FE|$ for the POE model.

Station	Wilcoxon Signed-Rank Test	T Test
	p Value	p Value
Toowoomba campus (Main feed)	0.00001	0.00001
Ipswich campus (Main feed)	0.00076	0.00053
Springfield campus (A Block)	0.00018	0.00043

To further examine the success of the MPOE model over the POE model for G forecasting in the testing period, observed and forecasted values were plotted as ordinate and abscissa for each model (Figures 5 and 6), and models' absolute forecasted errors $|FE|$ (Figure 7). Time-series plots in Figure 5 show that the wavelet models have achieved greater accuracy (closer to observed) than the wavelet-free models for all sites.

Scatterplots (Figure 6) show the coefficient of determination (R^2) and the linear regression line ($G^{for} = aG^{obs} + b$ where a is the slope and b is the ordinate intercept) between the observed and forecasted values. The greater R^2 and values of a and b closer to 1 and 0, respectively, showed that the MPOE model outperformed the POE model when using each of the campus datasets. For the Ipswich campus dataset, the MPOE model yielded $R^2 = 0.93$, $a = 0.99$ and $b = 993.15$, in contrast to $R^2 = 0.42$, $a = 0.47$ and $b = 24601$ for the POE model.

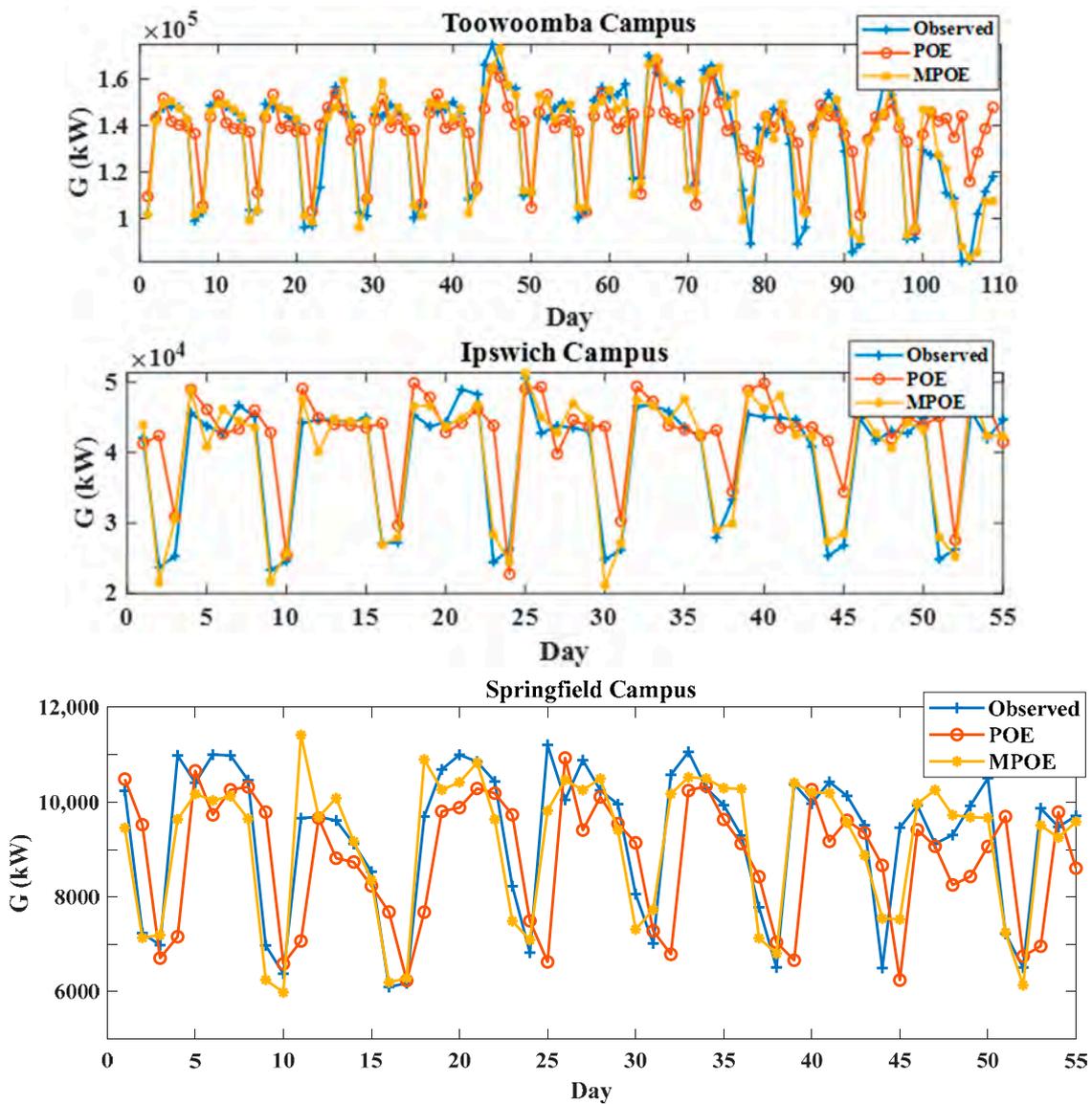


Figure 5. Observed vs. forecasted G data in the testing period with the proposed POE and MPOE models for the study sites.

Using boxplots (Figure 7) and all station datasets, the MPOE and POE models were compared based on their 25%, 50% and 75% quartiles (lower, middle, upper line of box) as well as maximum and minimum values. Of the forecasted and observed values, the mean and standard deviation of $|FE|$ and $|G_i^{for} - G_i^{obs}|$ closer to 1.00 respectively, the statistical error that the MPOE generated further improved the POE model significantly using a range of the comparison datasets. For the Ipswich campus dataset, the MPOE also outperformed the POE model as Figure 5–7.99 and 10.93 indicate that the MPOE model has achieved a better forecasting performance than the POE model by generating lower values from MAE, MAPE%, RMSE, and RRMSE% (Table 3), larger values from WI, ENS and LM (Table 3), significant p values (Table 4), closer forecasted values to the observed values (Figure 5), better values from R^2 , a and b (Figure 6) and lower $|FE|$ (Figure 7). The reason behind this is that the MODWT method has successfully addressed the non-stationarity issues in time-series datasets before running the POE model to enhance the forecasting accuracy.

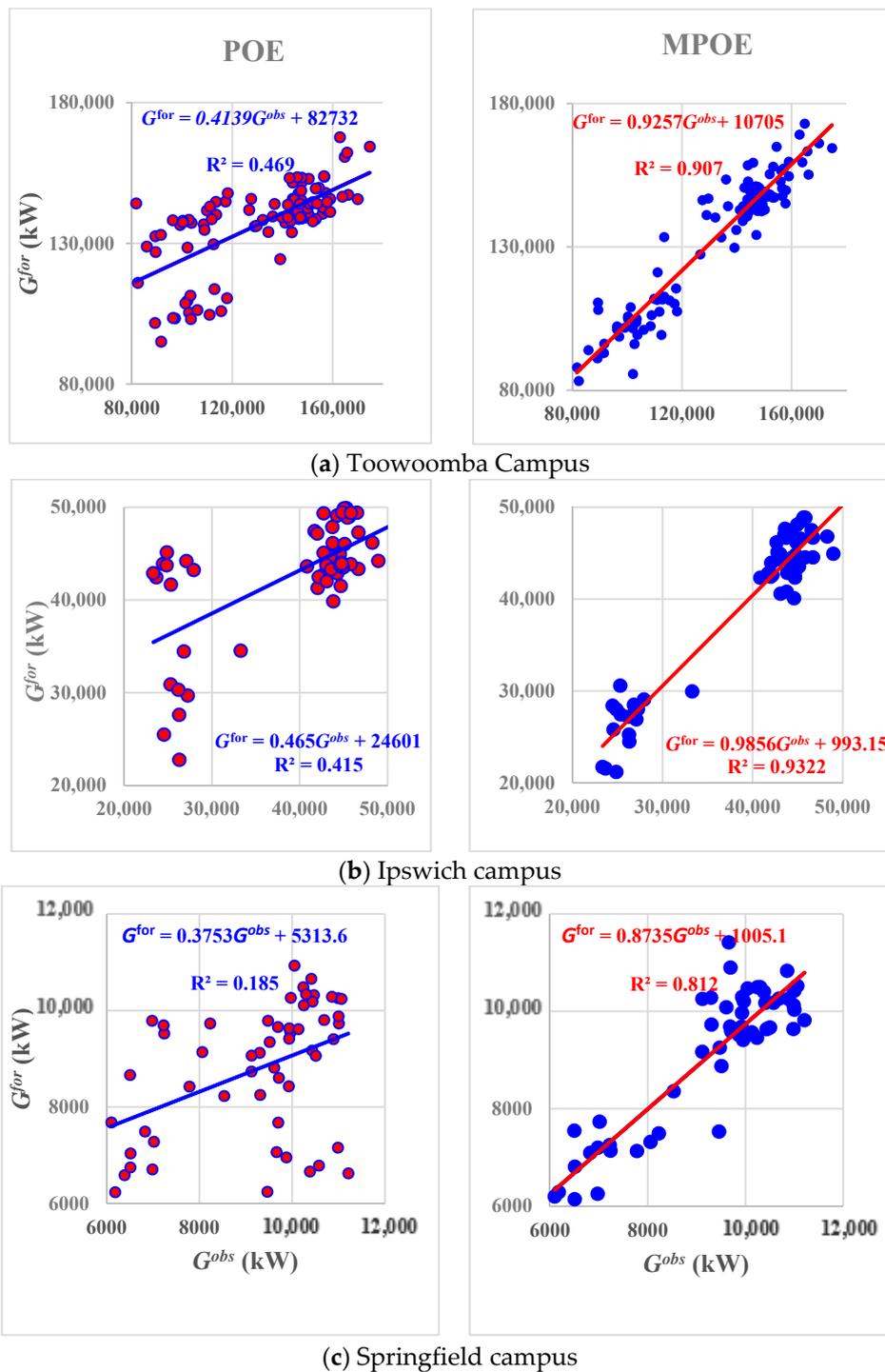


Figure 6. Scatterplots of the observed and forecasted G data in the testing phase with the optimal models of POE and MPOE. Equations of linear regression and the coefficient of determination are incorporated. (a) Toowoomba Campus. (b) Ipswich campus. (c) Springfield campus.

Using boxplots (Figure 7) and all station datasets, the MPOE and POE models were compared based on their 25%, 50% and 75% quartiles (lower, middle, upper line of box) as well as maximum and minimum values of the forecasted error statistic $|FE| = |G_i^{for} - G_i^{obs}|$. Consequently, the statistical error criteria generated for the MPOE model were significantly lower than those for POE.

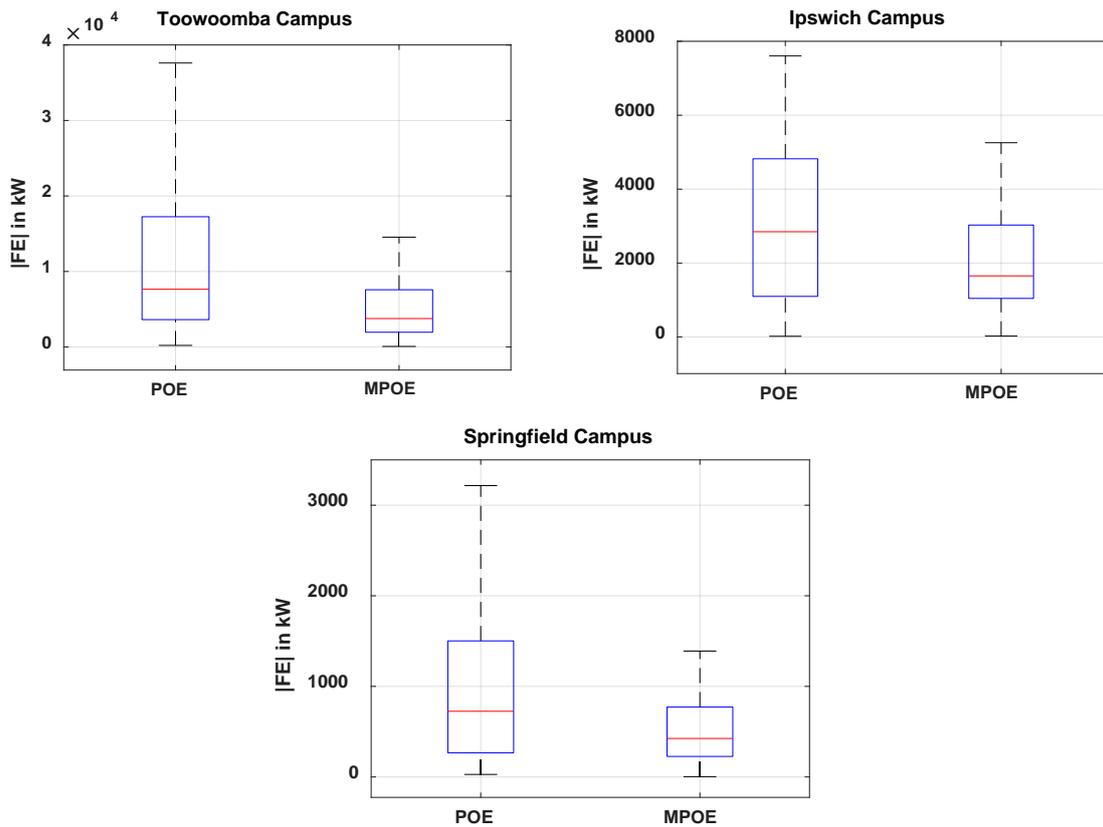


Figure 7. Boxplots of the absolute forecasted error |FE| in the testing dataset for the three study sites with the optimal models of POE and MPOE.

Figure 7. Boxplots of the absolute forecasted error |FE| in the testing dataset for the three study sites with the optimal models of POE and MPOE.

5. Challenges and Future Work

While this study was the first to apply the best suitable wavelet transforms to energy forecasting datasets, thereby achieving a high performance MPOE model, some limitations should be addressed in upcoming works, in particular, the incorporation of external datasets, such as climate variables, which can be downloaded from different sources (e.g., SILO [45], the European Centre for Medium Range Weather Forecasts (ECMWF) and global numerical weather prediction models [1,46] and NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) [47–50]). These can be decomposed together with lag values using MODWT. The OS-ELM model will be then fed by the wavelet and scaling components resulting from the climate and lag variables, to develop a very high-dimensional

5. Challenges and Future Work

model. Although this study has used a good amount of data (730 days and 366 days), a larger power grid model should be built and evaluated using larger datasets from electricity demand (G) to support national electricity markets. This could be achieved by testing the proposed method of this work with a larger study area or incorporating new datasets from the University of Southern Queensland (study area) when these data are available. However, given the large number of input variables that would be generated by MODWT, a method to select and narrow down the best input variables or a very fast model would be necessary to speed up the development step. Accordingly, different pre-processing techniques (e.g., iterative input selection (IIS) [51], grouping genetic algorithm (GGA) [52] or coral reef optimization (CRO) [53,54]), along with a fast forecasting method (e.g., deep learning strategy wavelet and scaling components resulting from the climate and lag variables, to develop a very high-dimensional model. Although this study has used a good amount of data (730 days and 366 days), a larger power grid model should be built and evaluated using larger datasets from electricity demand (G) to support national electricity markets. This could be achieved by testing the proposed method of this work with a larger study area or incorporating new datasets from the University of Southern Queensland (study area) when these data are available. However, given the large number of input variables that would be generated by MODWT, a method to select and narrow down the best input variables or a very fast model would be necessary to speed up the development step. Accordingly,

While this study was the first to apply the best suitable wavelet transforms to energy forecasting datasets, thereby achieving a high performance MPOE model, some limitations should be addressed in upcoming works, in particular, the incorporation of external datasets, such as climate variables, which can be downloaded from different sources (e.g., SILO [45], the European Centre for Medium Range Weather Forecasts (ECMWF) and global numerical weather prediction models [1,46] and NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) [47–50]). These can be decomposed together with lag values using MODWT. The OS-ELM model will be then fed by the wavelet and scaling components resulting from the climate and lag variables, to develop a very high-dimensional model. Although this study has used a good amount of data (730 days and 366 days), a larger power grid model should be built and evaluated using larger datasets from electricity demand (G) to support national electricity markets. This could be achieved by testing the proposed method of this work with a larger study area or incorporating new datasets from the University of Southern Queensland (study area) when these data are available. However, given the large number of input variables that would be generated by MODWT, a method to select and narrow down the best input variables or a very fast model would be necessary to speed up the development step. Accordingly,

6. Conclusions

This study has developed a new energy forecasting model by integrating wavelet transformation based on MODWT with the PACF-OS-ELM model to improve the forecasting accuracy of variables that would be generated by MODWT, a method to select and narrow down the best input variables or a very fast model would be necessary to speed up the development step. Accordingly,

electricity demand (G) data using the datasets from three regional campuses (Toowoomba, Ipswich, and Springfield) from the University of Southern Queensland (USQ). The MPOE model's testing phase accuracy of prediction was then evaluated and compared to that of its classical non-wavelet model equivalent (i.e., POE) using several statistical criteria including correlation coefficient (r), root-mean square error ($RMSE$), mean absolute error (MAE), relative root-mean square error ($RRMSE\%$), and relative mean absolute error ($MAPE\%$), Willmott's Index (WI), Nash–Sutcliffe efficiency coefficient (E_{NS}) and Legates and McCabe's Index (LM) as well as two statistical tests of Wilcoxon Signed-Rank test and T test. The MPOE model outperformed the POE model for all campus datasets.

Although better accuracy was yielded by the MPOE model developed, than the basic POE model, future work is needed to address some limitations associated with the data and methods used in this work. External datasets, such as climate variables and a pre-processing technique used to select the best inputs from those variables, such as IIS, could be employed to further reduce forecasting errors.

To sum up, accurate and reliable G forecasting can be supplied by the MPOE model, and can therefore help regional electricity markets to improve their system by delivering more precise decisions. However, improved model performance can be provided by future works that could address the challenges above.

Author Contributions: Conceptualization, M.S.A.-M.; methodology, M.S.A.-M.; software, M.S.A.-M.; validation, M.S.A.-M.; formal analysis, M.S.A.-M.; investigation, M.S.A.-M.; resources, M.S.A.-M.; data curation, M.S.A.-M.; writing—original draft preparation, writing—review and editing, M.S.A.-M., R.C.D. and Y.L.; visualization, M.S.A.-M.; supervision, R.C.D. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We thank Alicia Logan (Environmental Manager, campus services, University of Southern Queensland, Australia) for providing all the required data for this study. Our thanks also go to the Ministry of Higher Education and Scientific Research in the Government of Iraq for funding the first author's PhD project and the University of Southern Queensland for providing access to various academic services. Special thanks to Professor Jan Adamowski (Department of Bioresource Engineering, Faculty of Agricultural, and Environmental Science, McGill University, Canada) and Assistant Professor John Quilty (Department of Civil and Environmental Engineering, University of Waterloo, Canada) for advice on the wavelet technique. Finally, we also would like to thank Professor Jan Adamowski again and Dr Barbara Harmes (Language Centre, Open Access College, University of Southern Queensland, Australia) for providing help in proof-reading this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Acronyms

c_i	Threshold of i th hidden node
$coif_i$	Coiflets wavelet filter
db_i	Daubechies wavelet filter
$f(\cdot)$	SFLM activation function
fk_i	Fejer–Korovkin wavelet filter
$g_{j,l}$	j th level scaling filter
$h_{j,l}$	j th level wavelet filter
$int(\cdot)$	Nearest integer function
k	Number of hidden nodes in SFLM
kW	Kilowatts
r	Pearson's correlation coefficient
sym_i	Symlets wavelet filter
w_i	weight vectors linking i th hidden node with the input node (SFLM)
ANN	Artificial neural network
CRO	Coral reef optimization
DWT	Discrete wavelet transform
DWT-MRA	Discrete wavelet transform–multiresolution analysis

ECMWF	European Centre for Medium Range Weather Forecasts
ELM	Extreme learning machine
E_{NS}	Nash–Sutcliffe model efficiency coefficient
$ FE $	Absolute Forecasted error statistics
G	Electricity demand (kW)
$\frac{G_i^{for}}{G^{for}}$	i th forecasted value of G (kW) Mean of forecasted G values (kW)
POE	PACF-OS-ELM
R^2	Coefficient of determination
RMSE	Root-mean square error
RRMSE	Relative root-mean square error, %
SC	Scaling coefficient (MODWT)
SFLM	single-layer feed-forward neural network
$\frac{G_i^{obs}}{G^{obs}}$	i th observed value of G (Kw) Mean of observed G values (kW)
GGA	grouping genetic algorithm
H	SFLM's hidden layer output matrix
H^*	Inverse of H matrix
IIS	Iterative input selection
J	Decomposition level
L_j	Width of the j th level filters
LM	Legates and McCabe's Index
M	Hidden neuron size
MAE	Mean absolute error
MAPE	Mean absolute percentage error, %
MARS	Multivariate adaptive regression spline
MLR	Multiple linear regression
MODIS	Moderate Resolution Imaging Spectroradiometer (NASA)
MODWT	Maximum overlap discrete wavelet transform
MODWT-MRA	Maximum overlap discrete wavelet transform–multiresolution analysis
MPOE	MODWT-PACF-OS-ELM
MRA	Multiresolution analysis
N	Number of values in a data series
OS-ELM	Online sequential extreme learning machine
PACF	Partial autocorrelation function
SVR	Support vector regression
$V_{j,i}$	MODWT scaling coefficients
$W_{j,i}$	MODWT wavelet coefficients
WC1, WC2	MODWT wavelet coefficients
WI	Willmott's Index
WT	Wavelet transforms
ρ_i	weight vectors linking i th hidden node with the output node (SFLM)

References

1. Al-Musaylh, M.S.; Deo, R.C.; Adamowski, J.F.; Li, Y. Short-term electricity demand forecasting using machine learning methods enriched with ground-based climate and ECMWF Reanalysis atmospheric predictors in southeast Queensland, Australia. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109–293. [[CrossRef](#)]
2. Al-Musaylh, M.S.; Deo, R.C.; Adamowski, J.; Li, Y. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Adv. Eng. Informatics* **2018**, *35*, 1–16. [[CrossRef](#)]
3. Al-Musaylh, M.S.; Deo, R.C.; Li, Y.; Adamowski, J.F. Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting. *Appl. Energy* **2018**, *217*, 422–439. [[CrossRef](#)]

4. Ali, M.; Deo, R.C.; Downs, N.J.; Maraseni, T. Multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting. *Atmospheric Res.* **2018**, *213*, 450–464. [[CrossRef](#)]
5. Deo, R.C.; Wen, X.; Qi, F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl. Energy* **2016**, *168*, 568–593. [[CrossRef](#)]
6. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. *Renew. Sustain. Energy Rev.* **2019**, *113*, 109247. [[CrossRef](#)]
7. Areekul, P.; Senjyu, T.; Urasaki, N.; Yona, A. Neural-wavelet Approach for Short Term Price Forecasting in Deregulated Power Market. *J. Int. Counc. Electr. Eng.* **2011**, *1*, 331–338. [[CrossRef](#)]
8. Conejo, A.J.; Plazas, M.; Espinola, R.; Molina, A.B. Day-Ahead Electricity Price Forecasting Using the Wavelet Transform and ARIMA Models. *IEEE Trans. Power Syst.* **2005**, *20*, 1035–1042. [[CrossRef](#)]
9. Mohammadi, K.; Shamsirband, S.; Tong, C.W.; Arif, M.; Petković, D.; Chong, W.T. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers. Manag.* **2015**, *92*, 162–171. [[CrossRef](#)]
10. Tan, Z.; Zhang, Y.-J.; Wang, J.; Xu, J. Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Appl. Energy* **2010**, *87*, 3606–3610. [[CrossRef](#)]
11. Feng, Q.; Wen, X.; Li, J. Wavelet Analysis-Support Vector Machine Coupled Models for Monthly Rainfall Forecasting in Arid Regions. *Water Resour. Manag.* **2014**, *29*, 1049–1065. [[CrossRef](#)]
12. Maheswaran, R.; Khosa, R. Comparative study of different wavelets for hydrologic forecasting. *Comput. Geosci.* **2012**, *46*, 284–295. [[CrossRef](#)]
13. Sehgal, V.; Sahay, R.R.; Chatterjee, C. Effect of Utilization of Discrete Wavelet Components on Flood Forecasting Performance of Wavelet Based ANFIS Models. *Water Resour. Manag.* **2014**, *28*, 1733–1749. [[CrossRef](#)]
14. Tiwari, M.K.; Chatterjee, C. Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. *J. Hydrol.* **2010**, *394*, 458–470. [[CrossRef](#)]
15. Tiwari, M.K.; Adamowski, J. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resour. Res.* **2013**, *49*, 6486–6507. [[CrossRef](#)]
16. Zhang, B.-L.; Dong, Z.-Y. An adaptive neural-wavelet model for short term load forecasting. *Electr. Power Syst. Res.* **2001**, *59*, 121–129. [[CrossRef](#)]
17. Prasad, R.; Deo, R.C.; Li, Y.; Maraseni, T. Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmospheric Res.* **2017**, *197*, 42–63. [[CrossRef](#)]
18. Dghais, A.A.A.; Ismail, M.T. A comparative study between discrete wavelet transform and maximal overlap discrete wavelet transform for testing stationarity. *Int. J. Math. Comput. Phys. Electr. Comput. Eng.* **2013**, *7*, 1677–1681.
19. Quilty, J.; Adamowski, J. Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework. *J. Hydrol.* **2018**, *563*, 336–353. [[CrossRef](#)]
20. Barzegar, R.; Moghaddam, A.A.; Adamowski, J.; Ozga-Zielinski, B. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. *Stoch. Environ. Res. Risk Assess.* **2017**, *32*, 799–813. [[CrossRef](#)]
21. Barzegar, R.; Fijani, E.; Moghaddam, A.A.; Tziritis, E. Forecasting of groundwater level fluctuations using ensemble hybrid multi-wavelet neural network-based models. *Sci. Total. Environ.* **2017**, *599*, 20–31. [[CrossRef](#)]
22. Lan, Y.; Soh, Y.C.; Huang, G.-B. Ensemble of online sequential extreme learning machine. *Neurocomputing* **2009**, *72*, 3391–3395. [[CrossRef](#)]
23. Liang, N.-Y.; Huang, G.-B.; Saratchandran, P.; Sundararajan, N. A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. *IEEE Trans. Neural Networks* **2006**, *17*, 1411–1423. [[CrossRef](#)]
24. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]

25. Yadav, B.; Ch, S.; Mathur, S.; Adamowski, J. Discharge forecasting using an Online Sequential Extreme Learning Machine (OS-ELM) model: A case study in Neckar River, Germany. *Measurement* **2016**, *92*, 433–445. [[CrossRef](#)]
26. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: Cambridge, UK, 2000; p. 4.
27. Rathinasamy, M.; Adamowski, J.; Khosa, R. Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method. *J. Hydrol.* **2013**, *507*, 186–200. [[CrossRef](#)]
28. Seo, Y.; Choi, Y.; Choi, J. River Stage Modeling by Combining Maximal Overlap Discrete Wavelet Transform, Support Vector Machines and Genetic Algorithm. *Water* **2017**, *9*, 525. [[CrossRef](#)]
29. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. *A Practical Guide to Support Vector Classification*; Technical Report. Department of Computer Science and Information Engineering, University of National Taiwan: Taipei, Taiwan, 2003; pp. 1–12. Available online: <https://www.csie.ntu.edu.tw/~{cjlin/papers/guide/guide.pdf> (accessed on 15 January 2020).
30. Chai, T.; Draxler, R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]
31. Mohammadi, K.; Shamshirband, S.; Anisi, M.H.; Alam, K.A.; Petković, D. Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Convers. Manag.* **2015**, *91*, 433–441. [[CrossRef](#)]
32. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A refined index of model performance. *Int. J. Clim.* **2011**, *32*, 2088–2094. [[CrossRef](#)]
33. Willmott, C.J. *On the Evaluation of Model Performance in Physical Geography*; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 1984; pp. 443–460.
34. Willmott, C.J. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* **1982**, *63*, 1309–1313. [[CrossRef](#)]
35. Willmott, C.J. On the validation of models. *Phys. Geogr.* **1981**, *2*, 184–194. [[CrossRef](#)]
36. Dawson, C.; Abrahart, R.; See, L. HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ. Model. Softw.* **2007**, *22*, 1034–1052. [[CrossRef](#)]
37. LeGates, D.R.; McCabe, G.J. Evaluating the use of “goodness-of-fit” Measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [[CrossRef](#)]
38. Krause, P.; Boyle, D.P.; Båse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97. [[CrossRef](#)]
39. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290.
40. Li, M.-F.; Tang, X.-P.; Wu, W.; Liu, H. General models for estimating daily global solar radiation for different solar radiation zones in mainland China. *Energy Convers. Manag.* **2013**, *70*, 139–148. [[CrossRef](#)]
41. Heinemann, A.B.; Van Oort, P.; Fernandes, D.S.; Maia, A.D.H.N. Sensitivity of APSIM/ORYZA model due to estimation errors in solar radiation. *Bragantia* **2012**, *71*, 572–582. [[CrossRef](#)]
42. Zhang, Z.; Hong, W.-C.; Li, J. Electric Load Forecasting by Hybrid Self-Recurrent Support Vector Regression Model With Variational Mode Decomposition and Improved Cuckoo Search Algorithm. *IEEE Access* **2020**, *8*, 14642–14658. [[CrossRef](#)]
43. Dong, Y.; Zhang, Z.; Hong, W.-C. A Hybrid Seasonal Mechanism with a Chaotic Cuckoo Search Algorithm with a Support Vector Regression Model for Electric Load Forecasting. *Energies* **2018**, *11*, 1009. [[CrossRef](#)]
44. Hong, W.-C.; Dong, Y.; Lai, C.-Y.; Chen, L.-Y.; Wei, S.-Y. SVR with Hybrid Chaotic Immune Algorithm for Seasonal Load Demand Forecasting. *Energies* **2011**, *4*, 960–977. [[CrossRef](#)]
45. Jeffrey, S.J.; Carter, J.O.; Moodie, K.B.; Beswick, A.R. Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environ. Model. Softw.* **2001**, *16*, 309–330. [[CrossRef](#)]
46. Ghimire, S.; Deo, R.C.; Downs, N.J.; Raj, N. Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities. *Remote. Sens. Environ.* **2018**, *212*, 176–198. [[CrossRef](#)]
47. Deo, R.C.; Şahin, M. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew. Sustain. Energy Rev.* **2017**, *72*, 828–848. [[CrossRef](#)]

48. Wan, Z. MODIS land-surface temperature algorithm theoretical basis document (LST ATBD). Institute for Computational Earth System Science, Santa Barbara. 1999; p. 75. Available online: https://modis.gsfc.nasa.gov/data/atbd/atbd_mod11.pdf (accessed on 10 February 2020).
49. Wan, Z.; Zhang, Y.; Zhang, Q.; Li, Z.-L. Quality assessment and validation of the MODIS global land surface temperature. *Int. J. Remote. Sens.* **2004**, *25*, 261–274. [[CrossRef](#)]
50. MODIS. MODIS (Moderate-Resolution Imaging Spectroradiometer). 2018. Available online: https://modis.gsfc.nasa.gov/about/media/modis_brochure.pdf (accessed on 10 February 2020).
51. Galelli, S.; Castelletti, A. Tree-based iterative input variable selection for hydrological modeling. *Water Resour. Res.* **2013**, *49*, 4295–4310. [[CrossRef](#)]
52. Bueno, L.C.; Nieto-Borge, J.C.; García-Díaz, P.; Rodríguez, G.; Salcedo-Sanz, S. Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm—Extreme Learning Machine approach. *Renew. Energy* **2016**, *97*, 380–389. [[CrossRef](#)]
53. Salcedo-Sanz, S.; Pastor-Sánchez, Á.; Prieto, L.; Blanco-Aguilera, A.; García-Herrera, R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—Extreme learning machine approach. *Energy Convers. Manag.* **2014**, *87*, 10–18. [[CrossRef](#)]
54. Salcedo-Sanz, S.; Casanova-Mateo, C.; Pastor-Sánchez, Á.; Sánchez-Girón, M. Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization—Extreme Learning Machine approach. *Sol. Energy* **2014**, *105*, 91–98. [[CrossRef](#)]
55. Ghimire, S.; Deo, R.C.; Raj, N.; Mi, J. Deep Learning Neural Networks Trained with MODIS Satellite-Derived Predictors for Long-Term Global Solar Radiation Prediction. *Energies* **2019**, *12*, 2407. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Chapter 7: Conclusions and Future Scope

7.1 Conclusions

This study has developed a diverse range of data intelligent modelling techniques for electricity load (G) forecasting within Queensland, Australia. Several standalone, hybrid and high precision models were presented at multi-forecasting horizons including the short-term (half-hourly, hourly, 6-hourly, weekend, working days, public holidays and daily), and long-term (monthly). The adopted data intelligent procedures derived mainly from machine learning techniques, including data decomposition and model parameter optimisations selected in this work were: multivariate adaptive regression splines (MARS), support vector regression (SVR), autoregressive integrated moving average (ARIMA), M5 model tree, artificial neural networks (ANNs), multiple linear regression (MLR) and online sequential extreme learning machine (OS-ELM) models, improved versions of empirical mode decomposition with adaptive noise (ICEEMDAN) and maximum overlap discrete wavelet transform (MODWT), and grid search and hybrid particle swarm optimization (PSO), respectively.

In particular, the MARS, SVR with a grid search and ARIMA models were developed and evaluated in the first objective presented in Chapter 3 for short-term (i.e., half-hourly, hourly and daily) G forecasting using Queensland's area aggregated data obtained from the Australian Energy Market Operator. The study found that the MARS model performed well, outperforming both the SVR and ARIMA models for half-hourly and hourly horizons, while the best daily forecasting accuracy was achieved by the SVR model.

In contrast, using the same data in Chapter 3, the forecasting methodology was improved in the second objective (Chapter 4) in which the SVR model was constructed and integrated with a two-phase data intelligent methods of ICEEMDAN and PSO. The ICEEMDAN was applied to address the data non-stationary problem while the PSO was used to select the optimum parameters of the SVR model. Using different forecasting horizons from whole weeks, weekends, working days, public holidays, and monthly, the hybrid ICEEMDAN-PSO-SVR model was then evaluated against five alternative models of ICEEMDAN-MARS, ICEEMDAN- M5 model tree, PSO-SVR, MARS, and M5 model tree resulting in the highest performance generated by the proposed study model.

In addition, objective 3 (Chapter 5) presented a new model designed for G forecasting by studying the influences of external datasets on G prediction. The study used extensive variables from climate and atmospheric parameters. For eight stations located in southeast Queensland, the data were the six variables from Scientific Information for Land Owners (SILO) and 51 Reanalysis variables obtained from the European Centre for Medium-Range Weather Forecast (ECMWF) used to forecast G acquired from Energex. For 6 hours and daily G forecasting horizons, an artificial neural network (ANN) model was developed as the main model while multivariate adaptive regression spline (MARS), multiple linear regression (MLR), and autoregressive integrated moving average (ARIMA) were adopted as the comparison models developed in this research. According to the results, the study found that the ANN model was the most accurate for both forecasting horizons. Moreover, the hybrid ANN was assembled by combining three forecasts of the ANN, MARS, and MLR models to improve the forecasted accuracy and the bootstrapping (B) skill was integrated with the hybrid ANN model to build an ensemble model to explore the prediction errors.

Finally, a suitable and accurate approach of wavelet transformation (WT) in the field of energy forecasting was presented in the first time in objective 4 (Chapter 6) in this thesis. Accordingly, the maximum overlap discrete wavelet transformation (MODWT) technique was employed to decompose the lag variables extracted from the partial autocorrelation function (PACF). The decomposition data were then used to feed the online sequential extreme learning machine (OS-ELM) building the MODWT-PACF-OS-ELM (MPOE) model. Based on the datasets of the three university campuses at the University of Southern Queensland, Australia, the MPOE significantly outperformed the free wavelet model of PACF-OS-ELM (POE) for the daily forecasting horizon.

To sum up, various novel contributions were provided by this study in the development of data-intelligent predictive models for electricity demand forecasting. According to the results, the study found that the performances of the suggested models were relatively better with respect to standalone models. This means innovative new approaches were explored and the main contributions of the research could be summarized as follows:

- 1- The first contribution was to explore forecasting approaches that have never previously been undertaken in the field of G forecasting in Queensland,

Australia. The study firstly used MARS and SVR as the main models to be compared with the methods that were developed previously.

- 2- The second essential contribution was to develop a hybrid two phase model utilizing the ICEEMDAN algorithm to address the non-stationarity issues that were associated with the data and PSO algorithm to select the best parameters for the SVR model.
- 3- Another significant contribution, which has not previously been explored, to our knowledge, regarding this topic, was incorporating large datasets to forecast G . In this work, extensive sets of climate-based predictors, including both ground-based measurements and atmospheric reanalysis data obtained from the numerical weather forecasting model were used to forecast station-based electricity energy demand using the ANN model.
- 4- The important contribution here was the development of a suitable wavelet transformation that has been incorrectly used in many areas elsewhere. The MODWT technique was applied to improve the forecast accuracy of the OS-ELM model.
- 5- With the notion of real-time forecasting, the forecast error has been reduced gradually indicating the capability of the methods developed in this study.
- 6- Support was provided for the national electricity market in Australia specifically and worldwide possibly by addressing their energy security and sustainability challenges through providing most accurate forecasting tools that have the ability to reduce the forecasted error.

7.2 Limitations and Future Scope

Although this doctoral study has developed and evaluated many different predictive models for G forecasting, which performed quite accurately, the following limitations could be addressed in future research studies:

- 1- Seasonal and yearly forecasting horizons have not been considered in this study due to the limitation of the Queensland's G data that are currently available for 20 years only.
- 2- Different pre-processing techniques, such as iterative input selection (IIS) (Prasad et al. 2017), grouping genetic algorithm (GGA) (Cornejo-Bueno et al. 2016), or coral reef optimisation (CRO) (Salcedo-SanzCasanova-Mateo, et al. 2014; Salcedo-SanzPastor-Sánchez, et al. 2014) could be used to improve the

forecasting accuracy of G data through the selection of the models' best input variables.

- 3- The time consumed to run a model could be reduced using a fast forecasting method, for example, deep learning strategy (Ghimire et al. 2019) or long short term memory network (Zheng et al. 2017).
- 4- The ICEEMDAN technique could be extended into two-layer decomposition scenarios when the variational mode decomposition (VMD) (WangLuo, et al. 2017; WangWei, et al. 2017) is applied to further decompose high frequency generated by ICEEMDAN, into a number of variational modes. The three phase models, such as ICEEMDAN-VMD-PSO-SVR, OR ICEEMDAN-VMD-OS-ELM can then be developed.
- 5- Despite the PSO demonstrating good performance in selecting the parameters of the SVR model, it would be advantageous to include an improved PSO method (Chen, Su & Chen 2012), , multi-swarm PSO (MSPSO) (Liu et al. 2011) or develop a genetic algorithm (Shi et al. 2015) that could help to identify appropriate parameters for the SVR model.
- 6- Incorporating satellite data that can be extracted from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) and utilized as an alternative predictor dataset to develop subsequent G forecasting models. Hourly variables over a $0.05^\circ \times 0.05^\circ$ grid resolution can be generated from this source with two primary sensors (Terra and Aqua MODIS) (Wan 1999; Wan et al. 2004; Deo & Şahin 2017; MODIS 2018).
- 7- Since energy demand is a multivariate problem, social and population variables may need to be taken into account to further improve the accuracy of G forecasts.

In overall conclusion, towards the real-world forecasting problem, this thesis has made novel contributions in the field of energy using high precision electricity demand machine learning techniques. The forecasting errors have been gradually reduced while the data intelligent models have been used in this doctoral research. Consequently, it is important to note that the performance of the models developed in this study has provided significant ability to forecast energy demand in which the models should be explored by forecasters working in national electricity markets, particularly the Australian Energy Market Operator in Australia.

References

Note that the references presented here for Chapters 1, 2 and 7 and do not include the references from the published articles (Chapters 3 to 6). These references are provided in the reference sections of the respective articles.

Abbot, J & Marohasy, J 2012, 'Application of artificial neural networks to rainfall forecasting in Queensland, Australia', *Advances in Atmospheric Sciences*, vol. 29, no. 4, pp. 717-30.

Abbot, J & Marohasy, J 2014, 'Input selection and optimisation for monthly rainfall forecasting in Queensland, Australia, using artificial neural networks', *Atmospheric Research*, vol. 138, pp. 166-78.

Akay, D & Atak, M 2007, 'Grey prediction with rolling mechanism for electricity demand forecasting of Turkey', *Energy*, vol. 32, no. 9, pp. 1670-5.

Ali, M, Deo, RC, Downs, NJ & Maraseni, T 2018, 'Multi-stage hybridized online sequential extreme learning machine integrated with Markov Chain Monte Carlo copula-Bat algorithm for rainfall forecasting', *Atmospheric Research*, vol. 213, pp. 450-64.

Australian Bureau of Statistics, A 2019, *Regional Population Growth, Australia, 2017-18*, Australian Bureau of Statistics, Australia, <https://www.abs.gov.au/AUSSTATS/abs@.nsf/mf/3218.0>>.

Australian Energy Market Operator, A 2016, *Aggregated Price and Demand Data - Historical*, [The Australian Energy Market Operator is responsible for operating Australia's largest gas and electricity markets and power systems], The Australian Energy Market Operator, Australia, <http://www.aemo.com.au/>>.

Australian Government, A 2015, *Australian Energy Update*, Department of Industry and Science, Australia.

Box, GE & Jenkins, GM 1976, *Time series analysis: forecasting and control, revised ed*, Holden-Day.

Bunn, D & Farmer, ED 1985, 'Comparative models for electrical load forecasting'.

Chen, L-F, Su, C-T & Chen, K-H 2012, 'An improved particle swarm optimization for feature selection', *Intelligent Data Analysis*, vol. 16, no. 2, pp. 167-82.

Civelekoglu, G, Yigit, N, Diamadopoulos, E & Kitis, M 2007, 'Prediction of bromate formation using multi-linear regression and artificial neural networks', *Ozone: Science and Engineering*, vol. 29, no. 5, pp. 353-62.

Contreras, J, Espinola, R, Nogales, FJ & Conejo, AJ 2003, 'ARIMA models to predict next-day electricity prices', *IEEE transactions on power systems*, vol. 18, no. 3, pp. 1014-20.

Cornejo-Bueno, L, Nieto-Borge, J, García-Díaz, P, Rodríguez, G & Salcedo-Sanz, S 2016, 'Significant wave height and energy flux prediction for marine energy applications: A grouping genetic algorithm–Extreme Learning Machine approach', *Renewable Energy*, vol. 97, pp. 380-9.

Dee, DP, Uppala, S, Simmons, A, Berrisford, P, Poli, P, Kobayashi, S, Andrae, U, Balmaseda, M, Balsamo, G & Bauer, dP 2011, 'The ERA-Interim reanalysis: Configuration and performance of the data assimilation system', *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553-97.

Deo, RC & Şahin, M 2015, 'Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia', *Atmospheric Research*, vol. 153, pp. 512-25.

Deo, RC & Şahin, M 2017, 'Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland', *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 828-48.

Deo, RC, Wen, X & Qi, F 2016, 'A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset', *Applied Energy*, vol. 168, pp. 568-93.

Deo, RC, Kisi, O & Singh, VP 2017, 'Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model', *Atmospheric Research*, vol. 184, pp. 149-75.

Efron, B 1992, 'Bootstrap methods: another look at the jackknife', in *Breakthroughs in Statistics*, Springer, pp. 569-93.

Efron, B & Tibshirani, RJ 1994, *An introduction to the bootstrap*, CRC press.

Energex 2018, *Zone substation load data request form*, Energex, Australia, <https://www.energex.com.au/about-us/contact-us/forms/general-forms/zone-substation-load-data-request-form>>.

Erdogdu, E 2007, 'Electricity demand analysis using cointegration and ARIMA modelling: A case study of Turkey', *Energy policy*, vol. 35, no. 2, pp. 1129-46.

Fan, S & Chen, L 2006, 'Short-term load forecasting based on an adaptive hybrid method', *IEEE transactions on power systems*, vol. 21, no. 1, pp. 392-401.

Florens, M, Cairns, IH, Knock, S & Robinson, P 2007, 'Data-driven solar wind model and prediction of type II bursts', *Geophysical research letters*, vol. 34, no. 4.

Friedman, JH 1991, 'Multivariate adaptive regression splines', *The annals of statistics*, pp. 1-67.

Ghimire, S, Deo, RC, Raj, N & Mi, J 2019, 'Deep learning neural networks trained with MODIS satellite-derived predictors for long-term global solar radiation prediction', *Energies*, vol. 12, no. 12, p. 2407.

Govindaraju, RS 2000, 'Artificial neural networks in hydrology. II: hydrologic applications', *Journal of Hydrologic Engineering*, vol. 5, no. 2, pp. 124-37.

Grinham, A, Albert, S, Deering, N, Dunbabin, M, Bastviken, D, Sherman, B, Lovelock, CE & Evans, CD 2018, 'The importance of small artificial water bodies as sources of methane emissions in Queensland, Australia', *Hydrology and Earth System Sciences*, vol. 22, no. 10, pp. 5281-98.

Guo, X-C, Liang, Y-C, Wu, C-G & Wang, H-Y 2006, 'Electric Load Forecasting using SVMs', *2006 International Conference on Machine Learning and Cybernetics*.

Haida, T & Muto, S 1994, 'Regression based peak load forecasting using a transformation technique', *IEEE Transactions on Power Systems*, vol. 9, no. 4, pp. 1788-94.

Hsu, C-W, Chang, C-C & Lin, C-J 2003, 'A practical guide to support vector classification'.

Hu, Z, Bao, Y & Xiong, T 2013, 'Electricity load forecasting using support vector regression with memetic algorithms', *The Scientific World Journal*, vol. 2013.

Jeffrey, SJ, Carter, JO, Moodie, KB & Beswick, AR 2001, 'Using spatial interpolation to construct a comprehensive archive of Australian climate data', *Environmental Modelling & Software*, vol. 16, no. 4, pp. 309-30.

Kaboli, SHA, Selvaraj, J & Rahim, N 2016, 'Long-term electric energy consumption forecasting via artificial cooperative search algorithm', *Energy*, vol. 115, pp. 857-71.

Kaytez, F, Taplamacioglu, MC, Cam, E & Hardalac, F 2015, 'Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines', *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431-8.

Lan, Y, Soh, YC & Huang, G-B 2009, 'Ensemble of online sequential extreme learning machine', *Neurocomputing*, vol. 72, no. 13-15, pp. 3391-5.

Li, X & Li, C 2016, 'Improved CEEMDAN and PSO-SVR Modeling for Near-Infrared Noninvasive Glucose Detection', *Computational and mathematical methods in medicine*, vol. 2016.

Liang, N-Y, Huang, G-B, Saratchandran, P & Sundararajan, N 2006, 'A fast and accurate online sequential learning algorithm for feedforward networks', *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1411-23.

Liu, Y, Wang, G, Chen, H, Dong, H, Zhu, X & Wang, S 2011, 'An improved particle swarm optimization for feature selection', *Journal of Bionic Engineering*, vol. 8, no. 2, pp. 191-200.

MODIS 2018, *MODIS (Moderate-Resolution Imaging Spectroradiometer)*, NASA, In: NASA, editor. In the Internet, https://modis.gsfc.nasa.gov/about/media/modis_brochure.pdf>.

Nasr, G, Badr, E & Younes, M 2002, 'Neural networks in forecasting electrical energy consumption: univariate and multivariate approaches', *International Journal of Energy Research*, vol. 26, no. 1, pp. 67-78.

Pal, M & Deswal, S 2009, 'M5 model tree based modelling of reference evapotranspiration', *Hydrological Processes*, vol. 23, no. 10, pp. 1437-43.

Prasad, R, Deo, RC, Li, Y & Maraseni, T 2017, 'Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm', *Atmospheric Research*, vol. 197, pp. 42-63.

Queensland Government, Q 2019, *Queensland regions*, Queensland Government, Australia, <https://www.statedevelopment.qld.gov.au/regions/queensland.html>>.

Quezada, G, Grozev, G, Seo, S & Wang, C-H 2014, 'The challenge of adapting centralised electricity systems: peak demand and maladaptation in South East Queensland, Australia', *Regional environmental change*, vol. 14, no. 2, pp. 463-73.

Quilty, J & Adamowski, J 2018, 'Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework', *Journal of Hydrology*.

Şahin, M, Kaya, Y & Uyar, M 2013, 'Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data', *Advances in Space Research*, vol. 51, no. 5, pp. 891-904.

Salcedo-Sanz, S, Casanova-Mateo, C, Pastor-Sánchez, A & Sánchez-Girón, M 2014, 'Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization–Extreme Learning Machine approach', *Solar energy*, vol. 105, pp. 91-8.

Salcedo-Sanz, S, Pastor-Sánchez, A, Prieto, L, Blanco-Aguilera, A & García-Herrera, R 2014, 'Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization–Extreme learning machine approach', *Energy Conversion and Management*, vol. 87, pp. 10-8.

Shi, X, Huang, Q, Chang, J, Wang, Y, Lei, J & Zhao, J 2015, 'Optimal parameters of the SVM for temperature prediction', *Proceedings of the International Association of Hydrological Sciences*, vol. 368, pp. 162-7.

Suganthi, L & Samuel, AA 2012, 'Energy models for demand forecasting—A review', *Renewable and sustainable energy reviews*, vol. 16, no. 2, pp. 1223-40.

Tiwari, MK & Adamowski, J 2013, 'Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models', *Water Resources Research*, vol. 49, no. 10, pp. 6486-507.

Türkay, BE & Demren, D 2011, 'Electrical load forecasting using support vector machines', *Electrical and Electronics Engineering (ELECO), 2011 7th International Conference on*, IEEE, pp. I-49-I-53.

Vapnik, VN 1998, *Statistical learning theory*, Wiley, New York.

Wan, Z 1999, 'MODIS land-surface temperature algorithm theoretical basis document (LST ATBD)', *Institute for Computational Earth System Science, Santa Barbara*, vol. 75.

Wan, Z, Zhang, Y, Zhang, Q & Li, Z-L 2004, 'Quality assessment and validation of the MODIS global land surface temperature', *International journal of remote sensing*, vol. 25, no. 1, pp. 261-74.

Wang, D, Wei, S, Luo, H, Yue, C & Grunder, O 2017, 'A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine', *Science of The Total Environment*, vol. 580, pp. 719-33.

Wang, D, Luo, H, Grunder, O, Lin, Y & Guo, H 2017, 'Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm', *Applied Energy*, vol. 190, pp. 390-407.

Xydas, E, Marmaras, C, Cipcigan, LM, Jenkins, N, Carroll, S & Barker, M 2016, 'A data-driven approach for characterising the charging demand of electric vehicles: A UK case study', *Applied Energy*, vol. 162, pp. 763-71.

Yu, P-S, Chen, S-T & Chang, I-F 2006, 'Support vector regression for real-time flood stage forecasting', *Journal of Hydrology*, vol. 328, no. 3, pp. 704-16.

Zareipour, H, Bhattacharya, K & Canizares, C 2006, 'Forecasting the hourly Ontario energy price by multivariate adaptive regression splines', *2006 IEEE Power Engineering Society General Meeting*, IEEE, p. 7 pp.

Zheng, J, Xu, C, Zhang, Z & Li, X 2017, 'Electric load forecasting in smart grids using long-short-term-memory based recurrent neural network', *2017 51st Annual Conference on Information Sciences and Systems (CISS)*, IEEE, pp. 1-6.

Appendix A: Conference Paper

This appendix presents an exact copy of the published paper in the *2018 3rd International Conference on Power and Renewable Energy (ICPRE2018*, Vol. 64, p. 08001). The candidate attended this conference in Berlin, Germany on September 21-24, 2018.

Particle Swarm Optimized–Support Vector Regression Hybrid Model for Daily Horizon Electricity Demand Forecasting Using Climate Dataset

Mohanad S AL-Musaylh^{1,2}, Ravinesh C Deo¹, and Yan Li¹

¹School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, QLD 4350, Australia

²Management Technical College, Southern Technical University, Basrah, Iraq

Abstract. This paper has adopted six daily climate variables for the eleven major locations, and heavily populated areas in Queensland, Australia obtained from Scientific Information for Land Owners (SILO) to forecast the daily electricity demand (G) obtained from the Australian Energy Market Operator (AEMO). Optimal data-driven technique based on a support vector regression (SVR) model was applied in this study for the G forecasting, where the model’s parameters were selected using a particle swarm optimization (PSO) algorithm. The performance of PSO–SVR was compared with multivariate adaptive regression spline (MARS) and the traditional model of SVR. The results showed that the PSO–SVR model outperformed MARS and SVR.

1 Introduction

Electricity demand (G) forecasting is a purely fundamental yet a challenging optimisation task for improving business efficiency of the electricity industry. A relationship between the G data and temperature is clearly evident in winter and summer [1]. Hence, it would be significant to develop a forecasting model employing both the G and related climate input datasets.

In recent years, support vector regression (SVR), PSO algorithm, and multivariate adaptive regression splines (MARS) have been widely adopted in energy demand forecasting [1]. Those methods have been used to forecast G in [1, 2], however, the influences of the climate datasets are not incorporated yet.

The main contribution of this research paper is to improve the G forecasting accuracy by involving climate datasets and integrating the merits of the PSO algorithm with the SVR model. To evaluate the PSO-SVR model, the traditional methods of the SVR and MARS algorithms are also developed.

2 Theoretical Background

2.1 Support vector regression

A nonlinear regression problem can be solved by a SVR model, which is a machine learning method and pioneered by [3], below:

$$y = f(X) = \omega \cdot \phi(X) + b \tag{1}$$

where $X = \{x_i\}_{i=1}^{i=n} \in \mathcal{R}^N, y = \{y_i\}_{i=1}^{i=N} \in \mathcal{R}$ are the predictors and target variables, respectively. b is a constant, ω is the weighted vector, and $\phi(X)$ represents the mapping function employed in the feature space. A minimisation technique is used to estimate the coefficients ω and b as follows [3]:

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi_i^*) \tag{2}$$

$$\text{Subject to } \begin{cases} |y_i - (\langle w, x_i \rangle + b) \geq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{3}$$

where the smoothness of the function is determined by $\frac{1}{2} \|w\|^2$, C and ε are the model’s parameters and the nonnegative slack variables (ξ and ξ^*) demonstrate the distance between actual and equivalent boundary values of a function approximation. A nonlinear regression function can be expressed by Eq. 4 after applying Lagrangian multipliers and optimising conditions [3]:

$$f(X) = \sum_{i=1}^{i=N} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \tag{4}$$

where x_i and $x_j \in X$, and the term $K(x_i, x_j)$ denotes the kernel function. α_i and α_i^* are Lagrangian multipliers [3].

In this study, the radial basis function (RBF) was used in the processing of the SVR model as follows [4]:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

where the kernel width and inputs are represented by σ and x_i, x_j , respectively. The critical task for developing the SVR model with a good accuracy is to determine the three parameters which are kernel width (σ), the loss function (ϵ) and regulation (C) during the training period [5]. This is achieved through a hybrid method called particle swarm optimization (PSO) in section 2.3 below.

2.2 Multivariate adaptive regression splines

The relationship between X and y is demonstrated by the MARS model as follows [6]:

$$y = \hat{f}(X) = a_0 + \sum_{m=1}^M a_m BF_m(X) \quad (6)$$

where X and y are offered in Eq. 1, a_0 is a constant, $\{a_m\}_1^M$ are the model coefficients, M is the number of basis functions in MARS, and $BF_m(X)$ is a spline function defined as $C(X|s, t_1, t, t_2)$, where $t_1 < t < t_2$, and $s = \pm 1$.

The number of the basis functions for developing MARS model is determined through the Generalized Cross-Validation criterion (GCV) based on the mean square error (MSE) [6]:

$$GCV = MSE / \left[1 - \frac{\tilde{\sigma}(M)}{N}\right]^2 \quad (7)$$

where $MSE = \frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}(X_i)]^2$. $\tilde{G}(M) = C(M) + v \cdot M$, where v is a penalty factor with a characteristic value of $v = 3$. $C(M)$ is the number of parameters being fitted. In the training dataset, the lowest value of the GCV refers to the optimal MARS model.

2.3 Particle swarm optimization (PSO)

To select the best parameters of the SVR model, which are the regulation function (C), kernel width (σ) and loss function (ϵ), the PSO algorithm, first introduced by Kennedy and Eberhart [7, 8], is employed in this paper using the mean square error (MSE) as the fitness function as follows [9-11]:

$$V_{i,j}(k+1) = \omega * V_{i,j}(k) + c_1 * \text{rand}() * (P_{\text{best}_{i,j}}(k) - X_{i,j}(k)) + c_2 * \text{rand}() * (g_{\text{best}_j}(k) - X_{i,j}(k)) \quad (8)$$

$$X_{i,j}(k+1) = X_{i,j}(k) + V_{i,j}(k+1) \quad (9)$$

where $X_i = (X_{i1}, X_{i2}, \dots, X_{iD})^T$ is the i^{th} particle from the initial population of the size of $i = 1, 2, \dots, N$ and a dimension of $j = 1, 2, \dots, D$. $V_i = (V_{i1}, V_{i2}, \dots, V_{iD})^T$ is the

velocity of each particle X_i in the population. According to [9, 11], $\text{rand}()$ represents a random number between zero and one while the individual and global extreme values are represented by $P_{\text{best}_{i,j}}$ and g_{best_j} , respectively. The two values of c_1 and c_2 are usually within [2, 2.05], whereas ω can be defined as follows [9, 10]:

$$\omega = \omega_{\min} + \frac{(T_{\max} - T) * (\omega_{\max} - \omega_{\min})}{T_{\max}} \quad (10)$$

where ω_{\min} and ω_{\max} usually equal to 0.4 and 0.9; T and T_{\max} are the current and maximum iteration numbers, respectively [9].

3 Materials and Methods

3.1 Electricity demand data (G)

In this study, the G data were recorded half-hourly (48 times per day) in Megawatts (MW) for the state of Queensland, and these data were acquired from the Australian Energy Market Operator (AEMO) [12] for the period of 01-01-2015 to 31-12-2016 (dd-mm-yyyy). The 30-minute data periods were converted to daily terms by obtaining total values for each day. Figure 1 showed the plots of the actual G data series.

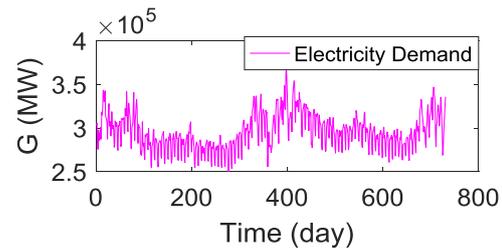


Fig. 1. Time-series of electricity demand (G MW) data.

3.2 Climate dataset

Historical climate datasets for the same period of the G data were obtained from Scientific Information for Land Owners (SILO) [13]. The data were collected for the main eleven stations, which contain the majority of the population of Queensland, that were shown in Fig. 2 and Table 1.

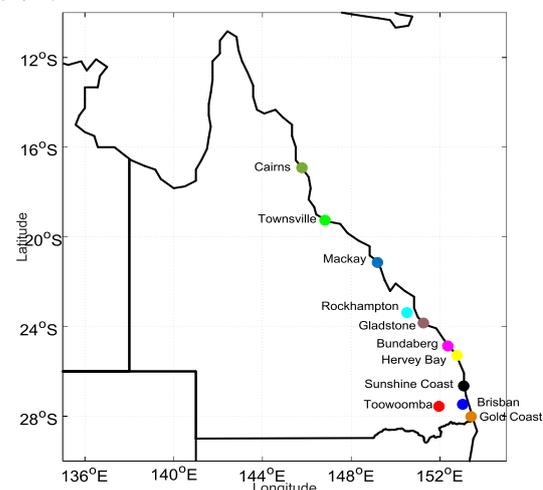


Fig. 2. Area map for the climate datasets.

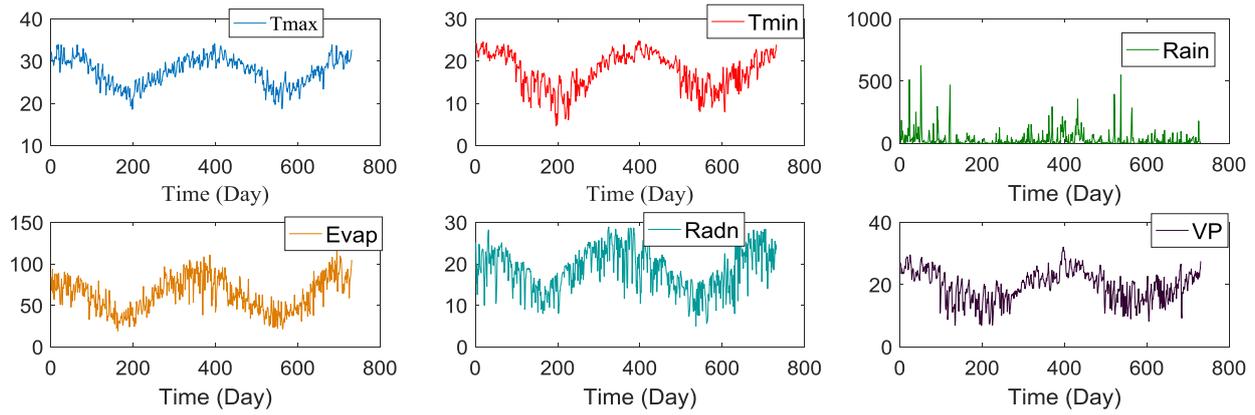


Fig. 3. Time-series of climate datasets used in this study.

Table 1. Population for each climate station

No.	Station	Population number
1	Cairns	240,190
2	Townsville	229,031
3	Mackay	43,364
4	Rockhampton	79,726
5	Gladstone	61,640
6	Bundaberg	92,897
7	Hervey Bay	53,035
8	Sunshine Coast	346,522
9	Brisbane	2,270,800
10	Gold Coast	555,721
11	Toowoomba	149,512
	Total	4,122,438

The population numbers were obtained from Australian Bureau of Statistics [14] where the total number of population resulted from the eleven stations in Table 1 is very close to the population of whole Queensland (4,883,739).

The input data were comprised of the time-series of maximum and minimum air temperature (T_{max} and T_{min}), rainfall (Rain) evaporation (Evap), solar radiation (Radn) and vapour pressure (VP). The datasets for whole Queensland were obtained by getting the average of T_{max} , T_{min} , Radn and VP and the total values of Rain and Evap of the eleven stations datasets. Those were used as the inputs of the models. Figure 3 showed the plots of those actual time series.

3.3 Forecast model development and validation

The climate variables in section 3.2 above were used to forecast the G data by developing three models: PSO-SVR, SVR and MARS. As there is no a single method for splitting data into training, validation and testing [5], the data were divided into subsets of 70% for training, 15% for validation and 15% for testing.

MATLAB-based Libsvm toolbox (version 3.22), developed by Chang and Lin [15], was used to build the SVR model. To develop a hybrid SVR model, the PSO algorithm (section 2.3) was used to select the optimal parameters based on the smallest value of MSE . To

evaluate the accuracy of the SVR model, the software packages version 1.13.0 was employed [16] for the MARS model.

The models were validated in Table 2 using the root-mean square error ($RMSE$, Eq. 11). The PSO-SVR model yielded the lowest $RMSE$, which indicated the best accuracy compared to the other models.

Table 2. Root-mean square error ($RMSE$; MW) in the validation dataset

Model	$RMSE$ (MW)
PSO-SVR	12730.52
SVR	12930.53
MARS	16701.61

3.4 Model performance evaluation

This study adopted a wide range of statistical error criteria in the testing period based on statistical indicators. Those are mean absolute error (MAE), root-mean square error ($RMSE$), relative error (%) based on MAE and $RMSE$ values ($MAPE$ and $RRMSE$), Willmott's Index (WI), the Nash-Sutcliffe coefficient (E_{NS}), and the Legates and McCabe Index (E_{LM}) [17-26], represented below:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^{i=n} |G_i^{for} - G_i^{obs}| \quad (12)$$

$$RRMSE = 100 \times \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}}{G_i^{obs}} \quad (13)$$

$$MAPE = 100 \times \frac{1}{n} \sum_{i=1}^{i=n} \left| \frac{G_i^{for} - G_i^{obs}}{G_i^{obs}} \right| \quad (14)$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (|G_i^{for} - \bar{G}^{obs}| + |G_i^{obs} - \bar{G}^{obs}|)^2} \right], \quad 0 \leq \quad (15)$$

$$WI \leq 1$$

$$E_{NS} = 1 - \left[\frac{\sum_{i=1}^{i=n} (G_i^{for} - G_i^{obs})^2}{\sum_{i=1}^{i=n} (G_i^{obs} - \bar{G}^{obs})^2} \right], \quad -\infty \leq E_{NS} \leq 1 \quad (16)$$

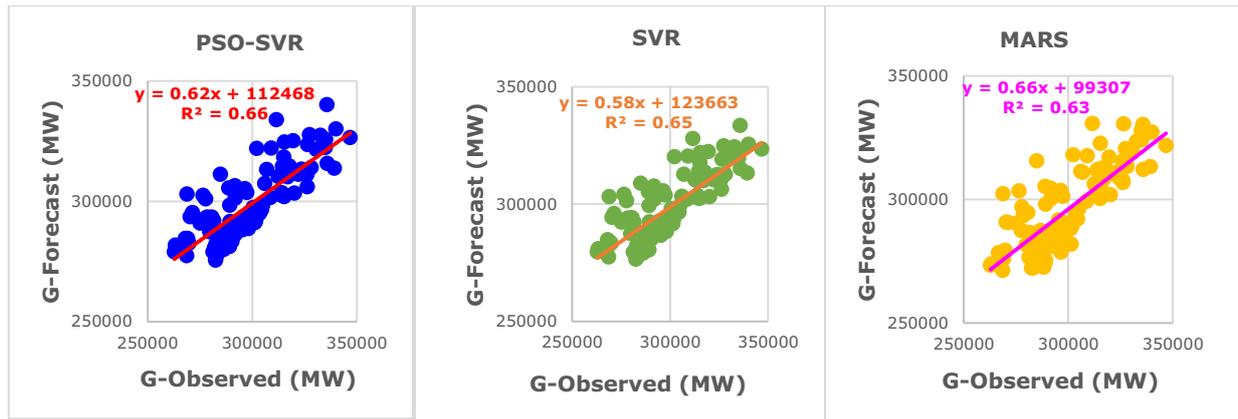


Fig. 4. Scatterplot of the G -forecasted vs. G -observed of electricity demand data in the testing period using the three models. The equation of linear regression line and the coefficient of determination are incorporated.

$$E_{LM} = 1 - \left[\frac{\sum_{i=1}^{i=n} |G_i^{obs} - G_i^{for}|}{\sum_{i=1}^{i=n} |G_i^{obs} - \bar{G}^{obs}|} \right], -\infty \leq E_{LM} \leq 1 \quad (17)$$

where G_i^{for} and G_i^{obs} are the i^{th} forecasted and observed values of G in the testing period, respectively; n is the total number of G_i^{for} or G_i^{obs} values, \bar{G}^{for} and \bar{G}^{obs} are the means of forecasted and observed values, respectively.

4 Results and Discussions

The performance of the PSO–SVR model for the daily forecast horizon was compared with traditional SVR and MARS models in the testing period. The results of the comparison indicated that the PSO–SVR yielded better performances (lowest $RMSE$, and MAE , as well as the largest WI , E_{NS} , and E_{LM}) than SVR and MARS models. Those values were summarized in Table 3.

Table 3. The performance evaluation of the models in the test period

Model	WI	E_{NS}	$RMSE$ (MW)	MAE (MW)	$MAPE$ (%)	$RRMSE$ (%)	E_{LM}
PSO–SVR	0.88	0.66	11716.97	9668.15	3.27	3.92	0.41
SVR	0.87	0.64	11909.00	9895.09	3.34	3.99	0.40
MARS	0.87	0.60	12612.22	11003.20	3.69	4.22	0.33

The scatterplots of G_i^{for} vs. G_i^{obs} and the model forecasted errors, $|FE| = |G_{FOR,i} - G_{OBS,i}|$ in the testing period for the three models were shown in Figs. 4 and 5, respectively. The lowest forecasted errors ($|FE|$) were shown by the PSO-SVR model in this study (Fig. 5). On the other hand, the highest correlation of determination (R^2) was achieved by the PSO-SVR model (Fig. 4). Overall, a significantly greater accuracy was attained by the PSO–SVR model than the other models.

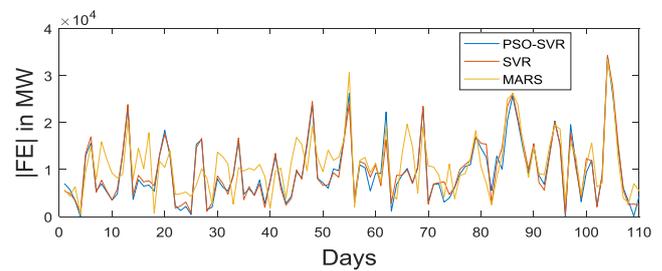


Fig. 5. Model forecasted errors, $|FE| = |G_{FOR,i} - G_{OBS,i}|$ in the testing period using the three models.

5 Concluding Remark

In this paper, a hybrid PSO–SVR model was proposed for daily G forecasting horizon in Queensland, Australia, where the model used data from Australian Energy Market Operator (AEMO) and Scientific Information for Land Owners (SILO). The MARS and the traditional method of SVR were also used in this research study to evaluate the main model. The results showed that the PSO–SVR outperformed the MARS and SVR models. As a result, the data-driven tool constructed by the PSO–SVR model is a powerful forecasting framework which can support the National Electricity Market (e.g., AEMO). Although the PSO–SVR model performed well in this paper, some challenges in model development section could be appeared. As the PSO algorithm needs a longer time to produce the SVR parameters, alternative methods, such as multi-swarm PSO and sine cosine algorithm may need to be used. In addition, the model could be improved using ensemble-based uncertainty testing by a bootstrapping technique. Those should be addressed in future studies.

Acknowledgments

Firstly, we would like to offer our thanks to the Australian Energy Market Operator (AEMO), Australian Bureau of Statistics, and Scientific Information for Land Owners (SILO) which provided all the required data. We are very grateful to the Ministry of Higher Education and Scientific Research in the Government of Iraq for funding the first author’s PhD project.

References

1. M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Advanced Engineering Informatics*, vol. 35, pp. 1-16, 2018.
2. M. S. Al-Musaylh, R. C. Deo, Y. Li, and J. F. Adamowski, "Two-phase particle swarm optimized-support vector regression hybrid model integrated with improved empirical mode decomposition with adaptive noise for multiple-horizon electricity demand forecasting," *Applied Energy*, vol. 217, pp. 422-439, 5/1/ 2018.
3. V. N. Vapnik, *Statistical learning theory* (no. Book, Whole). New York: Wiley, 1998.
4. J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1-4, pp. 85-105, 10// 2002.
5. R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Applied Energy*, vol. 168, pp. 568-593, 2016.
6. J. H. Friedman, "Multivariate adaptive regression splines," *The annals of statistics*, pp. 1-67, 1991.
7. J. Kenny, "Particle swarm optimization," in *Proc. 1995 IEEE Int. Conf. Neural Networks*, 1995, pp. 1942-1948.
8. R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, 1995, pp. 39-43: IEEE.
9. X. Li and C. Li, "Improved CEEMDAN and PSO-SVR Modeling for Near-Infrared Noninvasive Glucose Detection," *Computational and mathematical methods in medicine*, vol. 2016, 2016.
10. M. N. Alam, "Particle Swarm Optimization: Algorithm and its Codes in MATLAB," ed: ResearchGate, 2016.
11. M. Hasanipanah, A. Shahnazar, H. B. Amnieh, and D. J. Armaghani, "Prediction of air-overpressure caused by mine blasting using a new hybrid PSO-SVR model," *Engineering with Computers*, vol. 33, no. 1, pp. 23-31, 2017.
12. AEMO. (2016). *Aggregated Price and Demand Data - Historical*. Available: <http://www.aemo.com.au/>
13. S. J. Jeffrey, J. O. Carter, K. B. Moodie, and A. R. Beswick, "Using spatial interpolation to construct a comprehensive archive of Australian climate data," *Environmental Modelling & Software*, vol. 16, no. 4, pp. 309-330, 2001.
14. A. B. o. Statistics. (2018). *Census*. Available: <http://www.abs.gov.au/>
15. C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.
16. G. Jekabsons, "Adaptive Regression Splines toolbox for Matlab/Octave," *Version*, vol. 1, p. 72, 2013.
17. K. Mohammadi, S. Shamsirband, M. H. Anisi, K. A. Alam, and D. Petković, "Support vector regression based prediction of global solar radiation on a horizontal surface," *Energy Conversion and Management*, vol. 91, pp. 433-441, 2// 2015.
18. C. J. Willmott, S. M. Robeson, and K. Matsuura, "A refined index of model performance," *International Journal of Climatology*, vol. 32, no. 13, pp. 2088-2094, 2012.
19. C. J. Willmott, "On the evaluation of model performance in physical geography," in *Spatial statistics and models*: Springer, 1984, pp. 443-460.
20. C. J. Willmott, "Some comments on the evaluation of model performance," *Bulletin of the American Meteorological Society*, vol. 63, no. 11, pp. 1309-1313, 1982.
21. C. J. Willmott, "On the validation of models," *Physical geography*, vol. 2, no. 2, pp. 184-194, 1981.
22. C. W. Dawson, R. J. Abraham, and L. M. See, "HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts," *Environmental Modelling & Software*, vol. 22, no. 7, pp. 1034-1052, 2007.
23. D. R. Legates and G. J. McCabe, "Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation," *Water resources research*, vol. 35, no. 1, pp. 233-241, 1999.
24. P. Krause, D. Boyle, and F. Bäse, "Comparison of different efficiency criteria for hydrological model assessment," *Advances in Geosciences*, vol. 5, pp. 89-97, 2005.
25. T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature," *Geoscientific Model Development*, vol. 7, no. 3, pp. 1247-1250, 2014.
26. J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I—A discussion of principles," *Journal of hydrology*, vol. 10, no. 3, pp. 282-290, 1970.