# LONG-TERM ELECTRICAL ENERGY CONSUMPTION: FORMULATING AND FORECASTING VIA OPTIMIZED GENE EXPRESSION PROGRAMMING

## SEYED HAMIDREZA AGHAY KABOLI

## INSTITUTE OF GRADUATE STUDIES
## UNIVERSITY OF MALAYA
## KUALA LUMPUR

## 2018

# LONG-TERM ELECTRICAL ENERGY: CONSUMPTION FORMULATING AND FORECASTING VIA OPTIMIZED GENE EXPRESSION PROGRAMMING

## SEYED HAMIDREZA AGHAY KABOLI

## THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

## INSTITUTE OF GRADUATE STUDIES
## UNIVERSITY OF MALAYA
## KUALA LUMPUR

## 2018

# UNIVERSITY OF MALAYA

## ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Seyed Hamidreza Aghay Kaboli

Registration/Matric No: HHD120009

Name of Degree: Ph.D

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

LONG-TERM ELECTRICAL ENERGY CONSUMPTION: FORMULATING AND FORECASTING VIA OPTIMIZED GENE EXPRESSION PROGRAMMING

Field of Study: Power

I do solemnly and sincerely declare that:

(1)  I am the sole author/writer of this Work;
(2)  This Work is original;
(3)  Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4)  I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5)  I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6)  I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature:                           Date:

Subscribed and solemnly declared before,

Witness's Signature                              Date:

Name:

Designation:

# LONG-TERM ELECTRICAL ENERGY CONSUMPTION: FORMULATING AND FORECASTING VIA OPTIMIZED GENE EXPRESSION PROGRAMMING

## ABSTRACT

This study mathematically formulates the effects of two different historical data types, (i) electrical energy consumption in preceding years and (ii) socio-economic indicators (SEI) on electrical energy consumption (EEC) of ASEAN-5 countries, namely, Malaysia, Indonesia, Singapore, Thailand, and Philippines.

Firstly, a multi-objective feature selection approach is developed in this study to extract the most influential subsets of input variables from each historical data type (EEC and SEI) with maximum relevancy and minimum redundancy for long-term EEC modeling. In the developed feature selection approach, multi-objective binary-valued backtracking search algorithm (MOBBSA) is used as an efficient evolutionary search algorithm to search within different combinations of input variables and selects the non-dominated feature subsets, which minimize simultaneously both the estimation error and the number of features.

Then, in order to cope with the limitations of the existing artificial intelligence (AI) based methods, optimized gene expression programming (GEP) is applied to precisely formulate the relationships between historical data and EEC of ASEAN-5 countries. The optimized GEP as a recent extension of GEP approach is superior to other AI-based methods in giving an optimized explicit equation, which clearly shows the relationship between input historical data and EEC in different countries without prior knowledge about the nature of the relationships between independent and dependent variables. This merit is provided by balancing the exploitation of solution structure and exploration of its appropriate weighting factors through use of a robust and efficient optimization algorithm in learning process of GEP approach.

To assess the applicability and accuracy of the proposed method for long-term electrical energy consumption, its estimates are compared with those obtained from artificial neural network (ANN), support vector regression (SVR), adaptive neuro-fuzzy inference system (ANFIS), rule-based data mining algorithm, GEP, linear, quadratic and exponential models optimized by particle swarm optimization (PSO), cuckoo search algorithm (CSA), artificial cooperative search (ACS) algorithm and backtracking search algorithm (BSA). The simulation results are validated by actual data sets observed from 1971 until 2013. The results confirm the higher accuracy and reliability of the proposed method as compared with other artificial intelligence based models. On the basis of the favorable results obtained, it can be concluded that recent enhancements in AI-based approaches, as in this study, could result higher accuracy with the least complexity for long-term EEC forecasting.

Finally, future estimations of EEC in ASEAN-5 countries are projected up to 2030 by applying the rolling-based forecasting procedure on mathematical models derived from optimized GEP. Furthermore, EEC in ASEAN-5 countries is forecasted by autoregressive integrated moving average (ARIMA) model and first-order single-variable grey model (GM (1, 1)) and their forecasts are compared with those obtained by the proposed method.

**Keywords**: electrical energy consumption, forecasting, gene expression programming, optimization

**PENGELUARAN TENAGA LISTRIK TERJEMAHAN: MEMASUKKAN DAN MENINGKATKAN PROGRAMMING GENE OPTIMASI**

**ABSTRAK**

Kajian ini secara matematik merumuskan kesan dua jenis data sejarah yang berlainan, (i) penggunaan tenaga elektrik dalam tahun-tahun sebelumnya dan (ii) petunjuk sosioekonomi (SEI) mengenai penggunaan tenaga elektrik (EEC) bagi negara-negara ASEAN-5, Indonesia, Singapura, Thailand, dan Filipina.

Pertama, pendekatan pemilihan ciri multi-objektif dibangunkan dalam kajian ini untuk mengekstrak subset yang paling berpengaruh bagi pemboleh ubah input dari setiap jenis data sejarah (EEC dan SEI) dengan perkaitan maksimum dan redundansi minimum untuk pemodelan EEC jangka panjang. Dalam pendekatan pemilihan ciri yang maju, algoritma carian backtracking bernilai binary-objective (MOBBSA) digunakan sebagai algoritma carian evolusi yang cekap untuk mencari dalam kombinasi yang berbeza pembolehubah input dan memilih subset ciri yang tidak dikuasai, yang meminimumkan secara serentak kedua-dua anggaran kesilapan dan bilangan ciri.

Kemudian, untuk mengatasi batasan kaedah berasaskan kecerdasan buatan (AI) sedia ada, pengaturcaraan pengekspresian gen yang dioptimumkan digunakan untuk merumuskan hubungan antara data sejarah dan EEC negara-negara ASEAN-5. GEP yang dioptimumkan sebagai pendekatan baru GEP yang dioptimumkan adalah lebih tinggi daripada kaedah berasaskan AI yang lain dalam memberikan persamaan eksplisit yang dioptimumkan, yang jelas menunjukkan hubungan antara input data sejarah dan EEC di negara-negara yang berbeza tanpa pengetahuan terlebih dahulu tentang sifat hubungan antara bebas dan pembolehubah bergantung. Kelebihan ini disediakan dengan mengimbangi eksploitasi struktur penyelesaian dan penerokaan faktor penimbang yang

sesuai melalui penggunaan algoritma pengoptimuman yang mantap dan cekap dalam proses pembelajaran pendekatan GEP.

Untuk menilai kebolehgunaan dan ketepatan kaedah yang dicadangkan untuk penggunaan tenaga elektrik jangka panjang, anggarannya dibandingkan dengan yang diperoleh daripada rangkaian saraf tiruan (ANN), sokongan vektor regresi (SVR), sistem inferensi neuro-fuzzy adaptif (ANFIS) algoritma perlombongan data berasaskan peraturan, model GEP, linier, kuadrat dan eksponen yang dioptimumkan oleh pengoptimuman swarm partikel (PSO), algoritma carian pintar (CSA), algoritma carian koperasi buatan (ACS) dan algoritma carian mundur (BSA). Hasil simulasi disahkan oleh set data sebenar yang diperhatikan dari tahun 1971 hingga 2013. Hasilnya mengesahkan ketepatan dan keandalan yang lebih tinggi dari metode yang dicadangkan dibandingkan dengan model berdasarkan kecerdasan buatan yang lain. Berdasarkan hasil yang menggembirakan yang diperolehi, dapat disimpulkan bahawa peningkatan baru-baru ini dalam pendekatan berasaskan AI, seperti dalam kajian ini, dapat menghasilkan ketepatan yang lebih tinggi dengan kerumitan paling rendah untuk peramalan EEC jangka panjang.

Akhir sekali, anggaran masa depan EEC di negara-negara ASEAN-5 diunjurkan sehingga 2030 dengan menggunakan prosedur ramalan berasaskan rolling mengenai model matematik yang diperoleh daripada GEP yang dioptimumkan. Tambahan pula, EEC di negara-negara ASEAN-5 diramalkan oleh model purata bergerak bersepadu autoregressive (ARIMA) dan model kelabu tunggal-ubah tunggal (GM (1, 1)) dan ramalan mereka dibandingkan dengan yang diperolehi melalui kaedah yang dicadangkan.

**Keywords**: penggunaan tenaga elektrik, peramalan, pengaturcaraan gen gen, pengoptimuman

# ACKNOWLEDGEMENTS

First and foremost, I give glory to Almighty Allah for sound health and preserving my life during the course of this thesis. I would like to take this opportunity to express my sincere gratitude to my supervisors, Prof. Nasrudin Bin Abd Rahim and Associate Prof. Jeyraj A/l Selvaraj for their valuable suggestions, motivation, enthusiasm, and the immense continuous support rendered to me during my PhD research. I would like to extend my appreciation to UM Power Energy Dedicated Advanced Centre (UMPEDAC) for the rich research resources.

Appropriative words could not be found to express sincere appreciation to my parents for their endless patience, selfless loves, unconditional supports, and constant encouragements. Moreover, I would like to thank all my family especially my lovely wife, grandmother, sister, and brother for their motivational supports. It has been a long road, but they have been with me the whole time.

This thesis is dedicated to all of them.

My best wishes to them,

Seyed Hamidreza Aghay Kaboli

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

**LIST OF SYMBOLS AND ABBREVIATIONS**

| | | |
|---|---|---|
| $y_{k;x_g^k}$ | : | Fitness of $(x_g^k)$ |
| $(\xi_k, \xi_k^*)$ | : | Slack variables |
| $(A_i, B_i)$ | : | Fuzzy sets |
| $(x, y)$ | : | Inputs to node $i$ |
| $(\alpha_k, \alpha_k^*)$ | : | Nonnegative Lagrange multipliers |
| $\langle W, x_k \rangle$ | : | Vector inner product of the predictors |
| $\hat{x}^{(0)}(t + H)$ | : | H-step ahead predicted values |
| $\widehat{y_k}$ | : | Estimated output of the regression function |
| $(\beta_k, \beta_K^*)$ | : | Lagrangian multipliers |
| $\overline{w}$ | : | Normalized firing strength of a rule |
| $\mu_{Ai}$ | : | Membership function for $Ai$ fuzzy sets |
| $\mu_{Bi}$ | : | Membership function for $Bi$ fuzzy sets |
| $L(\beta)$ | : | Lévy distribution function |
| $\Gamma(.)$ | : | Gamma distribution function |
| $\overline{\overline{EEC}}$ | : | Normalized electric energy consumption |
| $:=$ | : | Update operation |
| $\partial$ | : | Adjustable parameter in Gaussian RBF |
| $b$ | : | Intercept of the regression function |
| $C$ | : | Positive constant regularization parameter |
| $c_1, c_2$ | : | Acceleration factors |
| d | : | Number of differences (I) that are needed to make the series stationary |
| $EEC\ (t)_{observed}$ | : | Observed electricity consumption at year t |

| $EEC(t)$ | : | Predicted electricity consumption at year t |
|---|---|---|
| $f$ | : | Activation function |
| $F$ | : | Wiener process |
| $f_i(x, y; p_i, q_i, r_i)$ | : | Output of the Sugeno type FIS |
| $f_j{}^{max}$ | : | Minimum value of the $j^{th}$ objective function |
| $f_j{}^{min}$ | : | Maximum value of the $j^{th}$ objective function |
| $g$ | : | Transfer function |
| $g+1$ | : | Next generation |
| $g_{best}$ | : | Overall best value |
| GM (1, 1) | : | First-order single-variable grey model |
| $h$ | : | Head length |
| $h_{0j}$ | : | Weight assigned to the bias unit of $j^{th}$ neuron in output layer |
| $h_{ij}$ | : | Connection strength between $i^{th}$ neuron in last hidden layer and $j^{th}$ neuron in output layer |
| $k$ | : | Time step |
| $key$ | : | Memory to track the origin of Predator in each iteration |
| $L$ | : | Hidden layer |
| $low_j$ | : | Lower search space limits of $j^{th}$ variable |
| $M$ | : | Binary integer-valued matrix |
| $m$ | : | Number of objectives |
| $map$ | : | Binary integer-valued matrix |
| $mixrate$ | : | Control parameter of BSA |
| $Mutant$ | : | Initial form of trial population |
| $N$ | : | Standard normal distribution |
| $\eta$ | : | Lagrangian multipliers |

| | | |
|---|---|---|
| $n_{max}$ | : | Number of arguments for the function that takes the most arguments |
| $n_j$ | : | Neuron in the first hidden layer |
| $n_{Lj}$ | : | Output signal $j^{th}$ neuron of total (N) neuron in hidden layer |
| $n_{max}$ | : | Maximum number of iterations |
| $Np$ | : | Number of nests |
| $nPop$ | : | Population size of host nests |
| $nVar$ | : | Number of respective optimization variable |
| $oldP$ | : | Historical population |
| $Ø_p$ | : | Parameter of the AR model |
| p | : | Number of time lags for the autoregressive model |
| $P$ | : | Control parameter of ACS |
| $p_a$ | : | Probability of an alien egg |
| $P_{best}$ | : | Previous best position |
| $permuting$ | : | Random shuffling function |
| q | : | Order of the moving average |
| $Q_{1,i}$ | : | Membership grade of a fuzzy set |
| r | : | PPMCC |
| $rand$ | : | Distributed random numbers |
| $randi$ | : | Random selection function |
| $randp$ | : | Random perturbation |
| $randperm$ | : | Random permutation function |
| $r_s$ | : | Spearman's rank correlation coefficient |
| $S$ | : | Sigmoid function |
| $t$ | : | Tail length |
| $T$ | : | Generated offspring at the end of crossover process |

| | | |
|---|---|---|
| $U$ | : | Uniform distribution function |
| $up_j$ | : | Upper search space limits of $j^{th}$ variable |
| $v_n$ | : | Velocity of a particel |
| $W$ | : | Weight vector of the regression function |
| $w_{0j}$ | : | Activation threshold |
| $w_i$ | : | Initial weight |
| $w_{ij}$ | : | Connection strength between neurons |
| $w_{max}$ | : | Maximum boundary of inertia weight |
| $w_{min}$ | : | Minimum boundary of inertia weight |
| $w_n$ | : | Inertia weight |
| $x$ | : | Mutation matrix |
| $X^{(0)}(t)$ | : | Non-negative sequence |
| $x^{best}$ | : | Previous best nests |
| $x_g^{best}$ | : | Nest with highest productivity among all nests in $g^{th}$ generation |
| $x_g^k$ | : | Randomly selected host egg in $g^{th}$ generation |
| $x_i$ | : | Input unit |
| $x_k$ | : | $k^{th}$ element in n -dimensional input vector |
| $x_n$ | : | Position of a particel |
| $y_i$ | : | Productivity of $i^{th}$ individual |
| $y_{i;x}$ | : | Productivity of $i^{th}$ host nest |
| $y_{i;\alpha}$ | : | Productivity of $i^{th}$ sub-superorganism related to $\alpha$ superorganism |
| $y_{i;\beta}$ | : | Productivity of $i^{th}$ sub-superorganism related to $\beta$ superorganism |
| $y_k$ | : | Observed response values |
| $Y_{t-p}$ | : | Time-lagged value |

| | | |
|---|---|---|
| z | : | Output of ANFIS model |
| $\alpha$ | : | Learning rate |
| $\beta$ | : | Distribution factor |
| $\varepsilon$ | : | Degree of tolerable errors |
| $\theta_q$ | : | Parameters of the MA model |
| $\sigma_{share}$ | : | Sharing parameter |
| $\psi\,(x_k)$ | : | Kernel function |
| 1-AGO | : | First-order accumulated generating operation |
| AAGR | : | Average annual growth rate |
| ABC | : | Artificial bee colony |
| ACO | : | Ant colony optimization |
| ACS | : | Artificial cooperative search |
| AERN | : | ASEAN energy regulators' network |
| AI | : | Artificial intelligence |
| AISO | : | APG independent system operator |
| ANFIS | : | Adaptive neuro-fuzzy inference system |
| ANN | : | Artificial neural network |
| APG | : | ASEAN power grid |
| AR | : | Auto regressive |
| ARIMA | : | Auto regressive integrated moving average |
| ASEAN | : | Association of Southeast Asian Nations |
| BBL | : | One barrel of crude oil |
| BBSA | : | Binary-valued BSA |
| BP | : | Back propagation |
| BS | : | Binary search |
| BSA | : | Backtracking search optimization algorithm |

| | | |
|---|---|---|
| CD | : | Crowding distances |
| CDE | : | Carbon dioxide emissions |
| CSA | : | Cuckoo search algorithm |
| CSS | : | Charged system search |
| DE | : | Differential evolution |
| DGs | : | Distributed generators |
| EE | : | Energy exchange |
| EEC | : | Electrical energy consumption |
| ENS | : | Efficient non-dominated sorting |
| ETs | : | Expression trees |
| EXP | : | Export of goods and services |
| FCM | : | Fuzzy c-means |
| FIFO | : | First-in, first-out |
| FIS | : | Fuzzy inference system |
| GA | : | Genetic algorithm |
| GDP | : | Gross domestic product |
| GEP | : | Gene expression programming |
| GM | : | Gray model |
| GP | : | Genetic programming |
| GSA | : | Gravitational search algorithm |
| HAPUA | : | Heads of ASEAN power utilities/authorities |
| HS | : | Harmony search |
| HVAC | : | High-voltage alternating current |
| HVDC | : | High-voltage direct current |
| IAGO | : | Inverse accumulated generating operation |
| ICA | : | Imperialist competitive algorithm |

| | | |
|---|---|---|
| IMP | : | Import of goods and services |
| IR | : | Industrialization rate |
| KBES | : | Knowledge-based expert system |
| KFD | : | Kernel Fisher discriminant analysis |
| KGOE | : | Kilograms of oil equivalent |
| KKT | : | Karush-Kuhn-Tucker |
| KW | : | Kilowatt |
| LNG | : | Liquefied natural gas |
| M5-R | : | M5-Rules |
| MA | : | Moving average |
| MAPE | : | Mean absolute percentage error |
| MF | : | Membership function |
| MLP | : | Multi-layer perceptron |
| MMT | : | Million metric tons |
| MOBBSA | : | Multi-objective BBSA |
| MOBSA | : | Multi-objective BSA |
| MOPSO | : | Multi-objective particle swarm optimization |
| MTOE | : | Million tonnes of oil equivalent |
| MW | : | Megawatts |
| NN | : | Neural network |
| NSGA | : | Non-dominated sorting genetic algorithm |
| O&M | : | Operation and maintenance |
| PCA | : | Principal component analysis |
| POE | : | Price of energy |
| POP | : | Population |
| PP | : | Power purchase |

| PPMCC | : | Pearson product-moment correlation coefficient |
| PSO | : | Particle swarm optimization |
| PV | : | Photovoltaic |
| QP | : | Quadratic programming |
| RACF | : | Residuals autocorrelation function |
| RBF | : | Radial basis function |
| RE | : | Renewable energy |
| RMSE | : | Root mean square error |
| RNC | : | Random numeric constants |
| S&C | : | Separate-and-conquer |
| SA | : | Simulated annealing |
| SEI | : | Socio-economic indicators |
| SI | : | Stock index |
| SS | : | Sequential search |
| SVC | : | Support vector classification |
| SVM | : | Support vector machine |
| SVR | : | Support vector regression |
| T&D | : | Transmission and distribution |
| TCF | : | Trillion cubic feet |
| TEC | : | Total energy consumption |
| TLBO | : | Teaching learning based optimization |
| TWh | : | Terawatt-hour |
| UR | : | Urbanization rate |
| WB | : | World Bank |

**CHAPTER 1: INTRODUCTION**

## 1.1     Introduction

Today, existing grids are under pressure to deliver the growing demand for power, as well as provide a stable and sustainable supply of electricity. These complex challenges are driving the evolution of smart grid technologies. Since the smart grid is taken as the future power grid development goal. The construction of the smart grid will exert significant impacts on the electric power industry. In smart grid environment, the capacity of distributed generators (DGs), transmission and distribution (T&D) system's efficiency will be optimized, thus it brings a challenge to the grid's stability while storing the electricity for future use has lots of difficulty and requires huge investment. Improper and inaccurate forecasts on this area will lead to electricity shortage, energy resource waste, loss of profit due to the penalty paid for under/over estimate of electricity consumption and even grid collapse. Therefore, accurate electricity demand forecasting is essential to move towards the smart grid technology (J. Wang, Li, Niu, & Tan, 2012).

According to the time horizon, the electricity consumption forecasting is classified as short-term, medium-term and long-term forecasts. Short-term forecasting (several days ahead in hourly steps) has attracted substantial attention due to its importance for power system control, economic dispatch and the order of unit commitment in electricity markets. Midterm forecasting (several months ahead in weekly or longer steps) is especially interesting for companies operating in a deregulated environment, as it provides them with valuable information about the market need of energy, scheduling the maintenance of the units, the fuel supplies, electrical energy imports/exports. Long-term (years ahead in annual or longer steps) forecasting has been always playing a vital role in power system management and planning. The accuracy of long-term load forecast directly impacts on effectiveness of energy trading, system reliability, operation and maintenance (O&M) costs, T&D expanding, and generators scheduling. Moreover,

accurate long-term power load forecasting can provide reliable guidance for power grid development and power construction planning, which is also important for the sustainable development of any country.

Accurate forecasts are also a prerequisite for decision makers to develop an optimal strategy that includes risk reduction and improving the economic and social benefits. The accurate long-term load forecast gives the more realistic spectrum of future country's energy sources consumption for moving towards sustainable development in a globalizing world while the growing global population is driving an even greater increase in the electricity consumption.

Electrical energy consumption (EEC) reflects the degree of economic development, and much evidence supports a causal relationship between economic growth and energy consumption. Association of Southeast Asian Nations (ASEAN) is one of the largest economic zones in the world with rapid and relatively stable economic growth. In fact, ASEAN has experienced much lower volatility in economic growth since 2000 than the European Union (analysis, 2013). If ASEAN considered as a single economic entity, it would already rank as the sixth-largest economy in the world, trailing the US, China, Japan, Germany, and the United Kingdom (*ASEAN Community in Figures (ACIF), 2013 (6th ed.), Jakarta: ASEAN*, Retrieved 9 May 2015).

ASEAN is a major global hub of manufacturing and trade, as well as one of the fastest-growing consumer markets in the world (analysis, 2013). As the region seeks to deepen its ties and capture an even greater share of global trade, its economic profile is rising which directly reflects on EEC.

According to the World Bank (WB) data bank, ASEAN's electricity consumption has changed dramatically since the early 1970s with average annual growth rate of 8.58% that is almost two (2.7) time more than the average annual growth rate of the global EEC. Only the five largest economies in this area (ASEAN-5 countries); Malaysia, Indonesia,

Singapore, Thailand, and Philippines consumed 52.65 MTOE in 2013 as shown in Figure 1.1, which ranked the ASEAN-5 countries as the world's sixth-largest electricity consumer, behind the China, US, Japan, India, and Russia. So, long-term forecasting of EEC to manage a power system, and fulfill power requirements with consideration of economic growth in the future is one of the most critical and challenging issues for sustainable development of ASEAN-5 countries.
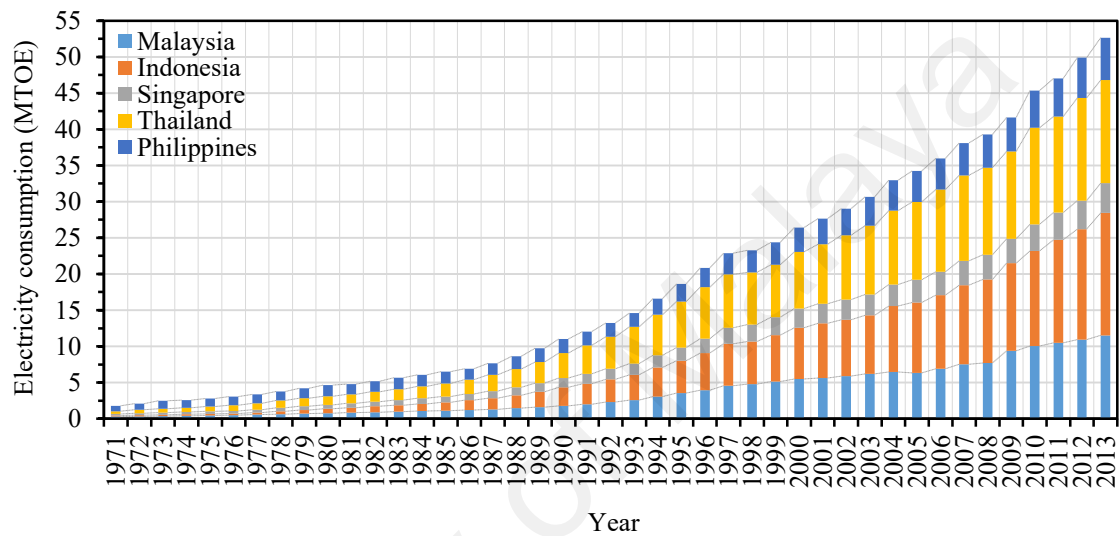


**Figure 1.1: The annual aggregate amount of EEC in ASEAN-5 countries from 1971 to 2013**

### 1.1.1    ASEAN Power Grid

ASEAN's high economic growth has rapidly increased electric energy consumption. Consequently, it has driven up the energy security risks. ASEAN recognizes the critical role of an efficient, reliable, and resilient electricity infrastructure in stimulating regional economic growth and development. To meet the growing electricity demand, huge investments in power generation capacity will be required.

Since the ASEAN-5 countries seeking to meet the expected growth in demand of the power sector in coming decades, investment in additional generating capacity and grids that is both sustainable and cost effective will be the biggest challenge.

At present, both the availability and the affordability of fuel supply are being prioritized over environmental sustainability; hence, fossil fuels, particularly coal and gas fired turbines, dominate the fuel mix. Efforts to use energy resources effectively are hampered by the uneven distribution of these resources and different levels of investment and economic development among ASEAN member countries. Sufficient financial resources, enabling governance environments, and regional coordination are critical drivers for reliable (available), sustainable, and affordable power systems. Secure, sustainable, and affordable electricity generation is vital to support regional economic growth in any region. Thus, these three daunting challenges have to be addressed when facing investment in this region (IEA, 2015). Together, these three challenges constitute the power supply trilemma as imply a trade-off when choosing one over the other. The energy trilemma is illustrated in Figure 1.2.
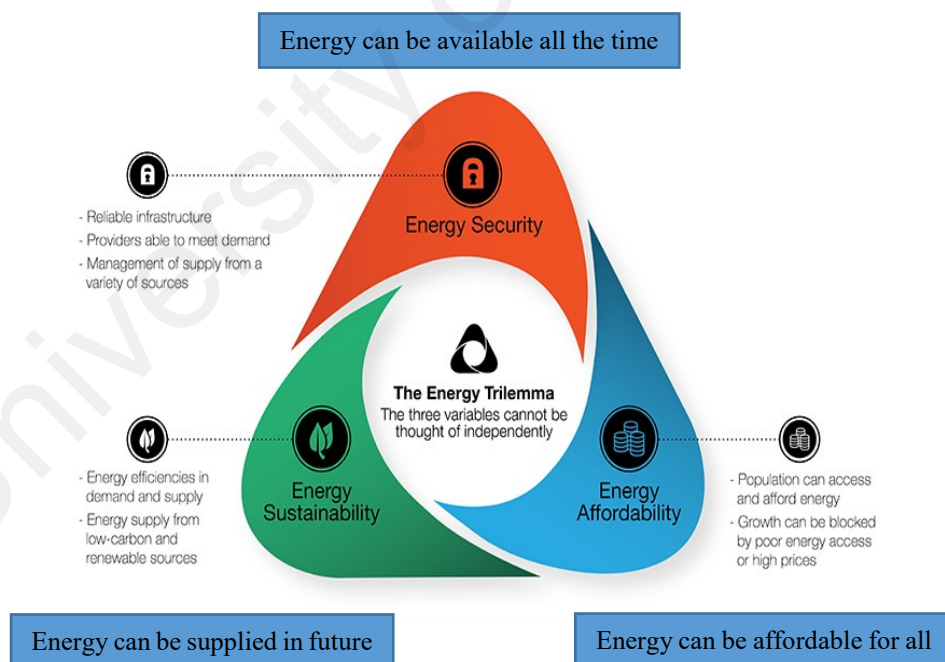


**Figure 1.2: The energy trilemma**

A regional ASEAN Power Grid (APG) would address the power supply trilemma by connecting countries with surplus power generation capacity to those facing a deficit; this would allow ASEAN countries to meet rising energy demand, improve access to energy services, and reduce the costs of developing an energy infrastructure. In recognizing the potential advantages to be gained from the establishment of integrated systems, the electricity interconnecting arrangements is mandated within the region through the APG. The primary aim of APG is to ensure regional energy security by promoting the effective utilization and sharing of resources for common regional benefit. In addition, the APG could help eliminate inefficient generation, lowering overall costs, and making more efficient electricity generation in the region. An interconnected power system could also further enhance the development and integration of variable renewable power generation capacity, which would bring benefits such as enhanced energy security and environmental sustainability. The APG will contribute to the creation of the provision for future energy trade, and mutually exploit the abundant energy resources within the region and reduce the dependency of fuel imports from other regions.

Energy resources, including fossil fuels and renewables, are abundant throughout the geographical region of ASEAN. The summary of energy resources of ASEAN-5 countries is reported in Table 1.1 (T. Ahmed et al., 2017). Indonesia is the largest oil producer, corresponding to its possession of the largest oil reserve in the region. However, Malaysia is the only net oil exporter of the region. Indonesia and Malaysia stand in the top of natural gas reserves. Thailand and Singapore are the net liquefied natural gas (LNG) importer in this region, while Malaysia and Indonesia are net exporters of LNG. Coal is most abundant fossil fuel in the region. Indonesia, Thailand, and Malaysia have the highest amount of coal reserves, respectively. Indonesia is the world largest steam coal exporters. While, Thailand, Malaysia, and Philippines are importers of steam coal.

The ASEAN-5 countries have an abundance of hydropower resources, including large, mini, micro, and pico hydropower plants. Indonesia and Malaysia have massive potentials for hydropower. Moreover, Philippines and Thailand possess great resources for hydropower generation, and they are actively developing this sector.

ASEAN-5 countries have a great potential for non-hydro based renewable generation. A significant supply of biomass energy is available in this region, from agricultural residues of rice husks, rice straw, corn cobs, sugarcane trash, cassava stalks, bagasse, as well as coconut and palm oil. Indonesia, Malaysia, and Thailand are the top three countries that have the highest theoretical biomass energy reserve, respectively, as Indonesia and Malaysia are the highest palm oil producers in the world and 40% of the Thai populations are actively depend on agriculture sector for livelihood. However, the technical and economic potential of biomass energy is much less due to difficulty of collecting these residues from its distributed geographic territory.

Philippines and Indonesia are the second and third largest geothermal power generators in the world, respectively. The rest of countries have not exploited their respective geothermal energy resources potential as of yet.

Solar is one of the most important and usable clean energy sources in the world, and due to the fact that ASEAN-5 countries are generally tropical, the region has the highest solar irradiation, at an average of 4.5 kW h/m$^2$, encompassing a significant area. Solar photovoltaic (PV) prospects and utilization of individual ASEAN countries has been reviewed in (Ismail, Ramirez-Iniguez, Asif, Munir, & Muhammad-Sukki, 2015) and shows that ASEAN countries have annual solar insolation level ranging from 1460 to 1892 kW h/m$^2$ per year. Consequently, Malaysia and Thailand are significantly advantaged when it comes to solar energy.

With the exception of Singapore, most ASEAN-5 countries have great potentials for onshore wind energy potentials. Thailand and Philippines have the highest theoretical wind energy potentials respectively. Furthermore, ASEAN-5 countries are generally located in coastal areas; hence, there is a great possibility for offshore wind energy generation. While all ASEAN-5 countries have high potential for harvesting offshore wind energy, it is necessary to exploit the offshore wind potentials for this region.

**Table 1.1: Energy resources in ASEAN-5 countries**

| Country | Oil (BBI)[1] | Gas (TCF)[2] | Coal (MMT)[3] | Hydro (MW) | Biomass (MW) | Geothermal (MW) | Solar (MW) | Onshore wind (MW) | Offshore wind (TW.h, 2030) |
|---|---|---|---|---|---|---|---|---|---|
| Malaysia | 3.42 | 84.40 | 1024.50 | 29500 | 29000 | – | 1412 | 2599 | 13.39 |
| Indonesia | 10 | 169.5 | 38000 | 75625 | 49810 | 29000 | 551 | 9300 | 21.34 |
| Singapore | – | – | – | – | – | – | – | – | 0.22 |
| Thailand | 0.16 | 12.20 | 1240 | 16655 | 22831 | – | 3000 | 190000 | 19.42 |
| Philippines | 0.28 | 4.60 | 346 | 13107 | 20 | 2047 | 350 | 76000 | 6.96 |

1. one barrel of crude oil (BBL)
2. trillion cubic feet (TCF)
3. million metric tons (MMT)

Since electricity cannot be cost effectively stored on a national scale, a country develops domestic electricity sources to achieve higher self-sufficiency. As it shows in Table 1.1, ASEAN-5 countries are unevenly endowed with power generation resources such as coal, natural gas, and hydro. Nevertheless, some countries in the region have more resources than required to meet domestic demand, others fail to develop sufficient electricity sources on their own due to resources shortages.

International power grid interconnection is a solution to this problem. It resolves difficulties in power resources endowment so, it allows a region to develop electricity infrastructure more efficiently than individual countries.

The continuing efforts of the ASEAN-5 countries in strengthening and restructuring their respective power market industry are oriented towards this direction. Electricity is produced through a mix of oil, gas, coal, hydro, geothermal and other renewable energy sources. Enhancing electricity trade across borders, through integrating the national

power grids, is expected to provide benefits of meeting the rising electricity demand and improving access to various energy resources.

The efficient utilization of clean energy resources to meet increasing electricity demand is imposing the integration of the electricity market and the construction of secure transmission mechanisms around the globe. Accordingly, the ASEAN-5 countries are integrating their large geographical power transmission infrastructure via APG.

ASEAN-5 countries have an abundance of renewable resources throughout their geographical region. However, the distribution is far from uniform, they have high potential to further harness renewable energy (RE), especially hydro, geothermal, biomass/biogas, wind, and solar power. Table 1.2 tabulates the present status of installed capacity of renewable generators in ASEAN-5 countries and their future targets (T. Ahmed et al., 2017). It can be observed that ASEAN-5 countries have less amount of installed capacity for renewables at present. Indonesia has the largest renewable power generation in this region with only 1353 MW installed capacity, followed by Thailand, Philippines, Malaysia, and Singapore. Recognizing the benefits that RE provides in terms of energy security, most countries have set individual targets and support schemes, which directly support the APG targets. According to Table 1.2, the target of RE generation in each country is utilizing the abundant renewable sources to generate the maximum amount of clean energy.

**Table 1.2: Installed capacity of renewable power generators for ASEAN-5 countries in 2013**

| Country | Installed Capacity (MW) | Target of RE generation |
|---------|-------------------------|-------------------------|
| Malaysia | 129 | 985 MW in 2015 (~ 5.5% of energy mix), 2080 MW in 2020 and 4000 MW in 2030 |
| Indonesia | 1353 | 17% of total primary energy consumption in 2025 and 25.9% in 2030 |
| Singapore | 10 | 4% of total generated electricity from RE sources in 2030 |
| Thailand | 984 | 13701 MW; 25% share of RE in 2021 |
| Philippines | 171 | 15234.3 MW in 2030 |

The primary advantages of system integration are the increase in security of supply and efficiency. Larger service territories allow for the pooling of generating resources, thus taking advantage of generation diversity. Therefore, APG supports access to multi-technology and geographically dispersed RE resources. Furthermore, system integration may boost renewable power generation as variable sources can be supported by flexible generation technologies. In the short term, this could lead to considerably greater exploitation of hydroelectric resources, while significantly higher targets for modern renewable energy can be achieved only in the medium to long-term (IEA, 2015).

The rapid growth of energy demand, caused by an increasing population as well as favorable economic growth rates, results in ASEAN-5 countries being faced with pressure on energy access and energy security. This is reinforced by the complicated archipelagic geography of the region. Even though they are a net exporter of energy, the countries differ very much in their reserves of fossil fuels and are dependent on imports of at least one fossil fuel. High fossil fuel dependency is due to the uneven distribution of renewables throughout the geographical region, high capital cost involvement of renewables generation, and the lack of transmission expansion planning for remotely located renewable generators. Mature renewable energy technologies such as hydro and geothermal are developed in the region but still have large potential for further expansion. New technologies like solar and wind start to see their deployment in recent years but their fraction in the total energy mix remains negligible.

Renewable power generations could be expedited by utilizing semi-shallow transmission expansion planning. In addition, the ASEAN power market integration via the establishment of APG could be another possible solution in meeting the increasing electricity demand from clean energy sources. The establishment of APG will create a sustainable and secure power system network, where investors can invest beyond borders

to renewable generators, and could easily transfer the generated power from cross border trades. In addition, APG will reduce the investment cost of electricity generation and increase the net savings of the ASEAN-5 countries.

The geographical map of APG interconnections is illustrated in Figure 1.3. APG have been divided into three regions, namely Eastern, Northern, and Southern regions. It can be observed from Figure 1.3 that among sixteen interconnection projects, some interconnection projects are already in operation, some are ongoing, and rests of the interconnection projects will be established in future. The status of existing, ongoing and future projects are given in Tables 1.3 – 1.6, respectively according to updates from Heads of ASEAN Power Utilities/Authorities (HAPUA) secretariat (T. Ahmed et al., 2017). Table 1.3 shows that seven projects of APG are in operation with cross border power transfer of 5032 –5192 MW, while, Table 1.4 illustrates that five projects of APG are under construction, which will allow 5589 MW of cross border power transfer. From Table 1.5 it can be seen that, another twelve projects of APG are in planning stage with a capacity of 24,829– 27,979 MW cross border power transfer.

APG as a flagship program is an initiative to construct a regional power interconnection to connect the region, first on cross-border bilateral terms, and then gradually expand to sub-regional basis and subsequently leading to a total integrated South East Asia power grid system. So, the long-term strategic goals of APG can be concisely summarized as follows:

- To facilitate and expedite the implementation of the ASEAN Interconnection Master Plan and to further harmonize technical standards and operating procedures as well as regulatory and policy frameworks among the ASEAN Member States.

- To achieve a long-term security, availability and reliability of energy supply, particularly in electricity through regional energy cooperation in Trans-ASEAN Energy Network.

- To optimize the region's energy resources towards an integrated ASEAN Power Grid system.

- To further harmonize all aspect of technical standard and operating procedure as well as regulatory frame works among member country.
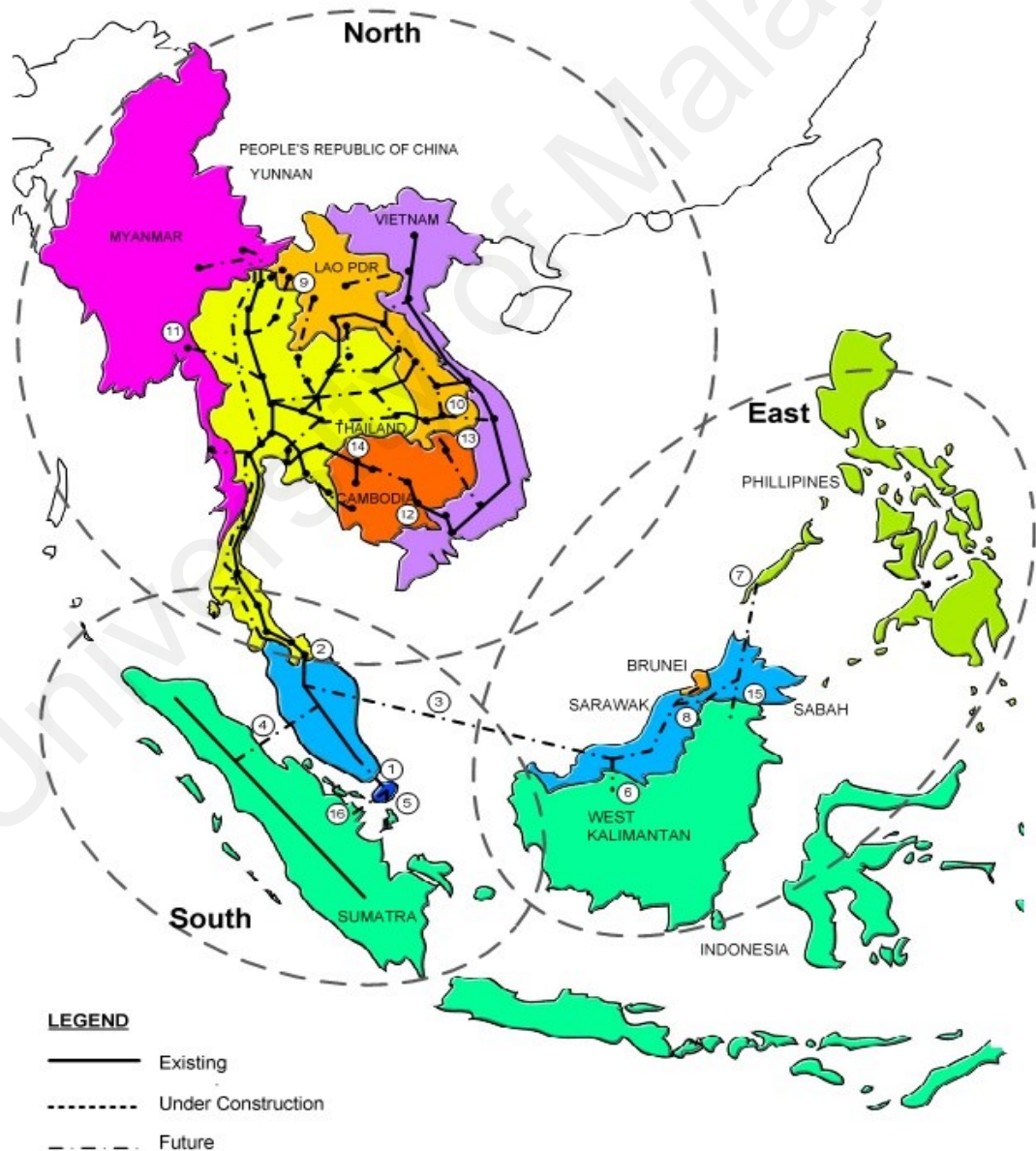


**Figure 1.3: Geographical map of APG interconnections**

## Table 1.3: APG existing projects

| No. | Project | System | Type | Capacity (MW) |
|---|---|---|---|---|
| 1 | **P. Malaysia –Singapore** | | | |
| | Plentong– Woodlands | HVAC: 230 kV | EE[1] | 450 |
| 2 | **Thailand - P.Malaysia** | | | |
| | Sadao - Bukit Keteri | HVAC: 132/11 5 kV | EE | 80 |
| | Khlong Ngae - Gurun | HVDC: 300 kV | EE | 300 |
| 6 | **Sarawak– West Kali-mantan** | | | |
| | Mambong – Bengkayang | HVAC: 275 kV | EE | 70-230 |
| 9 | **Thailand– Lao PDR** | | | |
| | Nakhon Phanom –Thakhek – Then Hinboun | HVAC: 230 kV | PP[2] | 220 |
| | Ubon Ratchathani 2 – Houay Ho | HVAC: 230 kV | PP | 126 |
| | Roi Et 2 - Nam Theun 2 | HVAC: 230 kV | PP | 948 |
| | Udon Thani 3– Na Bong – Nam Ngum 2 | HVAC: 500 kV | PP | 597 |
| | Nakhon Phanom 2 – Thakhek – Theun Hin-boun | HVAC: 230 kV | PP | 220 |
| | Mae M oh 3 – Nan 2 – Hong Sa # 1, 2, 3 | HVAC: 500 kV | PP | 1473 |
| 10 | **Lao PDR– Vietnam** | | | |
| | Xekaman 3 - Thanhmy | HVAC: kV | PP | 248 |
| 12 | **Vietnam– Cambodia** | | | |
| | Chau Doc – Takeo – Phnom Penh | HVAC: 230 kV | PP | 200 |
| 14 | **Thailand– Cambodia** | | | |
| | Aranyaprathet – Bantey Meanchey | HVAC: 115 kV | PP | 100 |
| | Total Capacity | | | 5032-5192 |

1: Energy Exchange (EE), 2: Power Purchase (PP)

## Table 1.4: APG on-going projects

| No. | Project | System | Type | Capacity (MW) |
|---|---|---|---|---|
| 2 | **Thailand -P.Malaysia** | | | |
| | Su – ngai kolok – Rantau Panjang | HVAC: 132/115 k V | EE[1] | 100 |
| 4 | **P. Malaysia – Sumatra** | | | |
| | Melaka - Pekan Baru | HVDC: kV | PP[2] & EE | 600 |
| 8 | **Sarawak– Sabah – Brunei** | | | |
| | Sarawak – Brunei | HVAC: 275 k V | EE | 2 * 100 |
| 9 | **Thailand– Lao PDR** | | | |
| | Udon Thani 3 – Na Bong – Nam Ngiep | HVAC: 500 kV | PP | 269 |
| | Ubon Ratchathani 3– Pakse – Xe Pien Xe Namnoi | HVAC: 500 kV | PP | 390 |
| | Khon Kaen 4 – Loei 2 – Xayaburi | HVAC: 500 kV | PP | 1220 |
| 10 | **Lao PDR– Vietnam** | | | |
| | Xekaman 1 - Ban Hat San - Pleiku | HVAC: 500 kV | PP | 1000 |
| | Nam Mo - Ban Ve | HVAC: 230 kV | PP | 100 |
| | Luang Prabang - Nho Quan | HVAC: 500 kV | PP | 1410 |
| 13 | **Lao PDR– Cambodia** | | | |
| | Ban Hat – Stung Treng | HVAC: 230 kV | PP | 300 |
| | Total Capacity | | | 5589 |

1: EE (Energy Exchange), 2: PP (Power Purchase)

**Table 1.5: APG future projects**

| No. | Project | System | Type | Capacity (MW) |
|---|---|---|---|---|
| 1 | **P. Malaysia – Singapore** Plentong – Woodlands (2nd link) | HVDC: kV | PP[1] | 600 |
| 2 | **Thailand– P. Malaysia** Khlong Mgae – Gurun (Addition) | HVDC: 300 kV | EE[2] | 300 |
| 3 | **Sarawak– P. Malaysia** | HVDC: kV | PP | 2 * 800 |
| 5 | **Batam – Singapore** | HVAC: kV | PP | 3 * 200 |
| 7 | **Philippines – Sabah** | HVDC: kV | EE | 500 |
| 8 | **Sarawak – Sabah – Brunei** Sarawak – Sabah | HVAC: 275 kV | PP | 100 |
| 9 | **Thailand– Lao PDR** | | | |
| | Nong Khai – Khoksa - at | HVAC: 230 kV | EE | 600 |
| | Nakhon Phanom – Thakhek | HVAC: 230 kV | EE | 510 |
| | Thoeng – Bo Keo | HVAC: 230 kV | EE | 315 |
| | Udon Thani 3 – Na Bong | HVAC: 500 kV | PP | 1040 |
| | Ubon Ratchathani 3– Pakse | HVAC: 500 kV | PP | |
| | Nan 2 – Tha Wang Pha – Nam Ou | HVAC: 500 kV | PP | |
| 10 | **Lao PDR – Vietnam** | | | |
| | Xekaman 1 - Pleiku 2 | HVAC: 230 kV | PP | 290 |
| | Luang Prabang – Nho Quan | HVAC: 500 kV | PP | 1600 |
| | Nam Mo - Ban Ve | HVAC: 230 kV | PP | |
| 11 | **Thailand – Myanmar** | | | |
| | Mai Khot – Mae Chan – Chiang Rai | HVAC: 230 kV | PP | 369 |
| | Hutgyi – Phitsanulok 3 | HVAC: 500 kV | PP | 1190 |
| | Mong Ton – Sai Noi 2 | HVAC: 500 - 800 kV | PP | 3150 - 6300 |
| | Myanmar– Thailand | HVAC: 500 kV | PP | 7000 |
| 12 | **Vietnam - Cambodia** Tay Ninh – Strung Treng | HVAC: 230 kV | PP | 465 |
| 14 | **Thailand – Cambodia** | | | |
| | Battambang – Prachin Buri 2 | HVAC: 230 kV | EE | 300 |
| | Stung Meteuk (Mnum) – Trat 2 | HVAC: 230 kV | PP | 100 |
| | Koh Kong - Thailand | HVAC: 500 kV | PP | 1800 |
| 15 | **E. Sabah– E. Kalimantan** Sipitang – East Kalimantan | HVAC: kV | EE | 200 |
| 16 | **Singapore – Sumatra** Sumatra – Singapore | HVDC: kV | PP | 600 |
| | Total Capacity | | | 24829 - 27979 |

1: EE (Energy Exchange), 2: PP (Power Purchase)

Two primary advantages of system integration are the increase in security of supply and efficiency. Larger service territories allow for the pooling of generating resources, thus taking advantage of the benefits of generation diversity. This diversity also has the ability to aggregate demand. Power systems can be integrated through coordination or complete consolidation. In the ASEAN context, complete consolidation is impractical, not least because of geographical factors, but also because of complete consolidation

would necessitate the establishment of a single market operator with authority that stretches across multiple jurisdictions, requiring changes in national laws. Consolidation is achievable, however, at a sub-regional level. Between the various sub-regions, coordination is a more efficient option for power sector integration. Thus, the efficient governance of APG can be achieved under both liberalized and regulated market.

Principally, development of the power sector needs a strong, reliable, and depoliticized governance framework. A precondition for such a governance framework is an independent and strong regulator. As regulator plays a pivotal role in a regional market, the ASEAN Energy Regulators' Network (AERN) is formally established among the ASEAN energy regulators to forge closer cooperation among ASEAN energy departments with a view of promoting sustainability and economic development of the region.

A regulatory agency (also regulatory authority, regulatory body, or regulator) is independent from other branches or arms of the governments for exercising autonomous authority over APG operation in a regulatory or supervisory capacity. It deals in the areas of administrative law, regulatory law, secondary legislation, and rule/policy making. Accordingly, the AERN must be formally separated from the executive branch (i.e. department of energy in each country), and governed by statute without executive political influence on the regulation process.

In liberalized markets, efficiency can only be obtained by having transparent procedures, fair grid access, and a substantial number of market players. The electricity prices for final consumers generally consist of the costs of generation, network, retailing, taxes and levies as well as profit margins. The market and regulatory system need to ensure that all these components are fully covered to stimulate future investment. Tariffs should be set in such a way to cover these costs. It is also critical to define and designate

the operation and maintenance responsibilities of each regulator early on, to avoid overlap and misunderstanding of roles. Additionally, matters pertaining to cross-border energy transfer must be managed in line with practice. Hence, APG independent system operator (AISO) and the AERN will work closely to address the technical, legal and economic issues of cross border interconnections for multilateral electricity trade in the region. Their key responsibilities include establishing electricity security regulations, allocating the cost of transmission development, revising network codes, system monitoring, allocating the interconnection capacity, providing the mechanisms to deal with congested capacity within the national power systems, facilitating the connection of new producers to the power system and providing the plan for future expanding of power system.

### 1.1.2    Overview of EEC in ASEAN-5 Countries

### 1.1.2.1    Malaysia

As illustrated in Figure 1.1, the electric energy consumption of Malaysia has been growing from 0.3 MTOE (26.63 KGOE per-capital) in 1971 to 11.53 MTOE (387.96 KGOE per-capital) in 2013 with the average annual growth rate of 9.19%. The Malaysia's compositions of gross power generation for the last five years (2009-2013) adapted from statistical report on electric power industry conducted by energy commission is illustrated in Figure 1.4 (Malaysia Energy Information Hub Unit, 2015). According to this report in 2009, 48.1%, 35.51%, 14.24%, and 2.18% of gross power generation composition had generated from natural gas, crude oil, coal, and hydropower respectively. On that year, 39.37%, 29.04%, 18.97%, 12.26%, 0.22%, and 0.13% of total generation were for industries, commercial sector, residential sector, grid loss (transmission loss plus power plants consumption), agricultural sector, and public transportation respectively. The electricity sales to various consuming sectors from 2009 until 2013 are shown in Figure 1.5.

Malaysia as a developing country has been subject to numerous perturbations on its economy. In recent years, infrastructure limitations such as concerns about energy consumption, scarcity of resources, fluctuation of fuel price, fluctuation on electricity consumption patterns, and economic crisis in this country have forced the government to move toward utilizing RE for sustainable development of power system. Hence, beyond year 2011, biodiesel, biomass, biogas, and solar power plants have been developed to redress the power system due to high-expected demand in the following years. The electricity consumption trend of Malaysia has been changing gradually due to the increase in population, urban life, and economic growth. The portions of electric energy consumption have changed to 41.71%, 30.06%, 16.62%, 8.14%, 0.28%, and 0.18% for industries, commercial sector, residential sector, grid loss, agricultural sector, and public transportation respectively while 44.06%, 35.79%, 16.61%, 2.96%, 0.33%, 0.21%, 0.04%, and 0.01% of the total electric power consumption have been supplied from natural gas, crude oil, coal, hydropower, biomass, biodiesel, solar, and biogas respectively in 2013.
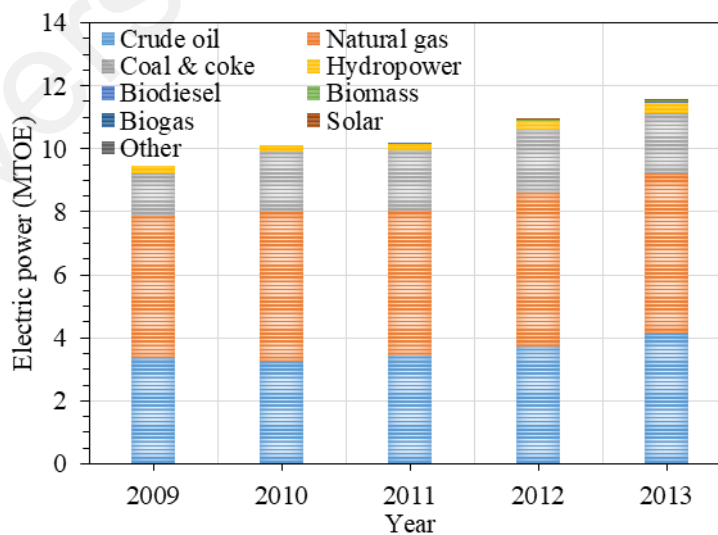


**Figure 1.4: Malaysia's composition of gross power generation**
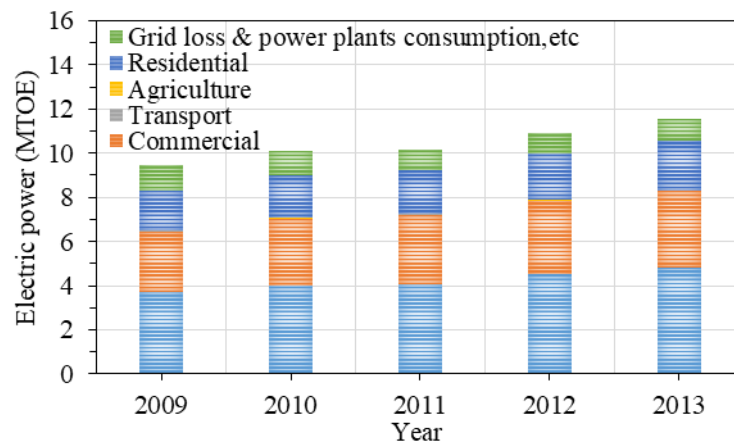
**Figure 1.5: Electricity sale to various consuming sectors in Malaysia**

### 1.1.2.2 Indonesia

Indonesia as a developing country has the largest population and economy in the ASEAN region. The electric energy consumption in this country has been growing from 0.14 MTOE (1.23 KGOE per-capital) in 1971 to 16.92 MTOE (67.73 KGOE per-capital) in 2013 with the average growth rate of 12.36%.

As illustrated in Figure 1.6, in 2011, 73.79% of its fuel mix came from fossil fuel sources (36.14% oil–fired, 19.39% gas-fired, and 18.26% coal-fired power plants). The remainder is made up of biomass (21.57%), hydroelectric (2.21%), geothermal (1.16%), and other renewables (1.26%). According to Figure 1.7, on that year, 40.55%, 34.1%, 24.59%, 0.67%, and 0.08% of total power generation were for household sector, industries, commercial sector, grid loss (transmission loss plus power plants consumption), and public transportation respectively (Ministry of Energy and Mineral Resources, 2015).

Indonesia is one of the leading exporters of steam coal in the world and also one of the largest exporters of LNG. Since 2004, the country's oil production has been declining and as a result insufficient to cover the oil demand. Indonesia has therefore become a net importer of oil. Thus, Indonesia is focusing on the use of locally available energy sources

such as coal-fired generation and its geothermal potential to increase its energy diversity and lessen its dependency on oil.

Indonesia is one of the two countries in the ASEAN region that has abundant geothermal resources. Additionally, it produces a noteworthy part of its power from biomass and hydro sources. According to Figure 1.6, in 2013, 26.45% of Indonesia's power production is covered by renewables (19% biomass, 2.39% hydroelectric, 4.41% biofuel, and 0.66% geothermal). The remainder is made up of fossil fuel sources (33.67% oil, 25.46% coal, and 14.41% gas).

While urban development is high, rural electrification faces a multitude of challenges. As shown in Figure 1.7, the portions of electric energy consumption have not changed significantly since last five years. Although the total electric energy consumption reached to 16.92 MTOE in 2013, the portions of electric energy consumption remained as 40.87%, 34.08%, 24.25%, 0.73%, and 0.07% for household sector, industries, commercial sector, grid loss, and public transportation respectively.
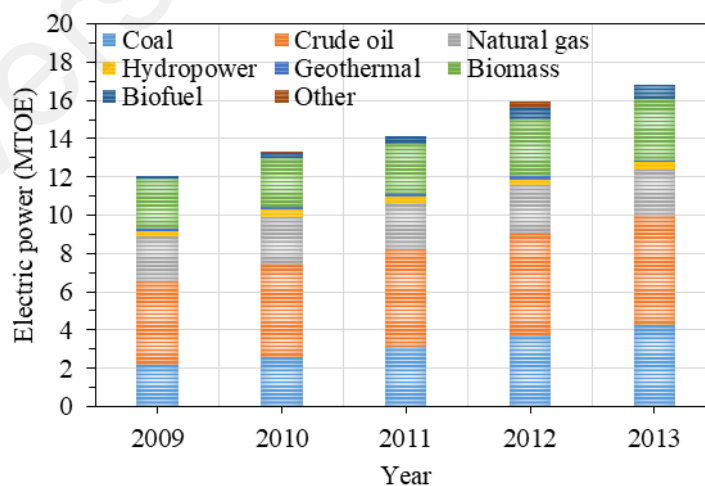


**Figure 1.6: Indonesia's composition of gross power generation**
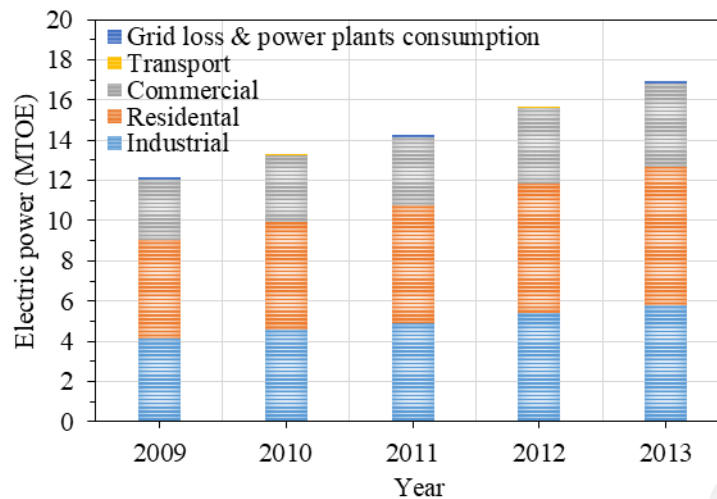
**Figure 1.7: Electricity sale to various consuming sectors in Indonesia**

### 1.1.2.3 Singapore

Given Singapore's mature economy as compared to other ASEAN countries, electric energy consumption growth is relatively slow. As shown in Figure 1.1, the electric energy consumption of Singapore has been growing from 0.21 MTOE (99.3 KGOE per-capital) in 1971 to 4.1 MTOE (760.01 KGOE per-capital) in 2013 with the average growth rate of 7.42%. Singapore is fully dependent on imported fuel resources for its power generation. As shown in Figure 1.8, approximately 90% of Singapore's electric energy consumption has been produced from gas-fired generation in 2013. Figure 1.9 illustrates the electricity sales to various consuming sectors in this country from 2009 until 2013. As shown in this figure, in 2013, 41.92%, 37.19%, 15.03%, 5.27%, and 0.58% of total power generation have been used for industries, commercial sector, household sector, public transportation, and grid loss (transmission loss plus power plants consumption) respectively (The Energy Market Authority (EMA), 2016).

Due to limited land area and natural endowments, Singapore is recognized as alternative energy disadvantaged country. Although, Singapore has no reserves of fossil fuels, it plays a major role as an oil trading and refining hub for the region.
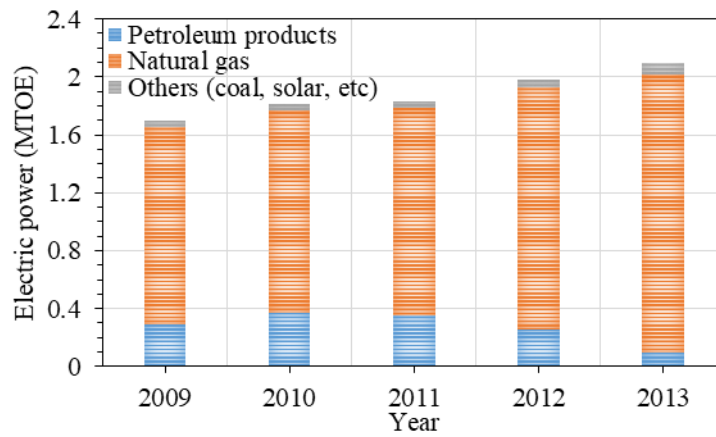
**Figure 1.8: Singapore's composition of gross power generation**
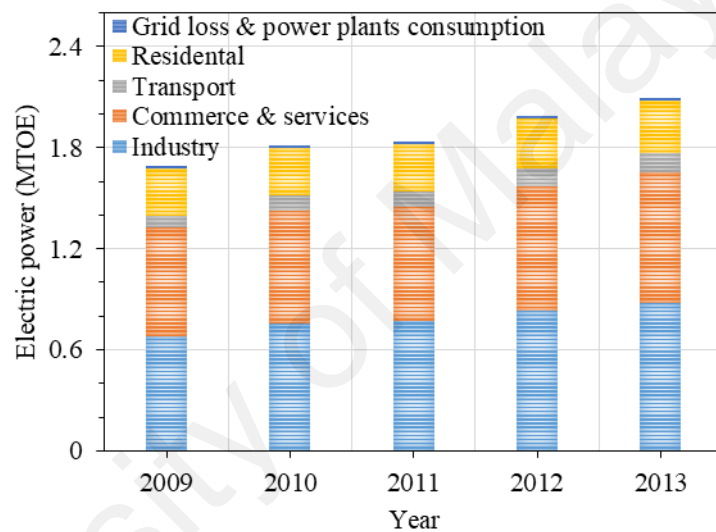


**Figure 1.9: Electricity sale to various consuming sectors in Singapore**

### 1.1.2.4  Thailand

Thailand's electric energy consumption has been growing from 0.39 MTOE (10.34 KGOE per-capital) in 1971 to 14.24 MTOE (212.45 KGOE per-capital) in 2013 with the average growth rate of 9.06%. Thailand's composition of gross power generation for the last five years (1992-2013) is reported in Figure 1.10. According to this report (Energy Policy and Planning Office (EPPO), 2015), in 2013, natural gas makes up 70.21% of power generation as domestic oil and coal reserves are very limited. The remainder is

made up of lignite (11.24%), coal (9.57%), RE (4.24%), hydropower (3.87%), oil (0.75%), and diesel (0.1%).

Dependency on gas-fired power generation makes Thailand vulnerable to fluctuations in the international market, and poses important concerns for electricity supply and power security. Thailand's national power development plan focuses on increasing green energy to maintain the security and adequacy of the power system. Hence, beyond year 2009, renewable power plants have been added to the power system due to high-expected demand in the following years. Since the production is not sufficient to cover country's energy needs, Thailand is a net importer of fossil fuels as well as electricity.

The electricity sales to various consuming sectors in this country from 2009 until 2013 are reported in Figure 1.11. As illustrated in this report, Thailand's electricity import has been 5.54% of total power generation in 2009. Against this backdrop, Thailand will play a major role in the APG as future interconnections with Lao PDR, Cambodia, Myanmar, and Malaysia, which have the potential to boost security of supply and present the opportunity of additional electricity imports. Increasing import capacity would help Thailand to decrease its gas dependency, decarbonize its electricity sector, and increase access to generation capacity.
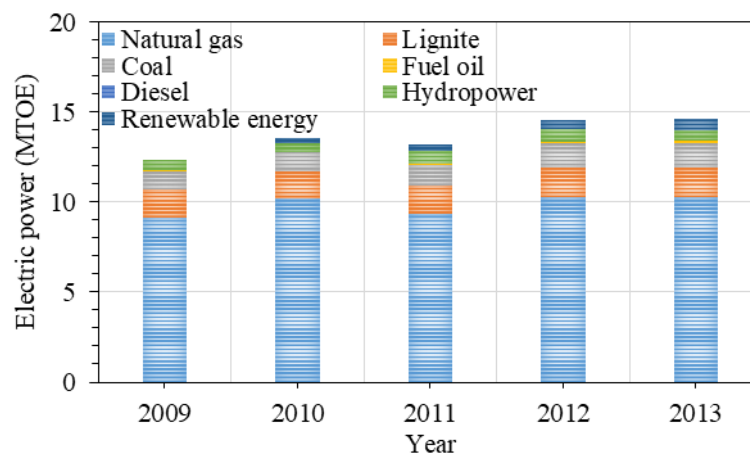


**Figure 1.10: Thailand's composition of gross power generation**
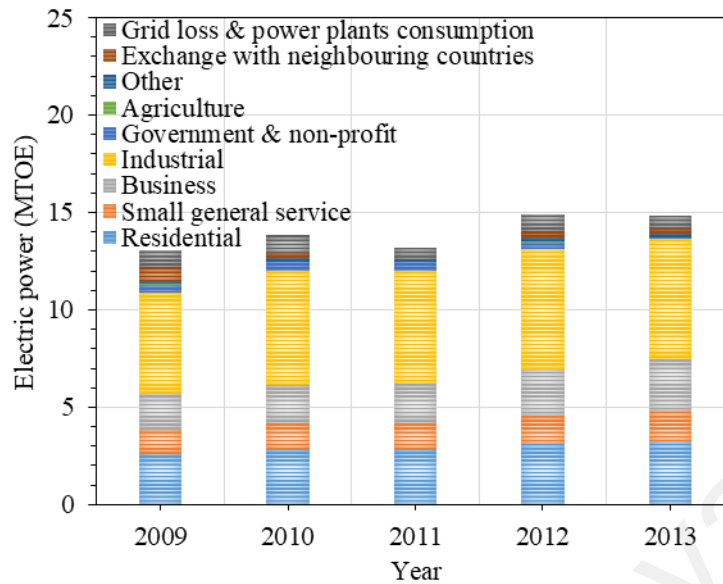
**Figure 1.11: Electricity sale to various consuming sectors in Thailand**

### 1.1.2.5 Philippines

The electric energy consumption of Philippines has been growing from 0.75 MTOE (20.27 KGOE per-capital) in 1971 to 5.85 MTOE (59.51 KGOE per-capital) in 2013 with the average growth rate of 5.19%. As shown in Figure 1.12 (Department of Energy, 2013), the power generation mix is relatively balanced among coal (34.79%), gas (28.8%), geothermal (13.25%), hydropower (12.52%), oil (10.48%), and other renewables (0.15%) in 2013. The country is the world's second-largest consumer of geothermal energy and has a high capacity for renewable energy. In 2009, geothermal provided at about 16.67% of its electric energy consumption, and still has potential for further substantial expansion of geothermal power generation.

The Philippines consists of over 7000 islands. Due to this complex geography, the government faces challenges for household electrification. The electricity sales to various consuming sectors from 2009 until 2013 are reported in Figure 1.13 (Department of Energy, 2013). As it reported, the portion of electric energy consumption has not grown significantly for residential sector in this country.
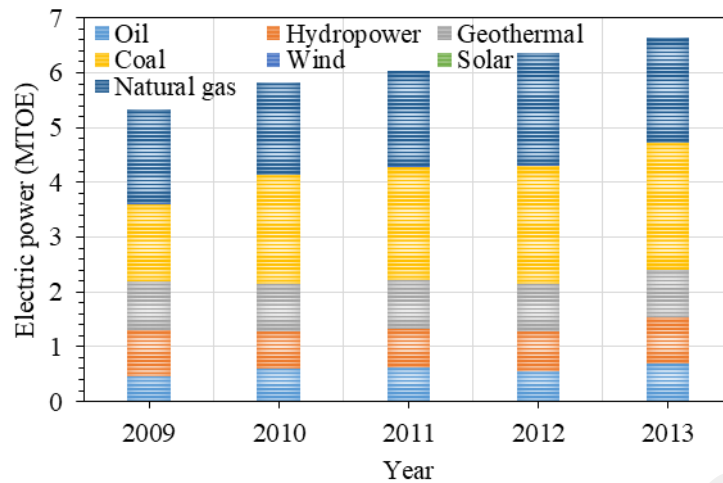
**Figure 1.12: Philippines's composition of gross power generation**
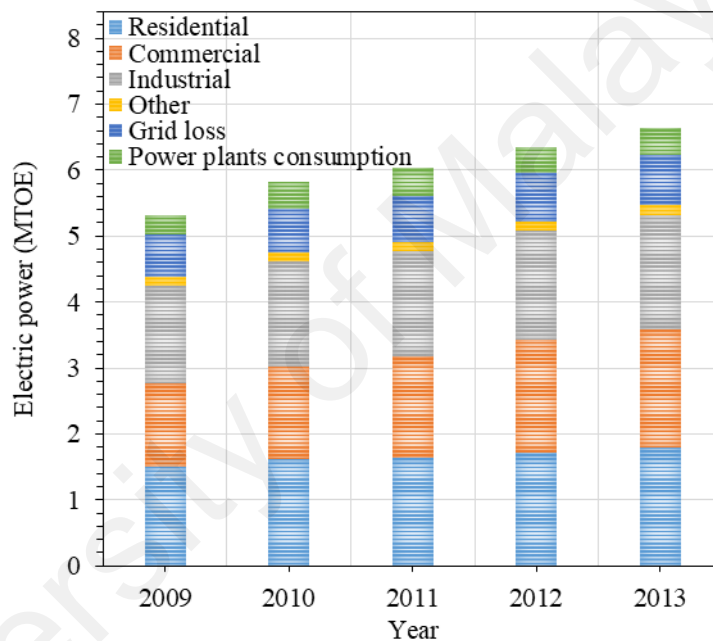


**Figure 1.13: Electricity sale to various consuming sectors in Philippines**

## 1.2    Problem Statement

The geographical distribution of the energy resources of ASEAN-5 countries is illustrated in Figure 1.14. It can be inferred from this figure that the integration of the energy market could enhance the utilization of the energy resources of the region. Though ASEAN-5 countries are rich in energy resources, meeting this increasingly electric energy demand at its regular business pace will be challenging. The uneven distribution of renewable energy resources and different levels of economic development among ASEAN-5 countries complicate efforts to effectively use energy resources to meet

demand not just between but also within countries. In particular, the electric power supply trilemma of the sustainability, availability and affordability of fossil fuels (oil, gas, coal) poses challenges for power supply in Southeast Asia. While growing regional power demand leads to increased competition for available energy resources and tightens the availability of conventional fuel supply, sustainability concerns call for increasingly cleaner energy supplies.

However, the APG is a solution for power supply trilemma in this region; it has been faced significant hurdles in creating a fully liberalized regional electricity market. Given the prevalence of subsidies, the absence of modern metering systems in most countries and a need to invest in economically unviable rural electrification from a utility viewpoint are the main difficulties in creating a fully liberalized regional electricity market. Moreover, ASEAN-5 countries vary greatly in their size, political, economy, geography, and national energy resources. Geographically, ASEAN-5 countries are mountainous and in many cases separated by large bodies of water, making some physical interconnections a technical and economic challenge. ASEAN-5 countries also vary considerably with regard to their power sector regulations, market structure, and technical characteristics. Both physical and institutional infrastructures need to be in place for regional energy cooperation to function properly. Due to the economic differences between the ASEAN-5 countries, not all the countries have the luxury to produce and distribute electricity easily. Nor will they all have the same abilities to distribute electricity consistently and the electricity access in this region varies greatly from country to country. Therefore, the electricity consumption rate in the ASEAN region is not uniform. Similarly, there is a large discrepancy between ASEAN-5 countries in installed capacity of power. The problem becomes more acute when the socio-economic conditions, legal, political, technical, and cultural diversity between ASEAN-5 countries are considered. It recognizes that even if a fully consolidated regional market may not be achievable in the

foreseeable future, the growth in electricity demand of ASEAN-5 countries can be more efficiently met by establishing long and medium-term regional plans to add capacity and expand or upgrade APG. Regional planning and cost sharing allow for larger investments than would otherwise be possible. Having clear and agreed upon plans can also translate to savings of both time and cost. For example, building a higher rated transmission line to cater for long-term forecasted demand growth is better than building lower rated lines, which would only have to be upgraded later to meet future demand.

As APG currently lacks enforcement, planning, and budgetary powers, ASEAN member states need to start conducting adequate studies and look at workable governance models to increase electricity trade across countries. Additionally, it is essential that the AISO and AERN have clear knowledge about future electric energy consumption of each country to expand the APG efficiently according to the long-term mutual benefit of interconnected countries. It is necessary to study each country individually since ASEAN-5 countries have different dependency on EEC with wide range of electric energy resources and have different level of economic development. The inherent differences among ASEAN-5 countries have important implications for the different power systems in terms of markets (pricing, impact of subsidies), governance frameworks (institutions, policies), electricity security (national resources, electrification, emergency), as well as region-wide initiatives, at both individual country and regional levels. With an almost linear trend in population growth and strong economic development, demand growth may soon surpass the available capacity for generation. So, these inherent differences among ASEAN-5 countries will become increasingly important both as challenges and as opportunities, which are summarized in Figure 1.14.

**Thailand**

Second largest electric energy consumer in ASEAN and heavily dependent on energy imports due to limited energy resources; aims to diversify electricity generation.

**Opportunities** :
* Resource of biomass
* High reserve of natural gas

**Challenges :**
* High dependency on imported energy
* Energy subsidies
* Utilization of green energy

**Malaysia**

Third largest electric energy consumer in ASEAN with relatively high per-capital consumption; it has a significant margin of capacity over demand. It exports oil and LNG significantly, but production is maturing. Additionally, Malaysia is the world's second largest producer of palm oil.

**Opportunities** :
* Refinery & petrochemical business
* LNG business

**Challenges :**
* Energy subsidies
* Utilization of green energy

ASEAN-5 Countries

**Philippines**

Fast rising electricity demand requires expanded supplies; strongly reliant on energy imports, thought it has the world's second largest installed geothermal electric capacity beyond the US and it has overtaken all other members of ASEAN in terms of installed wind energy capacity. Philippines consists of over 7000 islands, thus the electrification is very challenging in this country.

**Opportunities :**
* Geothermal energy
* Oil retail business
* Wind energy

**Challenges :**
* Limited petroleum resource
* Complex geography

**Singapore**

Strategically situated, it has become ASIA's key oil trading and refining hub (the third largest in the world) and could become a major gas hub.

**Opportunities :**
* Oil trading and refining hub

**Challenges :**
* High dependency on imported energy

**Indonesia**

Largest electric energy consumer in ASEAN, with massive scope for growth; it exports steam coal (the world's largest) and liquefied natural gas (LNG), and is an increasing importer of oil. It has the world's third largest installed geothermal electric capacity. Additionally, Indonesia is the world's largest producer of palm oil.

**Opportunities :**
* Coal business
* Palm oil plantation
* Exploration and production (E&P) business in oil and gas industry

**Challenges :**
* Rural electrification
* Energy subsidies
* Utilization of green energy

**Figure 1.14**: **Summary of energy challenges and opportunities in ASEAN-5 countries**

The outlook of electric energy consumption in ASEAN-5 countries highlights that energy optimization strategy and systematic management of power system are a matter of sustainable development necessity due to high-expected demand in the following years and variations in trend of electricity consumption in the preceding years. Reliable energy consumption forecasting can provide effective decision-making support on developing strategic plans for power system enterprises and establishing resourceful energy policies. Thus, ASEAN-5 countries should give high consideration to the electric energy consumption forecasting to prudently develop their national power grids and provide the plan for future expanding of APG. Owning to dependency of electric energy consumption

on various parameters, considering most influential input variables in long-term EEC modeling is a daunting task. In order to cope with the difficulties associated with energy consumption modeling, an efficient methodology is required to accurately model and forecast long-term EEC of ASEAN-5 countries. To address the difficulties involved in long-term energy consumption forecasting whereas, there is an implicit relationship among historical data, different techniques have been proposed to achieve a robust model with high accuracy. So far, the time series methods, artificial intelligence (AI) based approaches, and hybrid methods have been employed for long-term EEC forecasting. Despite the satisfactorily performance of AI-based approaches, and hybrid form of these methods for long-term EEC forecasting, the main shortcoming still is black-box problem that they do not provide the knowledge of process for obtaining a solution. Hence, they are not capable of generating a definite prediction equation based on the input historical data. In this regard, the application of new forecasting techniques, which can overcome the shortcomings of the existing AI-based approaches, is encouraged to precisely formulate the relationships between the historical data and EEC of ASEAN-5 countries.

## 1.3    Research Objectives

The main objectives of this study are:

- To develop optimized gene expression programming approach for coping with the limitations of existing AI-based approaches in long-term EEC forecasting.

- To develop multi-objective feature selection technique for extracting the most influential variables with minimum redundancy and maximum relevancy on EEC.

- To formulate the effects of two different historical data types, (i) EEC in preceding years and (ii) socio-economic indicators (SI) on EEC of ASEAN-5 counties.

- To access the accuracy of the proposed method for long-term EEC forecasting of ASEAN-5 countries and compare the estimations with different methodologies.

- To project the future EEC of ASEAN-5 countries up to 2030.

## 1.4     Scope of Study

The strength of the proposed technique lies in its ability to provide explicit forecasting expressions with better-fit solutions. The optimized GEP is used in this study to mathematically formulate the effects of two different input data sets (i.e. socio-economic indicators and electricity consumption in preceding years) on EEC of the ASEAN-5 countries from 1971 to 2014. Furthermore, the parallel comparison is carried out to evaluate the effectiveness of two different historical data types on long-term EEC forecasting. All historical data types in this study are adapted form World Bank (WB) data bank. The rolling-based forecasting procedure is implemented to anticipate the annual EEC of ASEAN-5 countries up to 2030. To assess the applicability and accuracy of the proposed method, its results are compared with AI-based approaches that have been applied for long-term energy consumption forecasting in a successful manner. All the simulations are evaluated in MATLAB (R2014a) environment on a personal computer, with a core-2 quad processor of 2.6 GHz clock speed and 2 GB RAM.

## 1.5     Organization of Thesis

The rest of the thesis is organized as follows:

First, the background on the concepts involved in this work and the related literature review are presented. In chapter three, long-term EEC forecasting models are described. In this chapter, the basic concept and flowchart of time series methods and AI-based approaches are briefly explained. Next, four metaheuristic optimization methods, namely PSO, CSA, ACS and BSA are implemented in learning process of GEP to form different optimized GEP approaches. In chapter four, the multi-objective feature selection is developed to extract the most influential subsets of input variables for long-term EEC modeling. Then the performance of optimized GEP is compared with other AI-based approaches. Further comparison is provided in this section to specify the most effective structure of input data sets. Moreover, EEC of ASEAN-5 countries are projected up to

2030 by applying different time series forecasting methods and their forecasts are compare with those obtained by proposed approach. Eventually, chapter five outlines the conclusion and the future works. A comprehensive list of reference is provided at the end of the thesis.

**CHAPTER 2: LITERATURE REVIEW**

**2.1     Introduction**

To address the difficulties involved in long-term energy consumption forecasting whereas, there is an implicit relationship among historical data, different techniques have been proposed to achieve a robust model with high accuracy.

In this framework, the techniques can be classified into statistical models including regression-based model and time series methods such as auto regressive integrated moving average (ARIMA) and gray model (GM); and artificial intelligence (AI) based approaches including artificial neural network (ANN), fuzzy method, support vector regression (SVR), knowledge-based expert system (KBES), metaheuristic techniques and genetic programming (GP) (Hernandez et al., 2014).

**2.2     Time Series Methods**

The ARIMA model is one of the most popular models for a time series forecasting when there is no missing sample within the time series and the time series is stationary (Yu, Zhu, & Zhang, 2012). Although ARIMA models are quite flexible as they can represent several different types of time series, namely pure auto regressive (AR), pure moving average (MA) and combined ARMA (AR and MA), their major limitation is the pre-assumed linear form of the model (Ardakani & Ardehali, 2014).

GM is a practical approach in time series forecasting due to its simplicity and ability to characterize an unknown system with incomplete information within the time series. The main principle of this approach to extract hidden information from incomplete data is to process the data indirectly through data mapping to the state space. In spite of this mechanism, the original GM is not suitable for long-term energy consumption forecasting due to disability of this approach to reflect the growth trends within different period into behavioral modeling of unknown systems (Hsu & Chen, 2003).

## 2.3 Artificial Intelligence-Based Approaches

In the past decade, the AI-based approaches are considered as enhanced alternatives to statistical models for energy consumption forecasting. AI-based approaches often guarantee a satisfying degree of estimation accuracy while independent and dependent variables faced too much fluctuation. Table 2.1 outlines the summary of AI-based approaches, which have been employed for long-term energy consumption forecasting.

ANN is the most widely used technique among the AI-based approaches, which has been applied in the field of energy management (Ekonomou, 2010; Kankal, Akpınar, Kömürcü, & Özşahin, 2011; Kialashaki & Reisel, 2014; Sözen, Arcaklioğlu, & Özkaymak, 2005). The capability of ANN to precisely learn, store, and recall information from experience, discover the relation between input and output variables, and extract various discriminators in complex environment makes this method especially attractive for long-term energy consumption forecasting.

SVR is another AI-based technique, which has been employed as a powerful predictive technique in energy consumption forecasting due to its ability to adapt and capture complex relationships in the input data (Hong, 2009). A significant advantage of SVR is that this method guarantees that the global minimum is found during the training phase, while ANN is trapped in local minima. Unlike ANN, SVR is less prone to over fitting due to independently of computational complexity to dimensionality of the input space. Moreover, SVR has a simpler geometric interpretation and it gives sparse solution (Ghelardoni, Ghio, & Anguita, 2013).

The fuzzy logic system successfully applied in (Elias & Hatziargyriou, 2009) for energy consumption forecasting. This approach is based on predefined rules (if-then) that lack the ability to learn and adapt them-self to new condition. In (Zahedi, Azizi, Bahadori, Elkamel, & Wan Alwi, 2013), the authors applied combination of fuzzy system and ANN

(Neuro-Fuzzy) to overcome this drawback. A specific approach in neuro-fuzzy development is the adaptive neuro-fuzzy inference system (ANFIS), which is considered to be a universal estimator capable for short, medium, and long-term energy consumption forecasting (Nadimi, Azadeh, Pazhoheshfar, & Saberi, 2010).

Furthermore, an effective practice to increase the forecasting accuracy and address the nonlinearity involved in long-term forecasting is integrating several techniques into a hybrid form. An integrated algorithm for forecasting energy consumption based on multi-layer perceptron (MLP) ANN, computer simulation and design of experiments is developed in (A. Azadeh, Ghaderi, & Sohrabkhani, 2008). The integration of fuzzy system and data mining approach is presented for energy consumption forecasting (A. Azadeh, Saberi, Ghaderi, Gitiforouz, & Ebrahimipour, 2008). Data mining approach is applied to extract the rules for constructing fuzzy system estimation in this study. A hybrid ANFIS and computer simulation is proposed to improve the accuracy of energy consumption forecasting (A. Azadeh, Saberi, Gitiforouz, & Saberi, 2009).

Despite the satisfactorily performance of ANN, SVR, ANFIS, and hybrid form of these methods for energy consumption forecasting, the main shortcoming still is black-box problem that they do not provide the knowledge of process for obtaining a solution. Hence, they are not capable of generating a definite prediction equation based on the input historical data (S. Roy, Ghosh, Das, & Banerjee, 2015).

In order to cope with the difficulties associated with energy consumption modeling, different forms of mathematical expressions optimized by metaheuristic methods have been proposed to use historical data for formulation the energy consumption.

Metaheuristic methods have been applied as efficient tools to provide realistic estimation models by optimizing the coefficients of predefined mathematical expressions.

These methods are characterized by stochastic nature, global search ability, and a large amount of implicit parallelism.

Different metaheuristic methods have been developed to provide more accurate and realistic estimation model through altering the way by which the weighting factors are determined. ant colony optimization (ACO) (Duran Toksarı, 2007), genetic algorithm approach (GA) (Canyurt & Ozturk, 2008), simulated annealing approach (SA) (Ozcelik & Hepbasli, 2006), particle swarm optimization (PSO) (Askarzadeh, 2014; Ünler, 2008), gravitational search algorithm (GSA) (Behrang, Assareh, Ghalambaz, Assari, & Noghrehabadi, 2011), artificial cooperative search (ACS) (Kaboli, Selvaraj, & Rahim, 2016), harmony search algorithm (HS) (Ceylan, Ceylan, Haldenbilen, & Baskan, 2008), artificial bee colony (ABC) (Kıran, Özceylan, Gündüz, & Paksoy, 2012b), charged system search (CSS) (Kaveh, Shamsapour, Sheikholeslami, & Mashhadian, 2012), imperialist competitive algorithm (ICA) (Nasab, Khezri, Khodamoradi, & Gargari, 2010) and cuckoo search algorithm (CSA) (Chang, Zhu, & Chen, 2015) are the metaheuristic optimization algorithms that have been satisfactorily applied in the field of energy (electricity, petroleum, oil, fossil fuels) consumption forecasting.

Lately, to promote the estimation accuracy combinatorial metaheuristic methods have been developed. These methods include a hybrid approach of ACO with PSO (Kıran, Özceylan, Gündüz, & Paksoy, 2012a), and a hybrid version of GA with PSO (Amjadi, Nezamabadi-Pour, & Farsangi, 2010; Assareh, Behrang, Assari, & Ghanbarzadeh, 2010). Furthermore, metaheuristic methods have been integrated with other techniques into a hybrid method to improve the performance of estimation methods. Recent hybrid solutions include ANN-TLBO (teaching learning based optimization) (Uzlu, Kankal, Akpınar, & Dede, 2014), ANN-PSO (Ardakani & Ardehali, 2014), ANN-GA (Ali Azadeh, Ghaderi, Tarverdian, & Saberi, 2007), SVR-DE (differential evolution) (J. Wang

et al., 2012), optimized GM (Hamzacebi & Es, 2014), ARIMA-PSO (Y. Wang, Wang, Zhao, & Dong, 2012) and Fuzzy-GA (Liu & Fang, 2013).

Although metaheuristic methods have been applied to provide realistic estimation models, the significant limitation still is that they require predefined knowledge about the structure of the existing relationship between the variables (fitness functions) which need to be ascertained beforehand (S. Roy et al., 2015). Furthermore, for all application of metaheuristic methods, the major deficiencies still are too many control parameters and quite sensitive to initial values of these parameters. Although the promising results are provided by combinatorial optimization methods, the proper integration point between the two metaheuristic methods is difficult to determine. Moreover, the inherent complexity of combinatorial optimization methods involves a non-negligible increase of the efforts for properly tuning the control parameters.

In order to cope with the limitations of the existing AI-based methods, genetic programming (GP) is employed to extract the relevant information of the corresponding problem and transform the derived information into a mathematical model. The capability of generating prediction equation without prior knowledge about the nature of the relationships between independent and dependent variables is a remarkable attribute of GP. Compared to metaheuristic methods, GP strategy provides a superior alternative for long-term energy consumption forecasting as it precludes the need to conform to the predefined fitness functions. Although the coupled GP with SA has been proposed in (Mostafavi, Mostafavi, Jaafari, & Hosseinpour, 2013) to enhance the efficiency of GP, still the bloat phenomenon (code growth problem) hampers the applicability of GP-based modeling system for energy consumption forecasting. The tendency of GP to complicate the model without any significant improvement in estimation accuracy is known as bloat phenomenon which slows the evolutionary search process, consumes computing

resources, and eventually impedes its ability and efficiency to discover better solutions (Whigham & Dick, 2010).

The gene expression programming (GEP) is an extension of GP (Rafieerad et al., 2016) , that incorporates advantages of GA into GP to avoid bloat. It takes the flexibility of GP and simplicity of GA hence it can be considered as an effective tool for long-term energy consumption forecasting. Although GEP has successfully modeled electricity consumption in (Mousavi, Mostafavi, & Hosseinpour, 2014), still more robust tool is required to formulate the energy consumption due to its inherent nonlinearity and complexity. Despite the flexible representation and efficient evolutionary process of GEP, it has difficulty discovering preferable and simplified function structure (Zhong, Ong, & Cai, 2015). Thus, a need has gradually emerged in contemporary meta-modelling paradigms to develop an approach for long-term energy consumption forecasting, which can combine the inherent capability of a GEP strategy in adapting information into a mathematical expression with the clarity of explicit closed form analytical representation of mathematical models optimized by metaheuristic methods. To bridge this gap, optimized GEP has been developed as an offshoot of the versatile evolutionary based meta-modelling techniques. A preliminary version of the optimized GEP approach was presented in (Zhong et al., 2015) to further improve the search performance, in which for each generation a metaheuristic optimization method is invoked on every chromosome in a population to tune up the constants of GEP. This approach was applied for mathematical modeling of benchmark problems and obtained results confirm the well performance of this method in comparison with other GP variants.

The optimized GEP is applied in this study as an enhanced approach to further improve forecasting precision. The mathematical model of the GEP can be manipulated in practical circumstances to provide a distinct advantage of a greater transparency and

simplicity in forecasting. This merit is provided by deploying a metaheuristic method over GEP models to determine the optimal weighting factors (best coefficients with minimum error).

Metaheuristic methods do not require precisely defined mathematical expressions and impose fewer mathematical requirements to obtain a highly accurate model. Hence, they have been employed as efficient tools to provide realistic estimation models by optimizing the weights of mathematical expressions developed by GEP. Several robust and efficient metaheuristic optimization methodologies including PSO, CSA, ACS and backtracking search optimization algorithm (BSA) which not only provide highly accurate results but also have relatively simple implementation procedure with fewest control parameters are applied in learning process of GEP to form three different variants of optimized GEP for long-term EEC forecasting.

**Table 2.1: Summary of studies on long-term energy consumption forecasting for various countries via AI-based methods**

| Methods | Forecasting for | Period | Model | Input variables | MAPE |
|---|---|---|---|---|---|
| **Fuzzy regression** | Energy demand of the Greek power system (Elias & Hatziargyriou, 2009) | 1987-2003 | - | Number of customers, Energy consumption, Temperatures | 0.70% |
| | Energy consumption of Iran(A. Azadeh, Saberi, & Seraj, 2010) | 1995-2005 | - | Previously observed values | 0.82% |
| **ANN** | Energy consumption of Turkey(Kankal et al., 2011) | 1980-2014 | - | GDP (gross domestic production), POP (population), IMP (import), EXP (export) | 1.22% |
| | Energy demand in the industrial sector of US (Kialashaki & Reisel, 2014) | 1980-2030 | - | GDP, Price of energy carriers | 0.57% |
| | Net energy consumption of Turkey(Sözen et al., 2005) | 1975-2003 | - | POP, IMP, EXP, Installed capacity, Gross generation | 2.14% |
| **ANN-MLP** | Energy consumption of Greek(Ekonomou, 2010) | 2005–2015 | - | GDP, Temperature, Installed capacity, Electricity consumption | 1.37% |
| **ANFIS** | Electricity consumption of G7 (Nadimi et al., 2010) | 2008-2015 | - | GDP,POP | 1.49% |
| **SVM** | Regional electric load of Taiwan(Hong, 2009) | 1981-2000 | - | Previously observed values | 1.29 |
| **KBES** | Load of fast developing utility(Kandil, El-Debeiky, & Hasanien, 2002) | 1981-2007 | - | Years, Temperature, Ramadan | 1.33% |
| **GP** | Electricity consumption of Turkey (Karabulut, Alkan, & Yilmaz, 2008) | 1994-2010 | - | Previous demand, Climate | 1.16% |

| Data mining | Electricity demand of Turkey (KÜÇÜKDENİZ, 2010) | 1980-2025 | - | Previously observed values | 3.25% |
|---|---|---|---|---|---|
| GEP | Electricity demand of Thailand (Mousavi et al., 2014) | 1986-2009 | Expression | GDP, POP, EXP, Stock index | 0.37% |
| ACO | Energy demand of Turkey (Duran Toksarı, 2007) | 1979-2025 | Linear, Quadratic | GDP, POP, IMP, EXP | 1.07% |
| GA | Fossil fuels demand of Turkey (Canyurt & Ozturk, 2008) | 1980-2020 | Quadratic | GDP, IMP, EXP | 2.09% |
| SA | Petroleum demand of Turkey (Ozcelik & Hepbasli, 2006) | 1990-2020 | Linear, Quadratic | GDP, Vehicle ownership | 2.2% |
| PSO | Energy demand of Turkey (Ünler, 2008) | 1979-2025 | Linear, Quadratic | GDP - POP - IMP - EXP | 0.83% |
| | Electricity demand of Iran (Askarzadeh, 2014) | 1982-2030 | Quadratic, Exponential | GDP, POP, IMP, EXP | 1.16% |
| GSA | Oil demand of Iran (Behrang et al., 2011) | 1981-2030 | Linear, Exponential | GDP, POP, IMP, EXP, Number of vehicles | 1.14% |
| ACS | Electric energy consumption of Iran (Kaboli et al., 2016) | 1992-2030 | Linear, Quadratic, Exponential, Logarithmic | GDP, POP, IMP, EXP, Stock index | 0.75% |
| HS | Transport energy demand of Turkey (Ceylan et al., 2008) | 1970-2025 | Linear, Quadratic, Exponential | GDP, Vehicle kilometers | 13.41% |
| ABC | Electricity demand of Turkey (Kıran et al., 2012b) | 1979-2025 | Linear, Quadratic | GDP, POP, IMP, EXP | 2.26% |
| CSS | Transport energy demand of Iran (Kaveh et al., 2012) | 1968-2030 | Linear, Exponential | GDP, Number of vehicles | 3.31% |
| ICA | Energy demand of Iran (Nasab et al., 2010) | 1986-2017 | Exponential, Quadratic | GDP, POP, EXP | 1.14% |

**Combinatorial methods**

| Methods | Forecasting for | Period | Model | Input variables | MAPE |
|---|---|---|---|---|---|
| PSO-ACO | Energy demand of Turkey (Kıran et al., 2012a) | 1979-2025 | Linear, Quadratic | GDP, POP, IMP, EXP | 1.03% |
| PSO - GA | Electricity demand of Iran (Amjadi et al., 2010) | 1980-2006 | Linear, Nonlinear | GDP, POP, Number of customers, Electricity price | 0.98% |
| | Oil demand of Iran (Assareh et al., 2010) | 1981-2030 | Linear, Exponential | GDP, POP, IMP, EXP | 1.36% |
| Fuzzy-data mining | Electricity demand of Iran (A. Azadeh, Saberi, et al., 2008) | 1995-2005 | - | Previously observed values | 2% |
| ANN-TLBO | Energy demand of Turkey (Uzlu et al., 2014) | 1980-2020 | - | GDP, POP, IMP, EXP | 1.50% |
| ANN-PSO | Electricity demand of Iran and US (Ardakani & Ardehali, 2014) | 1967-2030 | - | GDP, POP, IMP, EXP | 1.51% |
| ANN-GA | Electricity consumption of Iran (Ali Azadeh et al., 2007) | 1981-2008 | Logarithmic, Linear | Price, Number of customers, Value added | 3.68% |
| Simulated-ANN | Electrical energy consumption of Iran(A. Azadeh, Ghaderi, et al., 2008) | 1994-2005 | - | Previously observed values | 1.8% |
| GRNN-FOA | Power load of Beijing city and China(H.-z. Li, Guo, Li, & Sun, 2013) | 1978-2012 | - | Previously observed values | 1.15% |
| Simulation-ANFIS | Electricity consumption of Iran(A. Azadeh et al., 2009) | 1994-2005 | - | Previously observed values | 1.55% |

| | | | | | |
|---|---|---|---|---|---|
| **SVR-DE** | Electricity demand of Beijing (J. Wang et al., 2012) | 1987-2008 | - | Previously observed values | 1.1% |
| **LS-SVM-FOA** | Electricity consumption of China(H. Li, Guo, Zhao, Su, & Wang, 2012) | 1998-2011 | - | Previously observed values | 1.03% |
| **Optimized GM** | Electricity consumption of Turkey (Hamzacebi & Es, 2014) | 1945-2025 | - | Previously observed values | 3.28% |
| **ARIMA- PSO** | Electricity demand of china (Y. Wang et al., 2012) | 2006-2010 | - | Previously observed values | 2.19% |
| **Fuzzy-GA** | Electrical load of Shanghai (Liu & Fang, 2013) | 1990-2010 | - | Previously observed values | 7.45% |
| **GP-SA** | Electricity demand of Thailand (Mostafavi et al., 2013) | 1986-2009 | Nonlinear, Expression | GDP, POP, EXP, Stock index | 0.5% |

\* Mean absolute percentage errors (MAPEs) are on testing period of each model.

**CHAPTER 3: RESEARCH METHODOLOGY**

**3.1      Introduction**

All applied methods in this study are presented in this chapter. First, the time series forecasting methods namely; autoregressive integrated moving average (ARIMA) and first-order single-variable grey model (GM (1, 1)) are briefly descried. Then, the basic concept and flow chart of artificial intelligence-based techniques such as artificial neural network (ANN), support vector regression (SVR), adaptive neuro-fuzzy inference system (ANFIS), and gene expression programming (GEP) are explained. Next, four different robust and efficient metaheuristic optimization methods, namely; particle swarm optimization (PSO), cuckoo search algorithm (CSA), artificial cooperative search (ACS) algorithm and backtracking search algorithm (BSA) are applied in the second learning process of GEP to form optimized GEP approach. Eventually, multi-objective backtracking search algorithm (MOBSA) is developed and its procedure is described.

**3.2      Time Series Forecasting Methods**

**3.2.1    Autoregressive Integrated Moving Average**

Among the statistical forecasting techniques, autoregressive integrated moving average (ARIMA) is a high-precision non-structural method for time series forecasting when there is no missing data  within the time series. Particularly in time series analysis, an ARIMA model is considered as a "filter" that tries to separate the time series from the noise, and the time series is then extrapolated either to predict future points in the series or to better understand the data. Non-seasonal ARIMA models are generally denoted as ARIMA (p,d,q), where parameters **q**, **d**, and **p** are non-negative integers. The parameter **p** represents the number of time lags for the autoregressive model (i.e. AR (p)), the parameter **d** denotes the number of differences (I) that are needed to make the series stationary and **q** is the order of the moving average  (i.e. MA (q))  part (Yuan, Liu, & Fang, 2016).

The autoregressive (AR) part of the ARIMA model with order p is of the form:

$$Y_t = \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \cdots + \emptyset_p Y_{t-p} + e_t + c \qquad (3.1)$$

where $\emptyset_1, \emptyset_2, \ldots, \emptyset_p$ are the parameters of the model and the independent variables $Y_{t-1}, Y_{t-2}\ldots, Y_{t-p}$ are time-lagged values of the forecast variable. As the forecasts are only dependent on observed values in the previous time periods, this model named auto regression.

The moving average (MA) part of the ARIMA model consists of the past errors as the explanatory variable. A moving average model with order q is of the form:

$$Y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + c \qquad (3.2)$$

where $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model and $e_t, e_{t-1}, e_{t-2}, \ldots, e_{t-q}$ are white noise error terms.

An autoregressive (AR (p)) model is coupled with a moving average (MA (q)) model to form an ARMA (p,q) (autoregressive moving average) model for stationary series. For non-stationary series, differencing is added to ARMA model. Differencing is a method to stabilize the mean of the series, eliminate seasonality, and consequently make the series stationary. The first difference between consecutive observations is calculated according to Eq. (3.3) to mathematically difference the data.

$$\Delta Y_t = Y_t - Y_{t-1} \qquad (3.3)$$

Finally, the general form of ARIMA model is formulated in Eq. (3.4) which requires at least $p+d$ presamples to initialize the time series.

$$\Delta_d Y_t = c + \emptyset_1 \Delta_d Y_{t-1} + \emptyset_2 \Delta_d Y_{t-2} + \cdots + \emptyset_p \Delta_d Y_{t-p} + e_t - \theta_1 e_{t-1}$$
$$- \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \qquad (3.4)$$

### 3.2.2 First-Order Single-Variable Grey Model

Grey model (GM) is a practical approach in time series forecasting due to its simplicity and ability to generate, excavate, and extract useful information from a gray system. A system with limited number of available observations and partially known information is considered as a gray, hazy, or uncertain system. The main principle of this method to predict the behavior of gray systems is to process the data indirectly through data mapping to the state space.

In the context of grey systems theory, GM (p, q) represents a grey model, where q denotes the number of variables and p denotes the order of the difference equation. Although, various types of grey prediction models can be formed, the GM (1, 1) as a first-order single-variable grey model is superior to other grey prediction models due to its high computational efficiency (Zhao & Guo, 2016).

GM (1, 1) is a time series-forecasting model, which does not need any prior knowledge such as probability distribution of the input data, and it can only be used in positive data sequences with minimum number of four observations. Since all the primitive data points are positive in this study, GM (1, 1) can be used to forecast the future values. GM (1, 1) model is formed to forecast the future values according to the five sequential steps:

- i-   First-order accumulated generating operation (1-AGO)
- ii-  Build GM (1, 1) model
- iii- Last square estimation method
- iv-  Whitenization process (first-order grey differential equation)
- v-   Inverse accumulated generating operation (IAGO)

The modelling procedure of GM (1, 1) is demonstrated in Figure 3.1. Firstly, the first-order accumulating generation operator (1-AGO) is applied on the non-negative time

sequence of primitive data obtained from the system in order to smooth the randomness and form the GM (1, 1) model. Then, the first-order grey differential equation is solved to obtain the H-step ahead predicted values of the primitive data. Finally, the inverse accumulating generation operator (IAGO) is applied to find the predicted values.

Eq. (3.5) denotes the initial time sequence of primitive data obtained from the system.

$$x^{(0)}(t) = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\}, \quad n \geq 4 \qquad (3.5)$$

where $n$ is the number of the observations and $x^{(0)}(t)$ is a non-negative sequence.

The first-order AGO is applied on a non-negative sequence $x^{(0)}(t)$, so the following monotonically increasing sequence $x^{(1)}(t)$ is obtained.

$$x^{(1)}(t) = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \ldots, x^{(1)}(n)\}, \quad n \geq 4 \qquad (3.6)$$

where

$$x^{(1)}(t) = \sum_{i=1}^{t} x^{(0)}(i), \quad t = 1,2,3, \ldots, n \qquad (3.7)$$

According to the Eq. (3.8), the mean sequence of consecutive neighbors in $x^{(1)}(t)$ denoted by $z^{(1)}(t)$ is obtained.

$$z^{(1)}(t) = \{z^{(1)}(1), z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\} \qquad (3.8)$$

where

$$z^{(1)}(t) = \frac{1}{2}\left(x^{(1)}(t) + x^{(1)}(t-1)\right) \qquad (3.9)$$

So, the basic form of GM (1, 1) model can be formulated according to Eq. (3.10).

$$x^{(0)}(t) + az^{(1)}(t) = b \qquad (3.10)$$

The last square estimation method is applied to obtain the GM (1, 1) model coefficients $\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix}$ as follow:

$$\hat{a} = [a, b]^T = (B^T B)^{-1} B^T Y \qquad (3.11)$$

where

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \qquad (3.12)$$

The whitening equation of GM (1, 1) model as formulated by Eq. (3.12) is solved to obtain the solution at time (t+1) according to Eq. (3.13).

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \qquad (3.13)$$

$$\hat{x}^{(1)}(t + 1) = \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-at} + \frac{b}{a} \qquad (3.14)$$

To obtain the predicted value of the primitive data at time (t + 1), the IAGO is applied to establish the following grey forecasting sequence.

$$\hat{x}^{(0)}(t) = (1 - e^a) \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-a(t-1)} \qquad (3.15)$$

where $\hat{x}^{(0)}(1) = x^{(0)}(1)$

Consequently, the predicted value of the primitive data at time (t + H) is obtained according to the Eq. (3.16).

$$\hat{x}^{(0)}(t + H) = (1 - e^a) \left( x^{(0)}(1) - \frac{b}{a} \right) e^{-a(t+H-1)} \qquad (3.16)$$

where $\hat{x}^{(0)}(t + H)$ denotes H-step ahead predicted values.

The original non-negative data sequence
$$x^{(0)}(t) = \{x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)\}$$

**1-AGO**

The monotonic increasing sequence
$$x^{(1)}(t) = \{x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \ldots, x^{(1)}(n)\}$$

**build GM (1, 1)**

Basic form of GM (1, 1)
$$x^{(0)}(t) + az^{(1)}(t) = b$$
where $z^{(1)}(t) = \frac{1}{2}\left(x^{(1)}(t) + x^{(1)}(t-1)\right)$

**last square estimation method**

The parameter vector
$$\hat{a} = [a, b]^T = (B^T B)^{-1} B^T Y$$
where $Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$, $B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$

**whitenization process**

The albinism differential equation
$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b$$

**solve**

The solution at time (t+1)
$$\hat{x}^{(1)}(t+1) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-at} + \frac{b}{a}$$

**IAGO**

The forecasting sequence
$$\hat{x}^{(0)}(1) = x^{(0)}(1)$$
$$\hat{x}^{(0)}(t) = (1 - e^a)\left(x^{(0)}(1) - \frac{b}{a}\right)e^{-a(t-1)}$$

**Figure 3.1: Modelling procedure of GM (1, 1)**

44

**3.3      Artificial Intelligence-Based Techniques**

**3.3.1      Artificial Neural Network**

Artificial neural networks (ANNs) are artificial networks inspired by nervous system which are considered as human attempts to understand what goes on in the biological neural networks (in particular the brain). ANNs simulate the learning process provided by nervous system with the hope of capturing the power of biological neural network, collectively performs tasks that even the supercomputers with high-level computational capacity have not been able to process (Landeras, López, Kisi, & Shiri, 2012).

The ANN is analogous to biological neural network, which consists of a highly interconnected network with very simple processors known as neurons. The neurons are linked by weighted connections that communicate to each other by sending the signals from one neuron to other neurons while the strength of weighted connections expresses the importance of each neuron input. Each neurons is associated with a transfer function, which describes how the weighted sum of a neuron input signal is converted to an output signal.

The main characteristic feature of ANNs is that these artificial networks acquire accumulated experience within learning process and respond to new conditions based on the knowledge gained within the learning process. In ANNs, the learning process is provided through repeated adjustments of numerical weights, thus the weighted connections are considered as basic means of long-term memory in these networks.

Although various topologies of neural networks (NNs), have evolved based on the training paradigm neuron arrangement and neuron connections. Among the various types of NNs, multilayer perceptron (MLP) and the radial basis function (RBF) network have been the most useful types of NN in different applications. The main differences between these two types of NN reside in the activation functions of the hidden layer. The activation

function belongs to the Gaussian family in RBF network whereas, the linear, logistic sigmoid and bipolar sigmoid (hyperbolic tangent) activation functions are used in MLP (Kankal et al., 2011).

Generally, there is a trade-off between higher robustness provides by RBF network and higher accuracy gains by MLP. Due to the non-linear nature of RBF network, it brings much more robustness to adversarial noise. Instead, MLP is an acronym for deep learning in NNs as it has multiple hidden layers to provide higher accuracy.

MLP as a feedforward NN has the generalizing ability to approximate essentially any function with high degree of accuracy, so it is considered as universal approximators.

As shown in Figure 3.2, the MLP architecture is composed of three layers:

i-      Input layer, where the data are introduced into the NN (source neurons)

ii-     Hidden layer(s) where the data are processed (computational neurons)

iii-    Output layer where the results of given inputs are obtained.
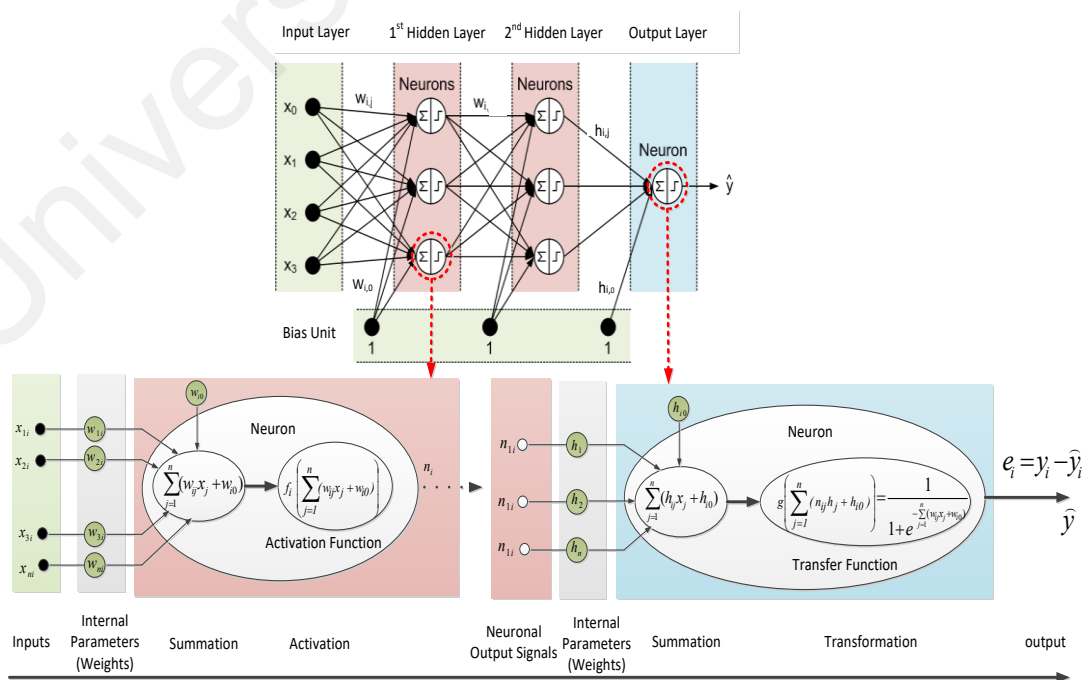


**Figure 3.2: The MLP architecture**

Each of these layers has several processing units, and each unit is fully interconnected with weighted connections to units in the subsequent layer. Each layer contains a number of nodes. Every input is multiplied by each of the nodes using its inter-connection weight. The output of each neuron is obtained by passing the sum of the product through an activation function, while the bias input (memory) is connected to each neuron to stabilize the origin of activation function for better learning.

MLP networks are usually applied to perform supervised learning tasks, which involve iterative training methods to adjust the connection weights within the network. Generally, several passes are required to attain a desired level of estimation accuracy. Adjustment of the correction weights is carried out using the standard error back-propagation algorithm, which minimizes the total error with the gradient decent method (Raza & Khosravi, 2015). Back-propagation is a systematic method for training MLP networks as its schematic diagram is briefly described in Table 3.1.

**Table 3.1: General structure of back-propagation algorithm**

1.  Initialize weights
**Repeat**
   2.  Apply a sample
      **Calculate the output of each neuron:**
         Summation
         Addition of bias
         Activation
      **end**
   3.  Calculate the output errors
   4.  Backpropagation of the error
   5.  Update weights and biases
**Until** termination criteria are satisfied;

- **Initialize weights**

The initial weights or numerical estimates of the connection strength between all neurons ($w_{ij}$) are assigned randomly. Furthermore the initial value of activation threshold ($w_{0j}$) is also assigned to each neuron randomly. This activation threshold is analogous to an independent term of the linear combination of the outputs from the previous neurons, and it is considered as a weight assigned to a fictitious neuron known as bias unit with an output value of 1. Therefore, the rule of the bias input (memory) is to shift the origin of activation function for better learning.

- **Apply a sample**

Apply the input vector $\{X_1, X_2, X_3, ..., X_k\}$ having desired output vector $\{y_1, y_2, y_3, ..., y_m\}$.

- **Feed-forward computation**

Starting from the first hidden layer and propagating toward the output layer. Each input unit ($x_i$) assigns initial weight ($w_i$) and broadcast this weight to all neuron in above layer (first hidden layer).

i.  Each neuron in the first hidden layer ($n_j$) sums its input weights

$$n\_in_j = \sum_{i=1}^{k} w_{ij}X_i + w_{0j} \tag{3.17}$$

ii.  The activation function ($f$) process the output signal of each neuron ($n_j$) in hidden layers

$$n_j = f(n\_in_j) \tag{3.18}$$

Generally the activation functions are one of the linear, logistic sigmoid and bipolar sigmoid (hyperbolic tangent) activation functions.

The linear activation function is:

$$n_j = f(n\_in_j) = n\_in_j \tag{3.19}$$

In a MLP network if the neurons have linear activation functions the capabilities of the network is no better than a single layer network with linear activation function. Thus, nonlinear activation functions (sigmoid functions) are used, which usually limit the output signal of each neuron to the values between two asymptotes.

The logistic sigmoid function as formulated below is the most widely used activation function in MLP.

$$n_j = f(n\_in_j) = \frac{1}{1 + e^{-(n\_in_j)}} \tag{3.20}$$

The hyperbolic tangent function as formulated by Eq. (3.21) is another sigmoidal function is used as activation function for neurons in hidden layer of MLP networks. The hyperbolic tangent function is closely related to the bipolar sigmoid function.

$$n_j = f(n\_in_j) = \frac{1 - e^{-(n\_in_j)}}{1 + e^{-(n\_in_j)}} = \frac{1 - e^{-(n\_in_j)}}{1 + e^{-(n\_in_j)}} \tag{3.21}$$

iii.  The output signal $j^{th}$ neuron of total (N) neuron in hidden layer (L) denoted by ($n_{Lj}$) is transferred to next hidden layer as follow:

$$n^L_j = f\left( \sum_{i=1}^{N} w_{ij} n_i^{L-1} + w_{0j} \right) \tag{3.22}$$

iv.  The output signal of each neurons in the last hidden layer are propagated toward the output layer as follow:

$$\hat{y}_j = g\left( \sum_{i=1}^{m} h_{ij} n_i + h_{0j} \right) \tag{3.23}$$

where ($g$), ($h_{ij}$) and ($h_{0j}$) are the activation function of output layer known as transfer function, the connection strength (weight) between $i^{th}$ neuron in last hidden layer and $j^{th}$ neuron in output layer, and the weight assigned to the bias unit of $j^{th}$ neuron in output layer respectively.

- **Calculate the output errors**

The error information term of each neuron in output layer is computed as follows:

$$\delta_j = \left(y_j - \hat{y}_j\right) g' \left(\sum_{i=1}^{m} h_{ij} n_i + h_{0j}\right) \tag{3.24}$$

- **Backpropagation of the error**

Propagate the error backward to the input layer through each hidden layer using the error information term.

The backward weight correction term from output layer to last hidden layer and its bias correction term are computed as follow:

$$\Delta h_{ij} = \alpha \delta_j n_i \tag{3.25}$$

$$\Delta h_{0j} = \alpha \delta_j \tag{3.26}$$

where $\alpha$ is learning rate:

The error information term of each neuron in last hidden layer is calculated from multiplying the summation of its backward weights correction by derivative of its activation function as follow:

$$\Delta_j = \left(\sum_{i=1}^{m} \delta_j h_{ij}\right) \times f' \left(\sum_{i=1}^{N} w_{ij} n_i + w_{0j}\right) \tag{3.27}$$

The backward weight correction term from the hidden layer ($L$) is sent to its hidden layer below ($L$-$1$) and its bias correction term are computed as follow

$$\Delta w_{ij} = \alpha \Delta_j n_j^{L-1} \tag{3.28}$$

$$\Delta w_{0j} = \alpha \Delta_j \tag{3.29}$$

The error information term of each neuron in hidden layer is calculated as follow

$$\Delta_{\delta j} = \left( \sum_{i=1}^{N} \Delta_j w_{ij} \right) \times f' \left( \sum_{i=1}^{N} w_{ij} n_j^{L-1} + w_{0j} \right) \tag{3.30}$$

The backward weight correction term from first hidden layer is sent to the input layer and its bias correction term are computed as follow:

$$\Delta_{\delta} w_{ij} = \alpha \Delta_{\delta j} X_i \tag{3.31}$$

$$\Delta_{\delta} w_{0j} = \alpha \Delta_{\delta j} \tag{3.32}$$

- **Update weights and biases**

Each neuron in the output layer updates its bias and weights as follow:

$$h_{ij}(t+1) = h_{ij}(t) + \Delta h_{ij} \tag{3.33}$$

$$h_{0j}(t+1) = h_{ij}(t) + \Delta h_{0j} \tag{3.34}$$

Each neuron in hidden layer updates its bias and weights as follow:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij} \tag{3.35}$$

$$w_{0j}(t+1) = w_{ij}(t) + \Delta w_{0j} \tag{3.36}$$

Each neuron in first hidden layer updates its bias and weights as follow:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta_{\delta} w_{ij} \tag{3.37}$$

$$w_{0j}(t+1) = w_{ij}(t) + \Delta_{\delta} w_{0j} \tag{3.38}$$

Eventually, Table 3.2 describes the Pseudo coding of back-propagation algorithm.

**Table 3.2: Pseudocode of back-propagation algorithm**

---

**Data**: X, y, Max epoch, Number of hidden layers, Number of neurons in hidden layers, Transfer function

**Result**: $\hat{y}$

**// Initialization**
1. The initial weights and biases are assigned randomly
2. **for** epoch ← 1 **to** Max epoch **do**
3.     **for** every pattern in the training set **do**
4.         present the pattern to the network
        **// Propagated the input forward through the network**
5.         **for** each layer in the network **do**
6.             **for** every node in each layer **do**
7.                 Calculate the weight sum of the inputs to the node
8.                 Add the threshold to the sum
9.                 Calculate the activation for the node
10.             **end**
11.         **end**
        **// Propagate the errors backward through the network**
12.         **for** every node in the output layer **do**
13.             calculate the error signal
14.         **end**
15.         **for** all hidden layers **do**
16.             **for** every neuron in each layer **do**
17.                 Calculate the node's signal error
18.                 Update each node's weight in the network
19.             **end**
20.         **end**
        **// Calculate Global Error**
21.         Calculate the Error Function
22.     **end**
23. **end**

---

### 3.3.2 Support Vector Regression

Kernel-based techniques e.g., Kernel principal component analysis (PCA), Kernel Fisher discriminant analysis (KFD), Bayes point machines, Gaussian processes and SVMs (support vector machines) represent a major development in machine learning algorithms. Kernel-based techniques map the data into a higher dimensional feature spaces in the hope that in the higher dimensional spaces the data are linearly separated or have better structure. SVMs as an extension to nonlinear model of the generalized portrait algorithm are the best-known member of Kernel-based techniques, which is able to either classify the input data or capture complex relationships in the input data. SVM that deal with function approximation and forecasting is just termed as SVR and SVM that deal with classification problems is just termed as support vector classification (SVC). With only a few minor modification, SVC is converted to SVR. So, SVM can be promoted to form SVR as a powerful function approximation technique based on statistical learning theory (Bian, Han, Du, Jaubert, & Li, 2016).

Considering the given data set as follow:

$$G = \{(x_k, y_k), \ k = 1,2,3,\dots,N\} \tag{3.31}$$

Where $x_k \in R^n$ denotes the $k^{th}$ element in n -dimensional input vector, $y_k \in R$ carrying the observed response values at time step $k$ with total number of $N$ observation. The linear SVR estimation function is expressed as follows:

$$\widehat{y_k} = f(x_k) = \langle W, x_k \rangle + b \tag{3.32}$$

where $b$ and $W$ are the intercept and weight vector of the regression function, respectively, $\widehat{y_k}$ is the estimated output of the model at time step $k$ and $\langle W, x_k \rangle$ denotes the vector inner product of the predictors.

In order to allow nonlinear modeling, SVR method uses Kernel trick by applying the Kernel function $\psi\ (x_k)$ that is s a non-linear mapping from the input space to a high dimensional feature space. So, the SVR estimation function is rewritten as follow:

$$\widehat{y_k} = f(x_k) = \langle W, \psi(x_k) \rangle + b \qquad (3.33)$$

In SVR method the attempt is to find a function that deviates from $y_K$ by a value no greater than $\varepsilon$ for every data point ($k$ =1, 2... N), while $W$ is as flat as possible. Thus, for better generalization performance and ensure that $W$ is as flat as possible, the norm of vector $W$ that measures the flatness of the function as formulated by Eq. (3.34) needs to be minimized:

$$minimize\ J(W) = \frac{||W||^2}{2} \qquad (3.34)$$

subject to all residuals having a value less than epsilon:

$$\forall\ k: |y_k - \widehat{y_k}| \leq \varepsilon \qquad (3.35)$$

The SVR is insensitive to small errors as it penalizes the errors that are greater than ε. This task is accomplished in SVR through use of ε-insensitive loss function as described by Eq. (3.36).

$$Loss_{\varepsilon(y_k, \widehat{y_k})} = \begin{cases} 0, & if\ |y_k - \widehat{y_k}| \leq \varepsilon \\ |y_k - \widehat{y_k}| - \varepsilon, & \text{otherwise} \end{cases} \qquad (3.36)$$

According to ε-insensitive loss function, if the forecasted value is within the ε– insensitive tube the calculated loss is equal to zero. Considering the ε-insensitive loss function leads to the objective function known as the regularized risk function. Then the task in SVR is to estimate the threshold value (*b*) and weight vector (*W*) that minimizes the following regularized risk function:

$$minimize \ R(C) = \frac{C}{N} \sum_{K=1}^{N} Loss_{\varepsilon(y_k, \widehat{y_k})} + \frac{\|W\|^2}{2} \qquad (3.37)$$

where the parameter $C$ is a positive constant regularization parameter, which determines the trade-off between the model flatness and the penalty imposed on forecasted values that lie outside the ε– insensitive tube. In SVR, since the parameter ε specifies the degree of tolerable errors (accuracy), the regularization parameter ($C$) specifies the balance between accuracy and generalization ability (Hong, 2010).

In this stage, it is possible that no SVR estimation function finds to satisfy the constraints for every data point ($k =1, \ 2... \ N$), so the feasible constraints are not guaranteed. To deal with other infeasible constraints, two slack variables ($\xi_k, \xi_k^*$) which represent the distance from actual values to the corresponding boundary values of ε– insensitive tube for each data point as shown in Figure 3.3.
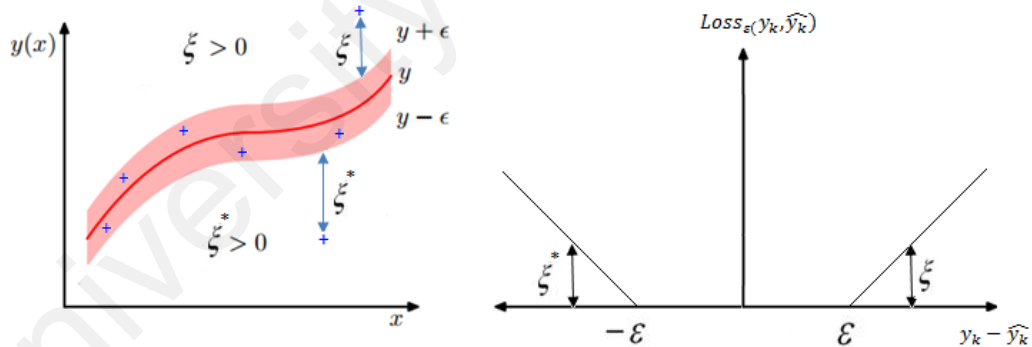


**Figure 3.3: The graphical representation of ε-insensitive loss function in SVR**

Although this approach is similar to the soft margin concept in SVC, the main difference comes in the slack variables used. SVR involves assigning two slack variables ($\xi_k, \xi_k^*$) for each data point, whereas in SVC for each data point one slack variable ($\xi_k$) is assigned. Including two slack variables, leads to the primal formula, which needs to be optimized as follow:

$$minimize\ R(W,\ _{\xi_k,\xi_k{}^*}) = \frac{C}{N}\sum_{K=1}^{N}(_{\xi_k+\xi_k{}^*}) + \frac{\|W\|^2}{2} \qquad (3.38)$$

S.t.

$$\forall\ k: y_k - \widehat{y_k} \leq \varepsilon + \ _{\xi_k} \qquad (3.39)$$

$$\forall\ k: \widehat{y_k} - y_k \leq \varepsilon + \ _{\xi^*{}_k} \qquad (3.40)$$

$$\forall\ k: _{\xi_k,\xi^*{}_k} \geq\ \ 0 \qquad (3.41)$$

To Improve the performance accuracy of SVR, the parameter $\varepsilon$ are replaced by a parameter ($v$), which is used to control the number of support vectors and training errors within the range of [0, 1]. After replacing ($\varepsilon$) by ($v$), the primal formula is transformed into

$$minimize\ R(W,\ _{\xi_k,\xi_k{}^*},\varepsilon) = C((v.\varepsilon) + (\frac{1}{N}\sum_{K=1}^{N}(_{\xi_k+\xi_k{}^*}))) + \frac{\|W\|^2}{2} \qquad (3.42)$$

S.t.

$$\forall\ k: y_k - \widehat{y_k} \leq \varepsilon + \ _{\xi_k} \qquad (3.43)$$

$$\forall\ k: \widehat{y_k} - y_k \leq \varepsilon + \ _{\xi^*{}_k} \qquad (3.44)$$

$$\forall\ k: _{\xi_k,\xi^*{}_k} \geq\ 0 \qquad (3.45)$$

To solve this constrained optimization problem (minimizing the primal formula) the primal Lagrangian form as formulated by Eq. (3.46) is used.

$$L(W,b,\ _{\xi_k,\xi_k{}^*},\alpha_k,\alpha_k^*,\beta_k,\beta_K^*,\eta) = \frac{\|W\|^2}{2} + C.v.\varepsilon + \frac{C}{N}\sum_{K=1}^{N}(_{\xi_k+\xi_k{}^*}) - \eta.\varepsilon -$$
$$\sum_{K=1}^{N}(\beta_k\xi_k + \beta_k^*\xi_k{}^*) - \sum_{K=1}^{N}\alpha_k(\xi_k + y_k - \widehat{y_k} + \varepsilon) - \sum_{K=1}^{N}\alpha_k^*(\xi_k{}^* + \widehat{y_k} - y_k + \varepsilon) \qquad (3.46)$$

where $\alpha_k, \alpha_k^*, \beta_k, \beta_K^*\ and\ \eta$ are Lagrangian multipliers.

Then, the saddle point of primal Lagrangian form that minimize over the primal variables ($_{\xi_k,\xi_k{}^*}$), ($b$) and ($W$) and, maximize over the Lagrangian multipliers ($\alpha_k, \alpha_k^*$), ($\beta_k, \beta_K^*$) $and$ ($\eta$) is fined. Thus, the derivatives of Lagrangian form with respect

to the primal variables are set to zero, which yields the following Karush-Kuhn-Tucker (KKT) complementarity conditions:

$$\frac{\delta L}{\delta W} = W - \sum_{K=1}^{N}(\alpha_k^* - \alpha_k)\psi(x_k) = 0 \tag{3.47}$$

$$\frac{\delta L}{\delta W} = C.\nu - \sum_{K=1}^{N}(\alpha_k^* + \alpha_k) - \eta = 0 \tag{3.48}$$

$$\frac{\delta L}{\delta W} = \sum_{K=1}^{N}(\alpha_k^* - \alpha_k) = 0 \tag{3.49}$$

$$\frac{\delta L}{\delta W} = \frac{C}{N} - \alpha_k^* - \beta_K^* = 0 \qquad k = 1,2,\dots,N \tag{3.50}$$

While the problem is convex and satisfies a constraint, the value of the optimal solution to the primal problem is calculated by the solution of the dual Lagrangian, in spite of the fact that the optimal values of the primal and dual Lagrangian should be equal, and their difference are known as duality gap. To obtain the dual formula, a Lagrangian function from the primal function by introducing two non-negative Lagrange multipliers $(\alpha_k, \alpha_k^*)$ for each observation $(x_k)$. Then, the dual Lagrangian is obtained by substituting Eqs. (3.47) – (3.50) into Eq. (3.46). This leads to the dual formula, which should be maximized as formulated below:

$$maximize \; L(\alpha) = \sum_{k=1}^{N}(\alpha_k - \alpha_k^*)y_k - \frac{1}{2}\sum_{K=1}^{N}\sum_{j=1}^{N}(\alpha_k - \alpha_j^*)(\alpha_k - \alpha_j^*)K_{kj} \tag{3.51}$$

subject to the constraints

$$\sum_{k=1}^{N}(\alpha_k - \alpha_k^*) = 0 \tag{3.52}$$

$$\sum_{k=1}^{N}(\alpha_k + \alpha_k^*) \leq C.\nu \tag{3.53}$$

$$\forall \, k: 0 \leq \alpha_k, \alpha_k^* \; \leq \frac{C}{N} \tag{3.51}$$

Where $K_{kj} = k(x_k,x_j) = \psi(x_k)^T \psi(x_j)$ is the inner product of two vectors $x_k$ and $x_j$ in the feature space $\psi(x_k)$ $and$ $\psi(x_j)$ respectively, known as the Kernel function which needs to meet Mercer's condition. The Mercer's theorem is positive semi-definite, meaning that Kernel matrix has only non-negative Eigen values. By applying positive definite Kernel, the optimization problem is converted to the convex optimization problem thus the unique solution is insured. In general, there are number of Kernels that can be used in SVM models. The most widely used Kernel functions are linear function, polynomial function, Gaussian RBF (radial basis function), exponential RBF, and sigmoid function.

- **Linear Kernel function**

The simplest Kernel function is the Linear Kernel that defined by the inner product $<x_k,x_j>$ as given below:

$$K_{kj} = k(x_k,x_j) = x_k{}^T x_j + c_0 \tag{3.55}$$

where, $c_0$ is an optional constant value.

- **Polynomial Kernel function**

The polynomial kernel is a non-stationary kernel function. Polynomial kernel function as formulated by Eq. (3.56) is well suited for problems where all the input data is normalized.

$$K_{kj} = k(x_k,x_j) = (ax_k{}^T x_j + c_0)^d \tag{3.56}$$

where, constant term $c_0$, the slope $a$, and the polynomial degree $d$ are the adjustable parameters for this Kernel function.

- **Gaussian RBF (radial basis function)**

The Gaussian RBF is formulated as follow:

$$K_{kj} = k(x_k, x_j) = \exp(-\frac{\|x_k - x_j\|^2}{2\gamma^2}) \qquad (3.57)$$

where the $\gamma$ is constant parameter, alternatively Gaussian RBF can be implemented as:

$$K_{kj} = k(x_k, x_j) = \exp(-\partial \|x_k - x_j\|^2) \qquad (3.58)$$

where, $\partial$ is the adjustable parameter that plays a major role in the performance of the Gaussian RBF. If this adjustable parameter overestimated, the Gaussian RBF behaves linearly and the higher-dimensional projection starts to lose its non-linear power. If this parameter underestimated, the Gaussian RBF starts to lose its regularization performance and the decision boundary become highly sensitive to the noise in training data set.

- **Exponential RBF**

The exponential kernel function as formulated by Eq. (3.59) is closely similar to the Gaussian RBF, while the square of the norm left out.

$$K_{kj} = k(x_k, x_j) = \exp(-\partial \|x_k - x_j\|) \qquad (3.59)$$

- **Hyperbolic Tangent Kernel function**

The Hyperbolic Tangent Kernel function is a sigmoid function comes from the ANNs field. Since, in artificial neurons the bipolar sigmoid function is used as an activation function, the Hyperbolic Tangent Kernel function is also known as MLP Kernel function. Generally, SVMs models are closely related to ANNs, and a SVM model using a MLP Kernel function is equivalent to a two-layer, perceptron ANN. Using MLP Kernel function in SVM is an alternative training method for ANNs, in which rather than by solving a non-convex, unconstrained optimization problem as in ANNs, in SVMs the

unknown parameters are obtained by solving a quadratic programming problem with linear constraints. Additionally, despite being only conditionally positive definite, MLP Kernel function as formulated by Eq. (3.60) shows acceptable performance in practice.

$$K_{kj} = k(x_k, x_j) = \tanh(a x_k^T x_j + c_0) \tag{3.60}$$

where, $c_0$ and $a$ are the adjustable parameters for this Kernel function. In general, a common value for tuning $a$ is (data dimension)$^{-1}$.

According to the nonlinear mapping of Kernel functions in the feature space, the regression parameter $W$ no longer need to be calculated explicitly. The maximization problem formulated by Eq. (3.51) is considered as a quadratic programming (QP) problem. By virtue of a quadratic programming problem, the global solution of two non-negative Lagrange multipliers ($\alpha_k, \alpha^*_k$) for each data set (k=1,2,…N) is guaranteed. The data point that vanishes the pair of non-negative Lagrange multipliers ($\alpha_k \alpha^*_k = 0$) lead to the spares model while the data point that does not vanish the pair of non-negative Lagrange multipliers ($\alpha_k \alpha^*_k \neq 0$) is known as support vector. For the sparse model ($\alpha_k \alpha^*_k = 0$) the QP problem is only need to be calculated for the support vectors. So, the optimal desired wight vector for the regression hyperplane is calculated as follow:

$$W = \sum_{k=1}^{N}(\alpha_k - \alpha_k^*) \ \psi(x_k) \tag{3.61}$$

and the bias $b$ for the regression hyperplane is calculated such that the following condition is satisfied for all the support vectors.

$$\varepsilon - y_k + \widehat{y_k} = 0 \tag{3.62}$$

Finally, the regression function can be written as

$$\widehat{y_k}(x) = \sum_{k=1}^{N}(\alpha_k - \alpha_k^*) \ k(x, x_k) + b \tag{3.63}$$

### 3.3.3 Adaptive Neuro-Fuzzy Inference System

The fuzzy logic approach is based on the predefined rules (if-then) that lacks the ability to learn and adapt them-self to a new condition. Thus to overcome this drawback authors in (J. S. R. Jang, 1993) hybridized a fuzzy inference system (FIS) with ANN to form ANFIS. The ANFIS methodology can be considered as an adaptive system in the form similar to ANN in which by training the system the parameters of the fuzzy membership functions (antecedent parameters) and the parameters of the fuzzy system output function (consequent parameters) are adapted. ANFIS possesses the advantage of both FIS and ANN and it has been solved the drawbacks of both systems, while the complicated procedures of neural networks are bypassed by applying linguistic variables of FIS system, and the lack of FIS is solved by applying the neural inference system which create the ability to learn and adapt them-self to new condition. Therefore, this approach is capable to simulate complex nonlinear mappings using fuzzy system with ANN learning, and it is considered as a universal estimator capable for short, medium, and long-term forecasting.

ANFIS was developed as an adaptive system with a set of fuzzy rules (if-then) and tunable membership function (MF) parameters in a training phase. During the training phase of ANFIS, two different parameters should be optimized to provide the learning procedures:

    i-      Antecedent parameters (the MF parameters)

    ii-     Consequent parameters (the fuzzy system output function)

As the consequent parameters are linear, to optimize these parameters the linear least-squares method is applied and to optimize the antecedent parameters similar to neural networks the backpropagation algorithm in conjunction with an optimization method such as gradient descent is applied.

Generally, five different layers construct the ANFIS structure while each layer consists of node functions and the inputs of the nodes in the present layer are obtained from previous layers (Tavana, Fallahpour, Di Caprio, & Santos-Arteaga, 2016). The consecutive layers of ANFIS structure are as follows: layer 1 is fuzzification (if-part), layer 2 is production part, layer 3 is normalization part, layer 4 is defuzzification (then-part), and eventually layer 5 is total output generation part. Figure 3.4 shows the structure of ANFIS with two independent variables ($x$ and $y$) as input and one dependent variable $f_{out}$ as an output.
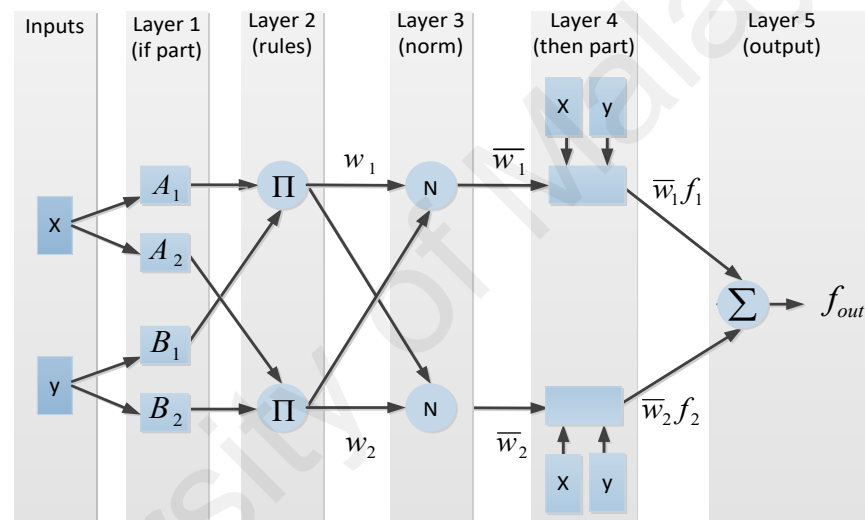


**Figure 3.4: The general structure of ANFIS**

For fuzzy inference systems, differing in the consequence of the set of fuzzy rules (if-then) and defuzzification procedures lead to two different types of fuzzy inference systems known as Mamdani type FIS and Sugeno type FIS.

In many respects, Mamdani type FIS is similar to Sugeno method. The fuzzifying the inputs data and executing the fuzzy operators are similar fuzzy inference process in both types. The main difference between Sugeno type FIS and Mamdani type FIS is the way the fuzzy inputs are converted to the crisp output. In Mamdani type FIS for computing the crisp output the defuzzification technique of a fuzzy output is used while in Sugeno type FIS the weighted average method is used. As the consequents of the rules are not

fuzzy in the Sugeno method, the interpretability and expressive power of Mamdani output are eliminated in this method. In comparison to Mamdani type FIS, Sugeno has faster processing time since instead of the time consuming defuzzification process the weighted average method is applied. Particularly, in decision support applications Mamdani method is widely applied, due to intuitive nature and the interpretable of the rule base provided in this method. Moreover, another difference between Sugeno and Mamdani type FIS is that Sugeno has no output membership functions whereas Mamdani FIS has output membership so, Sugeno method provides an output that is either linear (weighted) mathematical expression or a constant. Instead, Mamdani method provides an output that is a fuzzy set. ANFIS architectures representing both the Mamdani and Sugeno methods. In comparison to Mamdani type FIS, Sugeno has more flexibility in system design as latter can be integrated with ANFIS tool to model the systems more precisely (Svalina, Galzina, Lujić, & Šimunović, 2013).

Considering ANFIS with Sugeno type FIS, so the rule base of ANFIS contains fuzzy IF-THEN rules of a first order Sugeno type FIS are stated as:

Rule 1:  *If x is $A_1$ and y is $B_1$ then z is $f_1(x, y; p_1, q_1, r_1) = x\,p_1 + y\,q_1 + r_1$*

Rule 2:  *If x is $A_2$ and y is $B_2$ then z is $f_2(x, y; p_2, q_2, r_2) = x\,p_2 + y\,q_2 + r_2$*

where $f_i(x, y; p_i, q_i, r_i)$ is a first order polynomial function which represents the outputs of the Sugeno type FIS, $A_i$ and $B_i$ are the fuzzy sets, and $x$ and $y$ are two different input and $z$ is an output of ANFIS model.

In the ANFIS structure different layers consists of different node function. As shown in Figure 3.4, adaptive nodes which represent the adjustable parameter sets are denoted by squares whereas fixed nodes which represent the fixed parameter sets in the system are denoted by circles.

- **Layer 1**

Every node in this layer is an adaptive node with a node function as follow:

$$Q_{1,i} = \mu_{Ai}(x), \qquad i = 1,2 \tag{3.64}$$

$$Q_{1,i} = \mu_{Bi-2}(y), \quad i = 3,4 \tag{2.65}$$

where $x$ and $y$ are the inputs to node $i$ , $A_i$ and $B_i$ are linguistic labels, $\mu_{Ai}$ and $\mu_{Bi}$ are the membership functions for $Ai$ and $Bi$ fuzzy sets, respectively and $Q_{1,i}$ is the membership grade of a fuzzy set and considered as the output of node $i$ in the first layer which specifies the degree to the given input ($x$ or $y$) satisfies the quantifies.

Typically in ANFIS, the MF (membership function) for a fuzzy set can be any parameterized membership function, such as generalized Bell shaped function, Gaussian, trapezoidal or triangular.

A generalized Bell shaped MF (bell MF) is specified as folows:

$$\mu_A(x; a, b, c) = \frac{1}{1 + \{\frac{x-c}{a}\}^{2b}} \tag{3.66}$$

A Gaussian MF is specified as folows:

$$\mu_A(x; c, \sigma) = e^{-0.5(\frac{x-c}{\sigma})^2} \tag{3.67}$$

while $\sigma$ and c determined the width and center of Guassian MF, respectively.

A trapezoidal MF is specified as folows:

$$\mu_A(x; a, b, c, d) = max(min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0) \tag{3.68}$$

The parameters with $a < b \leq c < d$ specify the $x$ cordinates of the four corners for the underlying trapezoidal MF.

A triangular MF is specified as folows:

$$\mu_A(x; a, b, c) = max\left(min\left(\frac{x - a}{b - a}, \frac{c - x}{c - b}\right), 0\right) \tag{3.69}$$

The parameters with $a < b < c$ specify the $x$ cordinates of the three corners for the underlying triangular MF. Where in this layer, the parameters a,b,c,d and $\sigma$ are the antecedent parameters.

- **Layer 2**

Every node in this layer is a fixed node whose output is the product of all the incoming signals. In this layer through multiplication of input signals the firing strength of each rule is determined.

$$Q_{2,i} = w_i = \mu_{Ai}(x)\mu_{Bi}(y), \qquad i = 1,2 \tag{3.70}$$

where $w_i$ is output signal which represents the firing strength of a rule.

- **Layer 3**

Every node in this layer is a fixed node. In this layer, the firing strength provided in previous layer is normalized by computing the ratio of the i$^{th}$ rule's firing strength to the sum of all rules' firing strengths.

$$Q_{3,i} = \overline{w_i} = \frac{w_i}{w_1 + w_2} \qquad i = 1,2 \tag{3.71}$$

where $\overline{w}$ is output signal which represents the normalized firing strength of a rule.

- **Layer 4**

In this layer every node $i$ is adaptive with a node function.

$$Q_{4,i} = \overline{w_i} f_i \qquad\qquad i = 1,2 \qquad\qquad (3.72)$$

where $f_1$ and $f_2$ are the fuzzy IF-THEN rules as follows:

Rule1:  *If x is $A_1$ and y is $B_1$ then z=$f_1$(x, y; $p_1$, $q_1$, $r_1$)*

Rule2:  *If x is $A_2$ and y is $B_2$ then z=$f_2$(x, y; $p_2$, $q_2$, $r_2$)*

where $r_i$, $q_i$ and $p_i$ are the parameter set, referred to as the linear consequent parameters.

- **Layer 5**

This layer has only one fixed node that computes the overall output of ANFIS by summation of all incoming signals.

$$Q_{5,i} = f_{out} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} = \ overal\ output \qquad i = 1,2 \qquad (3.73)$$

The overall output is linear combination of the consequent parameters. Thus, the final output of ANFIS is expressed:

$$f_{out} = \overline{w_1} f_1 + \overline{w_2} f_2 = \frac{w_1}{w_1 + w_2}\ f_1 + \frac{w_2}{w_1 + w_2}\ f_2$$

$$(3.74)$$

$$= (\overline{w_1}x)p_1 + (\overline{w_2}x)p_2 + (\overline{w_1}y)q_1 + (\overline{w_2}y)q_2 + (\overline{w_1})r_1 + (\overline{w_2})r_2$$

Eventually, ANFIS applies a hybrid learning algorithm for parameters tuning. It utilizes the back propagation algorithm and the least squared method for updating the input MF parameters (antecedent parameters) in layer 1, and training the consequent parameters, respectively.

### 3.3.4    Gene Expression Programming

GP as a preliminary version of GEP (gene expression programming) is a supervised machine learning technique that creates sophisticated computer programs (mathematical models) to fit the experimental data using the principle of Darwinian natural selection (evolutionary change). Generally, GP is defined as an evolutionary algorithm that searches a program space instead of a data space to discover complex relationships among observed data. The computer programs developed by GP and its variants are based on parse trees that adapt and learn by changing their shapes, sizes, and structural architecture.

Unlike other AI techniques, GP does not suffer from black-box problem, due to its ability to generate explicit formulations (mathematical models) without assuming prior form of the existing relationship.

In GP the tendency for growing program size (depth of parse trees) as the search progresses without increasing in quality of results is known as bloat, which leads to producing nested functions. The bloat phenomenon causes a number of problems (e.g. large, complex, inefficient, and non-functional programs) that limit the application of GP systems. To overcome this drawback, the GEP has been proposed as an extension of GP.

GEP is extremely versatile and greatly surpasses the existing evolutionary methods, as the GEP inherited the expressive parse trees of varied sizes and shapes from GP; and it inherited the linear chromosomes of fixed length from GA to avoid bloat (Oltean & Grosan, 2003). The comparison between GEP, GP, and GA is summarized in Table 3.3. In GEP, the linear chromosomes of fixed length composed of genes structurally organized in a head and tail (i.e. work as genotype) and the parse trees (i.e. work as phenotype), developing a genotype-phenotype system.

**Table 3.3: Comparison of GEP technique with GP and GA**

| GEP | GP | GA |
|---|---|---|
| It is closely related to GP and GA. | Its chromosomes are non-linear, varying in shape as well as length, which are known as 'parse trees' | It has linear chromosomes of fixed length. |
| It inherited the expressive parse trees of varied sizes and shapes from GP; and it inherited the linear chromosomes of fixed length from GA. | At the same time, it uses a single entity working as genome (gene) and phenome (body). | Similar to GP, it uses a single entity working as genome and phenome at the same time. |
| It does not suffer from bloat problem. | It suffers from bloat problem. | - |
| It produces valid expressions. | Sometimes, it produces invalid expressions. | Sometimes, it can lead to invalid solutions. |
| It is well established beyond the replicator threshold. | It is not well established beyond the replicator threshold | It is not established beyond the replicator threshold. |
| It still has difficulty discovering suitable numeric constants for terminal nodes in the expression trees. | It has difficulty for constant creation procedure. | Unlike the GEP and GP, its genetic operators work on data space to optimize the given problem (i.e. optimization approach). |

The genotype-phenotype system in GEP benefiting from a simple genome to keep and transmit the genetic information and a complex phenotype to explore the environment and adapt to it much like a living organism. The computer programs generated by GEP are composed of multiple parse trees (expression trees) due to multigenic nature of its genotype-phenotype system, which allows evaluation of more complex programs comprised of several subprograms.

The genome of GEP consists of a linear, symbolic string (chromosome) with fixed length composed of one or more genes with equal size structurally organized in head and tail domains. Each allele of a gene is constructed from a predefined set of mathematical functions, variables, and numeric constants. The head domain contains mathematical functions as well as terminals and random numeric constants (RNC), while the tail domain can contain only terminals and RNC, which provides essentially a reservoir of terminals and pool of RNC (i.e. Dc domain) to ensure that all genes always correspond to

valid expression tree (ET). The arrangement of functions and terminals in the head and tail domains is known as structural architecture. The head length is defined by user, while the length of the tail is determined as

$$t = h \times (n_{max} - 1) + 1 \qquad (3.75)$$

where $t$, $h$, and $n_{max}$ represent the tail length, the head length, and the number of arguments for the function that takes the most arguments, respectively.

Once the head length and the set of functions have been set, the length of a gene is also decided accordingly. The genes of GEP can be transformed into an ET by the level-order traversal, and these genes within one chromosome are connected sequentially together by a user-specified linking function which is not shown in the chromosome. When expressed as trees, each gene in the chromosome codes for a sub-ET, and the sub-ETs add via the linking function to form the expression tree. In GEP, individual solutions are based on 'ETs' used to represent functional forms, which fit the experimental data. These ETs are made of 'nodes', which start with a rote node and correspond to branch points in the tree, or to 'leaves' that terminate the branches. The branch points are functional operators. The 'leaves' (terminals) are either variables or RNC. As well as the variables, RNC can also be used as terminals, giving added flexibility in evolution of the models.

It is easy to manipulate genetically the linear chromosomes of fixed length, without losing their functional complexity. Hence, in GEP, the creation of genetic diversity is extremely simplified, as the evolutionary operators work at chromosome level not expression trees (ETs); and only, the chromosomes are transmitted in the process of reproduction. Therefore, the pivotal insight of GEP consisted in the invention of Karva

notation (K-expression) as an unequivocal translation system to encode the language of chromosomes to the language of ETs.

Eq. (3.76) is considered as an example to clarify the genotype-phenotype system of GEP. An algebraic expression as in Eq. (3.76) is represented either by a K-expression or by an ET in Figure 3.5.

$$\sqrt{\frac{(a-d)}{(c+b)}} + sin\left(\frac{c}{b}\left(1.7+a\right)\right) \tag{3.76}$$

The Eq. (3.76) is coded by a chromosome (model) with two genes (terms), where the head length of each gene is four, the predefined mathematical functions are $\{\sqrt{\ }, sin, -, +, \times, /\}$ and the variables are $\{a, b, c, d\}$. According to Eq. (3.75) the tail length of each gene is specified to be five ($t=5$) while its head length is four ($h=4$) and the number of arguments for the predefined mathematical functions that takes the most arguments is equal to two ($n_{max}=2$).

The mathematical form of ET is expressed as a chromosome, which is the straightforward reading of the ET from left to right and from top to bottom. Eq. (3.76) can also be mapped as an ET. To map the encoded information within a gene into the sub-ET, the K-expression string is scanned gene by gene. Then the first position of each gene is placed to the root of the corresponding sub-ET. This mapping process continues sequentially until all leaf nodes in the ET are composed of elements from the terminal set. As illustrated in Figure 3.5 the first and second genes evolve simultaneously within the chromosome are interpreted by sub-ET$_1$ and sub-ET$_2$ respectively, which are joined by an addition function to form an ET structure. The ET consists of a group of nodes, which are shown by circles. Each node represents either a mathematical function or a variable.

The variables are leaf nodes and the predefined mathematical functions are branch points that govern the connectivity between the terminal nodes.
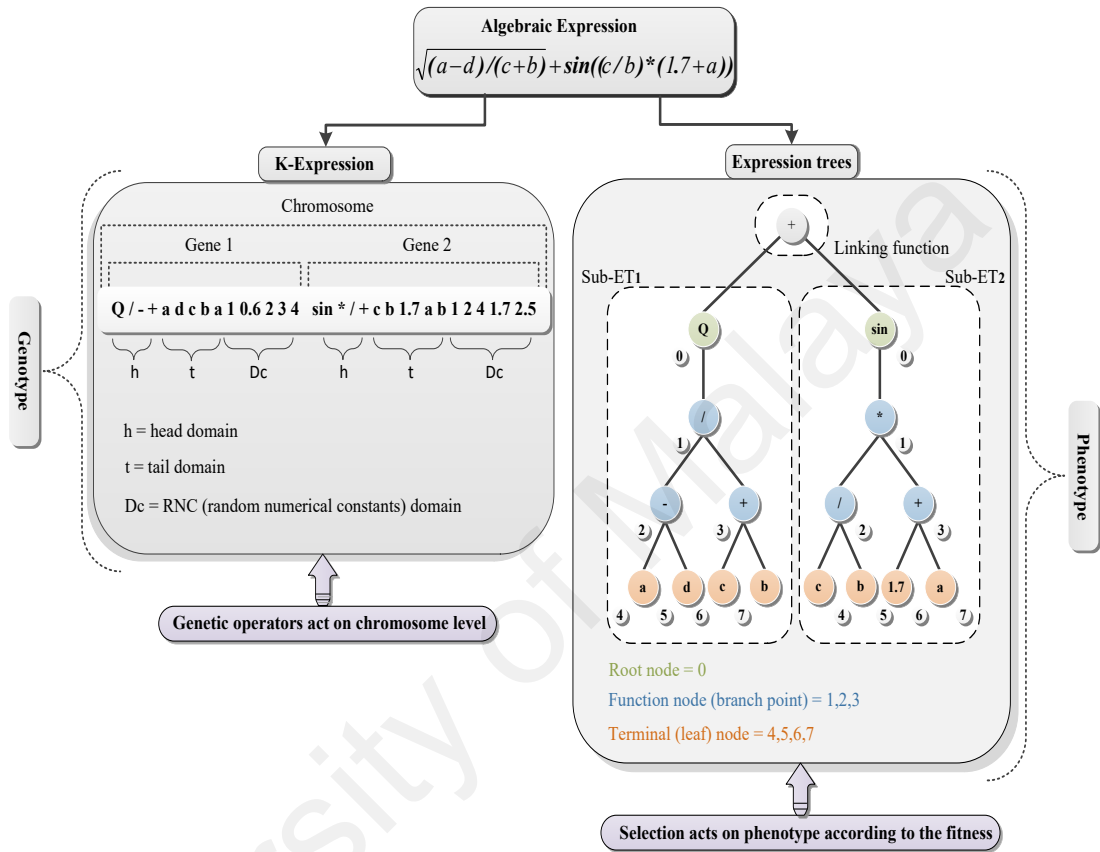


**Figure 3.5: GEP's genotype-phenotype system attached with considered mathematical equation**

The generic algorithm of GEP is shown in Table 3.4. GEP can be explained by dividing its evolution process into four major steps: initialization, elitist selection, reproduction, and termination.

**Table 3.4: General structure of GEP**

```
    1. Initialization
repeat
 | 2. Selection and Replication
 |  Reproduction
 | | 3. Mutation
 | | 4. IS transposition
 | | 5. RIS transposition
 | | 6. Gene transposition
 | | 7. Single or double crossover
 | | 8. Gene crossover
 | | 9. Inversion
 | end
until termination criteria are satisfied;
```

- **Initialization**

The initial chromosomes equal to the dimension of population are generated randomly to form the population of solutions. In GEP method, each individual of population (chromosome) composed of genes structurally organized in a head and tail to generate valid solution. The separation of genes in two parts (head and tail) implies Karva notation as a universal way of representing any mathematical or logical expression that can be represented as a parse tree with different sizes and shapes. Through Kerva notation, all chromosomes are translated to ETs, and then the solutions are executed to obtain their fitness values.

- **Selection and Replication**

The selection operator selects the programs for the replication operator to copy a chromosome with high fitness score into the new generation. In this stage of GEP for reducing the risk of losing fit individuals, the potentially useful individuals are selected into the next generation according to their fitness by roulette wheel selection (fitness proportionate selection) with elitism. The individuals with higher fitness scores will be

less likely to be eliminated in reproduction process, to guarantee the survival and cloning of the best chromosomes in new population.

The reproduction process genetically manipulates the population by conducting genetic operations on randomly selected chromosomes as illustrated in Table 3.5. Thus, in GEP, a chromosome might be modified by one or several operators at a time or not be modified at all. The pseudocode of GEP is provided in Table 3.5 to elucidate the operation of genetic operators in this method. The reproduction process in GEP includes the following genetic operations:

- **Mutation**

To increase the diversity of the population, mutation is placed at any location of a chromosome but it does not affect its structure, i.e., the alleles in the head change in to other functions or terminals, whereas alleles in the tail can be changed only into other terminals.

- **IS (insertion sequence) transposition**

To facilitate the evolution process in GEP, the transposition operators create simple repetitive sequences in the genome. The transposition operators randomly select the chromosome, the gene to be manipulated, the first position of transposon (transposable element), and its length. In IS transposition operation any sequence in the genome is inserted randomly at any position in the head of a randomly chosen gene expect for the first position (gene root).

- **Root insertion sequence (RIS) transposition**

In RIS transposition operation, any transposon with a function in the first position is inserted randomly at root of a randomly chosen gene.

- **Gene transposition**

This operator entirely transposes a randomly selected gene into the root of the chromosome.

- **Single and double crossover**

In 1-point recombination (single crossover) operation, the genetic material is exchanged at the randomly selected single location between a pair of randomly chosen chromosomes (parent) to generate new offspring. In 2-point recombination operation (double crossover), two randomly selected substrings are exchanged between two chromosomes.

- **Gene crossover**

This operator entirely exchanges two randomly selected genes in two chromosomes to generate new offspring.

- **Inversion**

By means of this operator, the allele in the gene head of a randomly selected chromosome is reallocated in the reverse order to facilitate the evaluation process.

- **Termination criteria**

The program executes the aforementioned genetic operations iteratively for a certain number of generations or until an explicit formulation has been found. When the stopping conditions are met, the best mathematical model in the form of expression tree is exported to the output.

**Table 3.5: Pseudocode of GEP algorithm**

---

**Input**: Generation $_{max}$, Population $_{size}$, Genes $_{numbers}$, Head $_{length}$, Function $_{set}$, Terminal $_{set}$, Constants $_{per\ gene}$, DC $_{limit}$ Crossover $_{rate}$, Mutation $_{rate}$, Inversion $_{rate}$, Transposition $_{rate}$

**Output**: Solution $_{Best-Cost}$, Solution $_{Best-ET}$

// **Initialization**//

1. population $\longleftarrow$ initialize population (Population $_{size}$, Genes $_{numbers}$, Head $_{length}$, Function $_{set}$, Terminal $_{set}$, Constants $_{per\ gene}$, DC $_{limit}$)

2. **for** each Solution $_i$ $\in$ population **do**

    // Translate the Chromosome into Expression Tree //

3.  | Solution $_{i\_ET}$ $\longleftarrow$ translate breadth first (Solution $_{i\_genes}$)

    // Execute the Corresponding Expression Tree//

4.  | Solution $_{i\_cost}$ $\longleftarrow$ execute (Solution $_{i\_ET}$)

5. **end**

    // **Elitist selection & Replication** //

6. Solution $_{Best}$ $\longleftarrow$ select best solution (population)

7. population $\longleftarrow$ copy Solution $_{Best}$

8. **while** stopping condition are not met **do**

    // **Parent Selection Process**//

9. parent $_i$ $\longleftarrow$ select parents (population)

10. parent $_j$ $\longleftarrow$ select parents (population)

    // **Crossover operator**//

11. offspring $_1$ $\longleftarrow$ crossover (parent $_i$, parent $_j$, Crossover $_{rate}$)

12. offspring $_2$ $\longleftarrow$ crossover (parent $_j$, parent $_j$, Crossover $_{rate}$)

    // **Mutation operator** //

13. offspring $_{1m}$ $\longleftarrow$ mutation (offspring $_1$, Mutation $_{rate}$)

14. offspring $_{2m}$ $\longleftarrow$ mutation (offspring $_2$, Mutation $_{rate}$)

    // **Inversion operator** //

15. offspring $_{1\_inversion}$ $\longleftarrow$ inversion (offspring $_{1m}$, Inversion $_{rate}$)

16. offspring $_{2\_inversion}$ $\longleftarrow$ inversion (offspring $_{2m}$, Inversion $_{rate}$)

    // **Transposition operator** //

17. offspring $_{1\_transposition}$ $\longleftarrow$ inversion (offspring $_{1\_inversion}$, Transposition $_{rate}$)

18. offspring $_{2\_transposition}$ $\longleftarrow$ inversion (offspring $_{1\_inversion}$, Transposition $_{rate}$)

    // **Traslate the Choromosme into Experession Tree**//

19. offspring $_{1\_ET}$ $\longleftarrow$ translate breadth first (offspring $_{1\_transposition}$)

20. offspring $_{2\_ET}$ $\longleftarrow$ translate breadth first (offspring $_{2\_transposition}$)

    // **Execute the Corresponding Expression Tree** //

21. offspring $_{1\_cost}$ $\longleftarrow$ execute (offspring $_{1\_ET}$)

22. offspring $_{2\_cost}$ $\longleftarrow$ execute (offspring $_{2\_ET}$)

    // **Roulette Wheel Selection** //

23. population $\longleftarrow$ population update RWS (offspring $_{1\_cost}$, offspring $_{2\_cost}$)

24. **end**

25. return to best soloution

---

### 3.3.5    Optimized Gene Expression Programming

The key issue in evolutionary modeling is balancing the exploitation of solution structure and exploration of its appropriate weighting factors. GEP lacks the ability to precisely generate and evolve empirical coefficients of generated mathematical model (Zhong et al., 2015). While evolution process of GEP has been developed more efficiently for exploitation of solution structure, it is difficult for evolution to come up with appropriate weighting factors. The RNC algorithm handles constant creation in GEP. To refine the weighting parameters of functional form expressed as a tree structure a metaheuristic optimization method is invoked over final solution of standard GEP, which is an optimized extension to GEP. The pivotal insight of the optimized GEP lies in the integration of a robust and efficient metaheuristic optimization method with GEP while optimization method with fixed fitness evaluation function as in GEP is invoked to optimize a parameter vector consisted of all weighting parameters in GEP chromosomes. Thus, learning process of optimized GEP is divided into two phases: in the first phase, the exploration of solution structure is provided by GEP; while in the second phase, the exploitation of its optimal weighting factors is provided by optimization method. The search power of metaheuristic method on parameter optimization tunes up the constant creation process in GEP and improves the prediction accuracy without further complicated model. Due to effectiveness and the potential benefits provided by optimized GEP, fewer control parameters (i.e. functional operators) can be selected to precisely adapt the information into a mathematical model.

Four different robust and efficient metaheuristic optimization methods, namely PSO, CSA, ACS, and BSA, which not only provide highly accurate results but also have simple implementation procedure with fewest control parameters, are examined to assess the most effective optimization algorithm for optimal training in the second learning process of optimized GEP.

### 3.3.5.1 Particle Swarm Optimization

PSO as a population-based metaheuristic technique has attracted significant attention to tackle the complexity of different optimization problems due to its simple concept with efficient search mechanism on real-valued numerical optimization (Sebtahmadi, Azad, Kaboli, Islam, & Mekhilef, 2017). It is inspired by particles moving around in the search space. As shown in Figure 3.6, each particle in PSO has a position ($x_n$) and a velocity ($v_n$). Each particle keeps track of the overall best value, $g_{best}$, as well as its previous best position, $P_{best}$. The new position of the particle is computed as follows:

$$v^{(i)}_{n+1} = w_n v^{(i)}_n + c_1 \, rand_1(P_{best}{}^{(i)}_n - x^{(i)}_n) + c_2 \, rand_2(g_{best}{}_n - x^{(i)}_n) \quad (3.77)$$

$$x^{(i)}_{n+1} = x^{(i)}_n + v^{(i)}_{n+1} \qquad (i = 1...nPop) \quad (3.78)$$
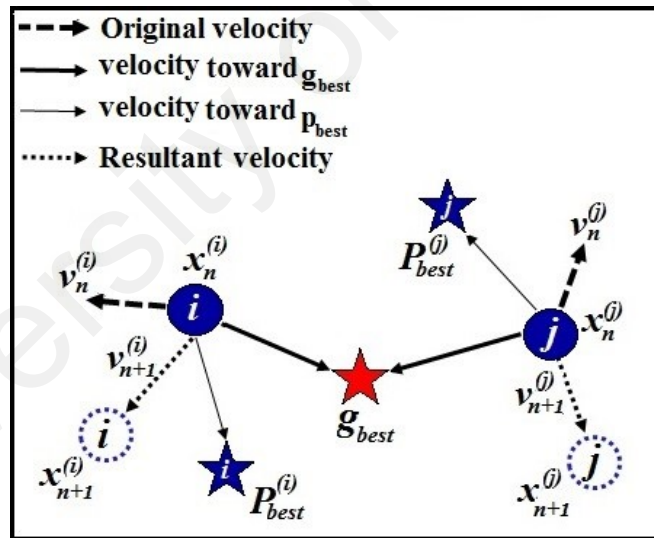


**Figure 3.6: The mechanism of particles ($i$ & $j$) movement toward the global position ($g_{best}$) within 2-dimensions search space**

where $rand_1$ and $rand_2$ are the random values distributed uniformly between 0 and 1, $c_1$ and $c_2$ are the acceleration factors, which determine the relative pull for particles towards $P_{best}$ and $g_{best}$. $w$ is the inertia weight that keeps a balance between global and local searches. The main structure and pseudocode of PSO is depicted in Tables 3.6 and 3.7 respectively.

**Table 3.6: General structure of PSO**

1. Initialization
   Generate random initial solution for particles
   Determine $g_{best}$ of swarm and $P_{best}$ of particles
2. For all particles,
   Generate new solution by using (3.77) and (3.78)
3. Update global $P_{best}$ and local $g_{best}$
4. Not stopping criterion go to step 2
5. Stop

To provide balance between exploration of the problem's search and exploitation of better results, instead of using a constant inertia weight an adaptive inertia weight is used. Global search requires larger step sizes at the beginning of the optimization process to determine the most promising areas. To provide local search ability, then the step size is reduced to focus only on that area. Thus, the inertia weight in PSO is adapted descendingly as a function of iteration ($n$) according to Eg. (3.79).

$$w_n = w_{max} - \left(\left(w_{max} - w_{min}\right)/w_{max}\right) n \qquad (3.79)$$

where $n_{max}$ is maximum number of iterations and $w_{min}$ and $w_{max}$ are the minimum and maximum boundary of inertia weight, respectively.

**Table 3.7: Pseudocode of PSO**

**Data**: nPop, nVar, max cycle, low, up, $C_1$, $C_2$, $w_{min}$, $w_{max}$
Result: $x_{best}$ | $g_{best}$ = f ($x_{best}$)
    **//Initialization**
1.   $g_{best, n=0} \approx \infty$
2.  **for** i ← 1 **to** nPop **do**
3.      **for** k ← 1 **to** nVar **do**
4.         $x_{i,k} \sim U$ (low$_k$, up$_k$)
5.      **end**
6.      $v_i = 0$
7.      $y_i = f(x_i)$
8.      $P_{best, i} := x_i$
9.      **if** $y_i < g_{best}$ **then** $g_{best} := y_i$ **end**
10. **end**
11. **for** n ← 1 **to** max cycle **do**
12.    **for** i ← 1 **to** nPop **do**
13.       $w = w_{max} - ((w_{max} - w_{min}) (n / \text{max cycle}))$
14.       $v_i = w\, v_i + (P_{best, i} - x_i) (c_1\, \text{rand}) + (g_{best} - x_i) (c_2\, \text{rand})$
15.       $x_i = x_i + v_i$
16.       **for** k ← 1 **to** nVar **do**
17.          **if** $x_{i,k} > $ up$_k \vee x_{i,k} < $ low$_k$ **then**
18.            $x_{i,k} = x_{i,k} - v_i$
19.            $x_{i,k} := \min (\max (x_{i,k}, \text{low}_k), \text{up}_k)$
20.          **end**
21.       **end**
22.       $y_i = f(x_i)$
23.    **end**
24.    **if** min f (x) < $g_{best}$ **then**
25.       $g_{best} := \min$ f (x)
26.       $x_{best} := \arg\min$ f (x)
27.    **end**
28. **end**

**3.3.5.2  Cuckoo Search Algorithm**

Cuckoo search algorithm (CSA) as a global optimization algorithm was inspired by the obligate brood parasitism of cuckoo and cowbird by laying their eggs in the nests of other host passerines to increase their reproductivity (Gandomi, Yang, & Alavi, 2013). The cuckoo and cowbird seek the parental care and nests of other host birds for their brood to fledge. The organisms with such breeding behavior are known as brood parasites. Although, the female parasitic cuckoos are very specialized in the mimicry in pattern and colors of eggs, some host bird can distinguish between their own eggs and intruding eggs. The host bird will either build a completely new nest in new location or simply throw alien eggs away if alien eggs are discovered. Eventually, the alien eggs are hatched to chicks by the host birds, if the alien eggs have not been detected.

CSA is based on the following idealized rules while the aim is to employ the cuckoo eggs as the potentially better solutions to replace with not-so-good solution in the nests.

- i-     Each egg in a nest represents a solution and a cuckoo's egg represents new solution.
- ii-    Each cuckoo lays one egg in a randomly chosen nest at a time.
- iii-   For the next generation the best nests with high quality of eggs are chosen.
- iv-    The host bird can discover the alien eggs with probability $P_a \in (0, 1)$. In this case, the host bird abundant its nest to build new one elsewhere.

CSA as a nature-inspired algorithm has the advantage of a simple implementation procedure and only a few control parameters. Despite its simple structure, it has been robust and effective algorithm with high probability of finding optimal parameters at numerical optimization. In CSA, cuckoos select a random nest to lay their eggs. Eggs with the highest quality (the better solution) are passed to the next generation by an elitist selection process. Alien eggs are detected by the host birds and thrown away or the nest

is discarded instead. The path and the position iterative formula of CSA comprises five stages: initialization, generating, evaluating, elitist selection, and alien egg discovery as presented in Table 3.8.

**Table 3.8: General structure of CSA**

| |
|---|
| 1. Initialization |
| **Repeat** |
| **lay the cuckoo eggs in the host nest** |
|   2. Generate the cuckoo egg with Lévy flight |
|   3. Evaluate the quality of cuckoo egg |
|   4. Elitist selection |
| **end** |
| 5. Alien egg discovery |
| **until** stopping conditions are met**;** |

- **Initialization**

The initial values of the i <sup>th</sup> host nest is defined by using Eq. (3.80).

$$
\begin{aligned}
&x_{i,j;g=0} \sim U(low_j, up_j) \quad , y_{i;x} = f(x_i) \\
&for \\
&i = \{1,2,3,\dots nPop\} \ , \ j = \{1,2,3,\dots nVar\}
\end{aligned}
\tag{3.80}
$$

where:

$nPop$ is population size of host nests;

$nVar$ is number of respective optimization variable;

$U$ represents the uniform distribution function;

$low_j$ and $up_j$ are lower and upper search space limits of j<sup>th</sup> variable;

$y_{i;x}$ are productivity of i<sup>th</sup> host nest;

$g$ is generation number;

- **Generate the cuckoo egg with Lévy flight**

The cuckoo's egg as a new solution is calculated based on the previous best nests ($x^{best}$) by Lévy flights. The Lévy flights is a random walk while the step-lengths are distributed acording to the heavy-tailed probability distribution. The new solution of CSA is calculated as follows:

$$x_{g+1}^{new} = x_g^{new} + \alpha \ rand_3 \ L(\beta)$$
(3.81)

where

$$L(\beta) = \upsilon \times \frac{\sigma_x(\beta)}{\sigma_y(\beta)} \times (x_g^{new} - x_g^{best})$$
(3.82)

where

$$\upsilon = \frac{rand_x}{|rand_y|^{(\frac{1}{\beta})}} \ , \ \sigma_x(\beta) = \left[ \frac{\Gamma(1+\beta)sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2})\beta(2^{(\frac{\beta-1}{2})})} \right]^{\frac{1}{\beta}} \ , \ \sigma_y(\beta) = 1$$
(3.83)

where

$L(\beta)$ is a Lévy distribution function describing randomly walked steps with an infinite variance and infinite mean, $\beta$ is the distribution factor ($\beta \in [0.3, 1.99]$);

$rand_3$ is a normally distributed random number in $[0, 1]$, $rand_x$ and $rand_y$ are two normally distributed stochastic variables with standard deviation $\sigma_x(\beta)$ and $\sigma_y(\beta)$, respectively;

$\Gamma(.)$ is the gamma distribution function, $x_g^{best}$ is corresponding to the nest with highest productivity among all nests in $g^{th}$ generation , and $a > 0$ is the step size;

- **Evaluate the quality of cuckoo egg**

The cuckoo selects a random host nest to lay its egg. If the quality of the generated cuckoo's egg in previous step is better than the quality of the randomly selected host egg, the new solution is kept for next generation. Therefore, in this step of CSA the quality of cuckoo's egg is compared with the quality of the host egg.

$$
\begin{aligned}
y_{k;x_g^{new}} &= f(x_g^{new}) \\
y_{k;x_g^k} &= f(x_g^k)
\end{aligned}
\tag{3.84}
$$

where $x_g^k$ is the randomly selected host egg in $g^{th}$ generation, and $y_{k;x_g^k}$ is its corresponding fitness.

- **Elitist selection**

If the fitness of host egg is worse than cuckoo's egg, the randomly selected host egg is replaced with cuckoo's egg. Figure 3.7 depicts the elitist selection process of CSA.
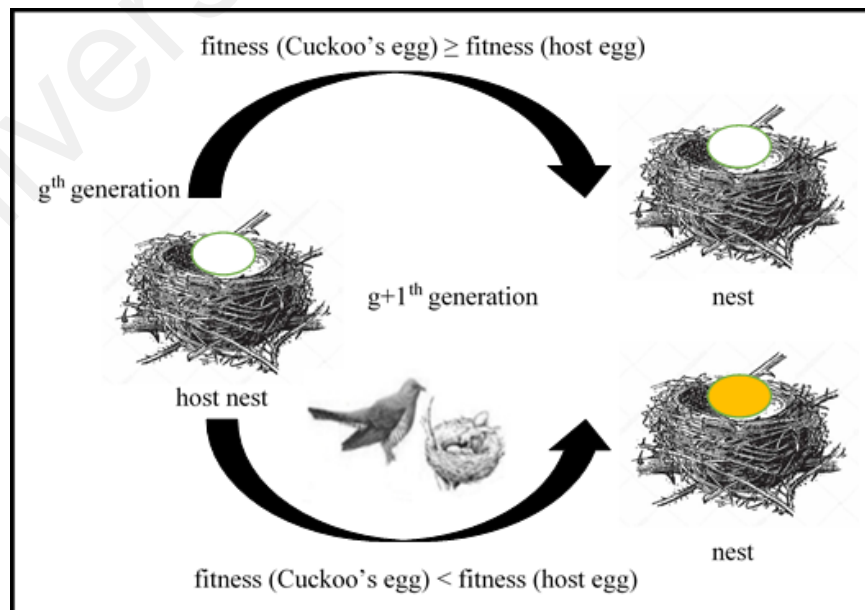


**Figure 3.7: The elitist selection process of CSA**

- **Alien egg discovery**

Discovery of an alien egg with the probability of $P_a \epsilon$ (0, 1) similar to the Lévy flights creates a new solution for the problem. The new solution due to this action is found out in the following way:

$$x_{g+1}^{dis} = x_g^{dis} + L \times rand_5 \left[ randp_1(x_g^{best}) - randp_2(x_g^{best}) \right] \qquad (3.85)$$

where

$$L = \begin{cases} 1 & if\ rand_4 < P_a \\ 0 & otherwise \end{cases} \qquad (3.86)$$

*rand* is the distributed random numbers in [0, 1] and *randp* is the random perturbation for positions of the nests in $x_g^{best}$. The same elitist selection process as in previous step is applied for this new solution. Eventually, the value of the fitness function is calculated and the nest corresponding to the best fitness function after all generation  is reported as the best nest ($x^{best}$).

The pseudocode of CSA algorithm is provided in Table 3.9.

**Table 3.9: Pseudocode of CSA**

**Data**: nPop, nVar, Max cycle, low, up, *f, ß, p_a*
Result: $x^{best} \mid y^{best} = f(x^{best})$
  **//Initialization of host nests**
1.  $y^{best}{}_{g=0} \approx \infty$
2. **for** i ← 1 to nPop **do**
3.     **for** j ← 1 **to** nVar **do**
4.       $x_{i,j} \sim U\ (low_j,\ up_j)$
5.     **end**
6.     $y_i = f(x_i)$
7. **end**
8.  **for** g ← 1 **to** max cycle **do**
9.     $x_{new}$ ← Generate a Cuckoo randomly by Lévy flight
10.    $x_k$    ← Choose a random nest among the host nests
11.       **if** $f(x_{new}) < f(x_k)$
12.          $x_k := x_{new}$
13.       **end**
14.    A fraction ($p_a$) of the worse nests are abandoned
15.    New nests are built via Lévy flights at new location
16.    Solutions/nests are ranked
17.    $m_{best} = min\ (f(x)) \mid best \in \{1, 2, 3,…, nPop\}$
18.       **if** $y < y^{best}$ **then**
19.          $y^{best} := m_{best}$
20.          $x^{best} := x_{best}$
21.       **end**
22. **end**

### 3.3.5.3   Artificial Corporative Search

A preliminary version of the ACS algorithm was presented in (Civicioglu, 2013a), in which applied for solving complex optimization problems and the results confirmed the well performance of this algorithm in comparison with other metaheuristic approaches.

As the superorganisms intelligently behave in nature, ACS drew inspiration from the cooperation and mutualism based biological interaction of two eusocial superorganisms living in the same habitat. A superorganism is any aggregate of individuals that behaves like an intelligent unified organism while the members of superorganism have a social cooperative instinct, and are unable to survive away from their superorganism for extended periods. There is a biological interaction based on amensalism between the superorganisms that have parasite-host or predator-prey relationship, which leads to coevolution, cooperation, or coextinction between the superorganisms.

The ACS is a new two-population search algorithm based on coevolution process. The ACS has been developed to overcome some of the drawbacks of metaheuristic approaches; e.g. too many control parameters and over sensitivity to initial value of these parameters, premature convergence and time-consuming computation due to deficient balance between exploitation of better results and exploration of the problem's search space.

The ACS has only one control parameter and is not over sensitive to the initial value of this parameter. Two operators; crossover and mutation, which provide the balance between exploitation and exploration, are utilized in this algorithm. These operators are unique and quite different from the structures of the crossover and mutation strategies defined in other evolutionary methods; e.g. GA and DE. The ACS has advantage of a memorization process to facilitate the exploration of the feeding areas. The ACS comprises seven stages: initialization, selection of Predator, selection of Prey, mutation,

crossover, boundary control, and export the best individual. The generic algorithm of ACS is shown in Table 3.10.

.
**Table 3.10: General structure of ACS**

| |
|---|
| 1. Initialization |
| **Repeat** |
| ⎸ 2. Selection of Predator |
| ⎸ 3. Selection of Prey |
| ⎸ **Finding active individuals** |
| ⎸ ⎸ 4. Mutation |
| ⎸ ⎸ 5. Crossover |
| ⎸ **end** |
| ⎸ 6. Boundary control |
| ⎸ 7. Export the best individual |
| **Until** stopping conditions are met |

- **Initialization**

In ACS algorithm, a superorganism consisting of random solutions of the related problem corresponds to an artificial superorganism migrating to more productive feeding areas. ACS algorithm contains two superorganisms; $\alpha$ and $\beta$ that have artificial sub-superorganisms equal to the dimension of the population. The initial values of the individuals of $i^{th}$ sub-superorganism belong to each superorganism ($\alpha$ and $\beta$) are defined by using Eq. (3.87).

$$
\begin{aligned}
&\alpha_{i,j;g=0} \sim U(low_j, up_j) \quad , y_{i;\alpha} = f(\alpha_i) \\
&\beta_{i,j;g=0} \sim U(low_j, up_j) \quad , y_{i;\beta} = f(\beta_i) \\
&for \\
&i = \{1,2,3,...nPop\} \ , \ j = \{1,2,3,...nVar\}
\end{aligned}
\tag{3.87}
$$

where:

*nPop* is the population size of individuals in the superorganisms $\alpha$ and $\beta$;

*nVar* is the number of respective optimization variable;

*U* represents the uniform distribution function;

*low$_j$* and *up$_j$* are the lower and upper search space limits of jth variable;

$y_{i;\alpha}$ and $y_{i;\beta}$ are the productivity of ith sub-superorganism related to $\alpha$ and $\beta$ superorganisms;

*g* is the generation number which express the coevolution level of the superorganisms;

- **Selection of Predator**

In this stage of ACS algorithm the coevolution process for both of superorganisms is provided, while the Predator sub-superorganism is determined randomly from two superorganisms in each generation through the 'if-then-else' decision rule according to Eq. (3.88).

$$
\begin{aligned}
&if \ a < b \Big|_{a,b\sim U(0,1)} then \ predator := \alpha, \ y_{predator} := y_\alpha, \ key := 1 \\
&else \ predator := \beta, \ y_{predator} := y_\beta, \ key := 2
\end{aligned}
\tag{3.88}
$$

where:

*:=* is the update operation;

*a* and *b* are randomly generated numbers;

*key* is a memory to track the origin of Predator in each iteration;

- **Selection of Prey**

In this stage of ACS algorithm, the similar decision rule provided in Eq. (3.88) is applied for random selection of Prey sub-superorganism from two superorganisms. To mimic the behavior of the natural superorganisms, the hierarchical sequencing between the individuals of Prey sub-superorganisms is permuted through a random shuffling function. The obtained Prey is used to determine the search direction in each generation according to Eq. (3.90). Subsequently, in this stage the Predator pursues the Prey for a period of time while they migrate towards more productive feeding areas which provide a memorization process to facilitate the exploration of the feeding areas.

$$if \ a < b \bigg|_{a,b \sim U(0,1)} \quad then \ prey := \alpha, \quad else \ prey := \beta$$
$$prey := permuting \ (prey)$$

$$(3.89)$$

- **Mutation**

The mutation process of ACS models biological interaction location $x$ among Predator and Prey sub-superorganisms according to Eq. (3.90) while the random walk function (Wiener process) is applied in mutation process to mathematically formulate the foraging behavior of sub-superorganisms. The local and global search capability of ACS algorithm is promoted, as the Prey sub-superorganisms is used to generate a mutation matrix $(x)$ by taking partial advantage of its experiences from previous generations.

$$x = predator + R.(prey - predator)$$
$$R = \ 4.a.(b\text{-}c) \bigg|_{a,b,c \sim U(0,1)}$$

$$(3.90)$$

- **Crossover**

A binary integer-valued matrix (*M*) that indicates the active individuals of the Predator are determined for crossover process according to Eq. (3.91). In ACS, active individuals stand for the individuals that participated in a migration at any time and only they can discover new biological interaction locations. During migration, the control parameter of ACS, determines the probability of corporation between individuals by selecting the biological interaction level within the crossover process. Eq. (3.91) elucidates the unique crossover strategy of ACS.

$$
\begin{aligned}
&M_{i,j} = 1, \\
&if \ a < (p.b)\Big|_{a,b \sim U(0,1)} \ then \ M_{i,randperm(randi(nVar))} = 0 \\
&else \ M_{i,randi(nVar)} = 0 \\
&if \ M_{i,j} > 0 \ then \ x_{i,j} := predator_{i,j} \\
&for \\
&i = \{1,2,3,...,nPop\} \ , \ j = \{1,2,3,...,nVar\}
\end{aligned}
\tag{3.91}
$$

where

*P* is the control parameter of optimization algorithm. ACS has only one control parameter; probability of biological interaction (*P*) which restricts the number of the active individuals of each sub-superorganism by controlling the number of individuals to be engaged in the crossover process. The experiments with different values of *P* exposed that ACS algorithm is not too sensitive to its control parameter, while *P* is in the range of [0.05 0.15].

*randperm(nVar)* is random permutation function that generates a row vector containing a random permutation of the integers from 1 to *nVar* inclusive.

*randi(nVar)* is random selection function that generates a pseudorandom integer between 1 and *nVar*.

- **Boundary control**

At the end of the crossover process, if the achieved interaction location *x* violates the habitat limits, the related interaction location is updated according to the boundary control mechanism developed in this stage.

$$if\ (x_{i,j} < low_j)\ or\ (x_{i,j} > up_j)\ then\ x_{i,j} \sim U(low_j, up_j) \tag{3.92}$$

- **Export the global minimum:**

Finally, the fitness value of the best individual is exported as global minimum and its position is considered as global minimizer according to the Eq. (3.93).

$$
\begin{aligned}
&if\ f(x_i) < y_{i;predator}\ then\ predator_i := x_i\ ,\ y_{i;predator} := f(x_i)\\
&if\ key=1\ then\ \alpha := predator\ ,\ y_\alpha := y_{predator}\\
&else\ \beta := predator\ ,\ y_\beta := y_{predator}\\
&y_g = min(y_{predator})\\
&if\ y_g < y_{g-1}\ then\ globalminimum := y_g\ ,\ globalminimizer := predator_g\ ,\\
&g = g+1\\
&for\\
&i = \{1,2,3,...,nPop\}\ ,\ g = \{1,2,3,...,gMax\}
\end{aligned}
\tag{3.93}
$$

Due to the probabilistic nature of ACS algorithm, the selected predator can be different in each generation. Thus, the cooperative/coevolution process is provided in ACS algorithm for both of the superorganisms. Moreover, the self-interaction process is provided in ACS algorithm while the origin of Predator and Prey sub-superorganisms are the same with each other in a generation.

Eventually, the pseudocode of ACS algorithm is provided in Table 3.11.

**Data**: nPop, nVar, Max cycle, low, up, f, P
Result: Globalminimizer | globalminimum = f (globalminimizer)
1. Superorganisms: $\alpha$, $\beta$
   **//Initialization**
2. globalminimum $_{g=0} \approx \infty$
3. **for** i $\leftarrow$ 1 **to** nPop **do**
4.     **for** j $\leftarrow$ 1 **to** nVar **do**
5.        $\alpha_{i,j}$, $\beta_{i,j} \sim U (low_j, up_j)$
6.     **end**
7.     $y_{i;\alpha} = f(\alpha_i)$
         $y_{i;\beta} = f(\beta_i)$
8. **end**
9. **for** g $\leftarrow$ 1 **to** max cycle **do**
       **// Selection**
10.    **if** rnd < rnd **then**
11.      Predator = $\alpha$, $y_{\textbf{Predator}} = y_\alpha$,    key=1
12.    **else**
13.      Predator = $\beta$, $y_{\textbf{Predator}} = y_\beta$,    key=2
14.    end
15.    **if** rnd < rnd **then** Prey = $\alpha$ else Prey= $\beta$ **end**
16.    Prey := permuting (Prey)
17.    **if** rnd < rnd **then** R=4.rnd.(rnd-rnd) **else** R $\sim \Gamma$ (4.rnd, 1) **end**
18.    $M_{1:nPop, 1:nVar} = 1$
19.    **for** q$\leftarrow$1 **to** nPop.nVar **do**
20.      **if** rnd < (P.rnd) **then** $M_{rndint\,(nPop),\;rndint\,(nVar)} = 0$ **end**
21.    **end**
22.    **if** rnd < (P.rnd) **then**
23.      **for** i $\leftarrow$ to nPop **do**
24.        **for** j $\leftarrow$ 1 to nVar **do**
25.          **if** rnd < (P.rnd) **then**
26.            $M_{i,j} = 1$
27.          **else**
28.            $M_{i,j} = 0$
29.          **end**
30.        **end**
31.      **end**
32.    **end**
33.    **for** i $\leftarrow$ 1 to nPop **do**
34.      if $\sum M_i$ = nVar **then** $M_{i,\,rndint\,(nVar)} = 0$ **end**
35.    **end**
       **// Mutation**
36.    x= Predator + R.(Prey – Predator)
37.    **for** i $\leftarrow$ 1 **to** nPop **do**
38.      **for** j $\leftarrow$ 1 to nVar **do**
          **// Crossover**
39.        **if** $M_{i,j} > 0$ **the** $x_{i,j} :=$ Predator $_{i,j}$ **end**
          **// Boundary control**
40.        **if** $(x_{i,j} < low_j) \vee (x_{i,j} > up_j)$ **then**
41.          $x_{i,j} :=$ rnd.( $up_j$ - $low_j$) + $low_j$
42.        **end**
43.      **end**
44.    end
       **// Selection**
45.    **for** i $\leftarrow$ 1 to nPop **do**
46.      **if** $f(x_i) < y_{i;\,Predator}$ **then** Predator $_i := x_i$ , $y_{i;\,Predator} := f(x_i)$ **end**
47.    **end**
48.    **if** key = 1 then
49.      $\alpha :=$ Predator, $y_\alpha := y_{Predator}$
50.    **else**
51.      $\beta :=$ Predator, $y_\beta := y_{Predator}$
52.    **end**
53.    $y_{best}$ = min $(y_{Predator})$ | best $\in$ {1, 2, 3,…, nPop}
54.    if $y_{best}$ < globalminimum **then**
55.       globalminimum := $y_{best}$
          globalminimizer := Predator $_{best}$
56.    **end**
57. **end**

### 3.3.5.4 Backtracking Search Algorithm

BSA is one of the most recently proposed evolutionary algorithms with a simple structure but high effectiveness in solving multimodal functions that enables it to easily adapt to different numerical optimization problems. This merit is provided by balancing exploitation of better results and exploration of the problem's search space through use of a single control parameter and two advanced crossover and mutation operators. Its strategy contains two advanced crossover and mutation operators for generating a trial population. These operators are unique and quite different from the structures of the crossover and mutation strategies defined in other evolutionary methods (e.g. GA and DE). BSA's strategies for generating trial populations and controlling the search-space boundaries and adapting the amplitude of the search-direction matrix provide effective exploration and exploitation capabilities.

BSA has been developed to overcome the drawbacks of metaheuristic methods (e.g. too many control parameters and over sensitivity to initial value of these parameters, premature convergence and time-consuming computation), since it has only one control parameter and is not overly sensitive to the initial parameter value. It also possesses a memory that allows it to take advantage of the experience gained from past generations when generating a trial population. In particular, it stores a randomly chosen population from previous generation in its memory for use in generating the search-direction matrix.

Statistical analysis in (Civicioglu, 2013b) confirm that BSA is a promising optimization method for solving high multimodal optimization benchmarks over different well-known evolutionary methods. BSA comprises six stages: initialization, selection-I, mutation, crossover, boundary control, and selection-II as presented in Table. 3.12.

**Table 3.12: General structure of BSA**

```
        1.  Initialization
Repeat
   |    2.  Selection I
   |    Generation of Trial-Population
   |  |    3.  Mutation
   |  |    4.  Crossover
   |    end
   |        5.      Selection II
Until stopping conditions are met;
```

- **Initialization**

This process of BSA initially scatters the population members in the solution space. The initial value of i $^{th}$ individual in the solution space is defined by using Eq. (3.94).

$$
\begin{aligned}
&P_{i,j;g=0} \sim U(low_j, up_j) \quad , y_i = f(P_i) \\
&for \\
&i = \{1,2,3,...nPop\} \, , \, j = \{1,2,3,...nVar\}
\end{aligned}
\tag{3.94}
$$

where:

$nPop$ is population size. $nVar$ is number of respective optimization variable. $U$ represents the uniform distribution function. $low_j$ and $up_j$ are lower and upper search space limits of j$^{th}$ variable. $y_i$ is productivity of i$^{th}$ individual. $g$ is generation number.

- **Selection-I**

In this stage of BSA algorithm, a historical population ($oldP$) which is utilized to determine the search-direction matrix is initialized. The historical population is initialized according to Eq. (3.95).

$$
oldP_{i,j} \sim U(low_j, up_j)
\tag{3.95}
$$

Then the historical population is redefined at each iteration through the 'if-then' decision rule (by comparing two random numbers *a* and *b*) according to Eq. (3.96). Subsequently, in this stage the population (*P*) pursues the historical population for a period of time (until it is changed), which provides a memorization process to facilitate the exploration of search space in this algorithm.

$$if \; a < b \bigg|_{a,b \sim U(0,1)} then \; oldP := P \tag{3.96}$$

where

$:=$ is the update operation.

*a* and *b* are randomly generated numbers.

Finally, the hierarchical sequencing between the individuals of historical population is permuted through a random shuffling function. The obtained *oldP* is used to determine the search-direction matrix in each generation according to Eq. (3.97).

$$oldP := permuting(oldP) \tag{3.97}$$

where:

*permuting (oldP)* is a random shuffling function. As a permuting function, it randomly changes the order of the individuals in historical population.

- **Mutation**

In the mutation step, an initial form of trial population (Mutant) is generated through the Eq. (3.98) while the Wiener process (*F*) is applied in mutation process to controls the amplitude of the search-direction matrix (*oldP - P*).

$$Mutant = P + F.(oldP - P)$$

$$F = 3.rndn \Big|_{rndn \sim N(0,1)} \tag{3.98}$$

where

*N* is standard normal distribution.

- **Crossover**

A binary integer-valued matrix (*map*) that indicates the active individuals of the trial population is determined for crossover process according to Eq. (3.99). In BSA, active individuals stand for the individuals that participate in a crossover at any time and only they can discover new solutions. The control parameter of BSA, determines the probability of corporation between individuals in the crossover process by selecting the active individuals of population. Eq. (3.99) elucidates the unique crossover strategy of BSA.

$$map_{i,j} = 1,$$

$$if \ a < b \Big|_{a,b \sim U(0,1)} \quad then$$

$$map_{i,u_{(1:\lceil mixrate.\, rnd \, . \, nVar \rceil)}} = 0 \Big|_{rnd \sim U(0,1), \ u = permuting(1,2,3,...,nVar)} \tag{3.99}$$

$$else$$

$$map_{i,randi(nVar)} = 0$$

$$T := Mutant$$

$$if \ map_{i,j} = 1 \ then \ T_{i,j} := P_{i,j}$$

$$for$$

$$i = \{1,2,3,...,nPop\}, \ j = \{1,2,3,...,nVar\}$$

where

*mixrate* is the control parameter of optimization algorithm. BSA has only one control parameter that controls the number of individuals to be engaged in the crossover process.

The control parameter of BSA (*mixrate*) varies from 0% to 100% of population size. The experiments with different values of *mixrate* exposed that BSA algorithm is not too sensitive to initial value of its control parameter.

*Permuting(1,2,...,nVar)* is a random permutation that refers to the act of randomly rearranging (reordering) all members.

*randi(nVar)* is a random selection function that generates a pseudorandom integer between 1 and *nVar*.

- **Boundary control**

At the end of the crossover process, if an individual in generated offspring ($T$) violates the boundary condition of the optimization problem, the related individual is updated according to the boundary control mechanism developed in this stage.

$$if\ (T_{i,j} < low_j)\ or\ (T_{i,j} > up_j)\ then\ T_{i,j} \sim U(low_j, up_j) \qquad (3.100)$$

- **Selection II:**

Finally, the fitness value of the best individual is exported as global minimum and its position is considered as global minimizer according to the Eq. (3.101).

$$
\begin{aligned}
&if\ f(T_i) < y_i\ then,\ y_i := f(T_i),\ P_i := T_i \\
&y_g = min(f(P_g)) \\
&if\ y_g < y_{g-1}\ then\ globalminimum := y_g,\ globalminimizer := P_g, \\
&g = g+1 \\
&for \\
&i = \{1,2,3,...,nPop\},\ g = \{1,2,3,...,gMax\}
\end{aligned}
\qquad (3.101)
$$

Eventually, the pseudocode of BSA algorithm is provided in Table 3.13.

**Table 3.13 Pseudocode of BSA**

**Data**: nPop, nVar, Max cycle, low, up, f, mixrate
Result: Globalminimizer | globalminimum = f (globalminimizer)
1. Superorganisms: α, β
   **//Initialization**
2. globalminimum $_{g=0} \approx \infty$
3. **for** i ← 1 **to** nPop **do**
4.    **for** j ← 1 **to** nVar **do**
5.      $P_{i,j}$, oldP $_{i,j} \sim U$ (low $_j$, up $_j$)
6.    **end**
7.    $y_i = f(P_i)$
8. **end**
9. **for** g ← 1 **to** max cycle **do**
      **// Selection-I**
10.    **if** rnd < rnd **then**
11.     oldP := P
12.    **end**
13.    oldP := permuting (pldP)
14.    **Generation of trial population**
        **//Mutation**
15.      mutant = P + 3.(normrnd(0,1)). (oldP - P)
        **//Crossover**
16.     map$_{1:nPop, 1:nVar}$ =1
17.     **for** i←1 **to** nPop **do**
18.       **if** rnd < rnd
19.        **then** map $_{i, u\,(1:\,[mixrate.\,rnd.\,D])}$ = 0 | u =randperm (nVar)
20.        **else** map $_{i,randi(nVar)}$ = 0
21.       **end**
22.     **end**
23.    T := mutant
24.    **for** i ← 1 **to** nPop **do**
25.     **for** j ← 1 **to** nVar **do**
26.      **if** map$_{i,j}$ = 1
27.      **then** $T_{i,j}$ := $P_{i,j}$
28.      **end**
29.     **end**
30.    **end**
31. **end**
       **// Boundary control**
32.    **for** i ← 1 **to** nPop **do**
33.     **for** j ← 1 **to** nVar **do**
34.      **if** $(T_{i,j} <$ low $_j$) ∨ $(T_{i,j} >$ up $_j$) **then**
35.      $T_{i,j}$ := rnd.( up $_j$ - low $_j$) + low $_j$
36.      **end**
37.     **end**
38.    **end**
       **// Selection-II**
39.    **for** i ← 1 **to** nPop **do**
40.     **if** $f(T_i) < y_i$ **then** $P_i$ := $T_i$ , $y_{i;}$ := $f(T_i)$ **end**
41.    **end**
42.    $y_{best}$ = min (f (P)) | best ∈ {1, 2, 3,…, nPop}
43.    **if** $y_{best}$ < globalminimum **then**
44.     globalminimum := $y_{best}$
        globalminimizer := $P_{best}$
45.    **end**
46. **end**

(a) *Multi-objective backtracking search algorithm*

In the context of multi-objective optimization, instead of unique solution, there is a Pareto optimal set corresponding to the optimal value of each objective. Considering the two solution from Pareto optimal set as denoted by $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)$ and $\partial = (\partial_1, \dots, \partial_N)$ and their corresponding objective functions represent by $f(\varepsilon) = (f_1(\varepsilon), \dots, f_m(\varepsilon))$ and $f(\partial) = (f_1(\partial), \dots, f_m(\partial))$. The solution $\partial$ is dominated by the solution $\varepsilon$, denoted by $f(\varepsilon) \prec f(\partial)$, if and only if the conditions described in Eq. (3.102) are satisfied. Therefore, the solution $\varepsilon$ is considered as a non-dominated solution.

$$\forall i \in \{1, \dots, m\}: \ f_j(\varepsilon) \leq f_j(\partial)$$
$$\exists i \in \{1, \dots, m\}: \ f_j(\varepsilon) < f_j(\partial)$$

(3.102)

For two objective functions denoted by $f_1$ and $f_2$, the Pareto optimal set is illustrated in Fig 1. Where the dominated solutions are represented by the gray circles and the Pareto optimal set of two objectives ($f_1$ and $f_2$) are represented by red circuits.
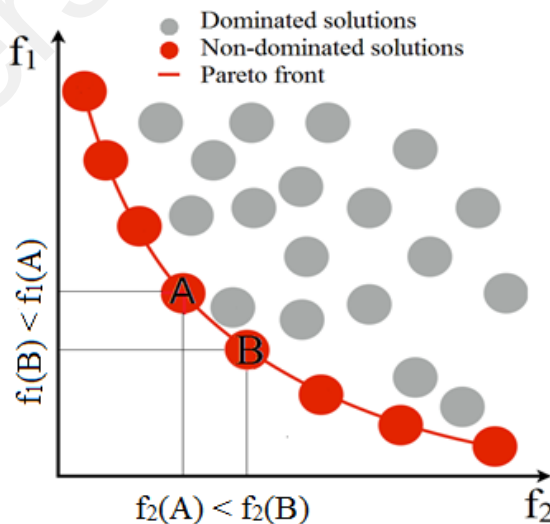


**Figure 3.8: The Pareto optimal set for the two objective functions (A and B are two sample from non-dominated solutions)**

- **External elitist archive**

In the context of evolutionary multi-objective optimization, finding non-dominated solutions within evolutionary process have been retained for different methods, so this sorting method in evolutionary multi-objective optimization algorithms have been enhanced over the last years. Initially the non-dominated sorting strategy proposed as an efficient selection strategy in multi-objective optimization (Holland & Goldberg, 1989). Later, Deb et al. (Srinivas & Deb, 1994) applied non-dominated sorting strategy in genetic algorithm and proposed non-dominated sorting genetic algorithm (NSGA) to solve the multi-objective optimization problems. Afterwards, Deb et al. (Deb, Pratap, Agarwal, & Meyarivan, 2002) developed a more efficient non-dominated sorting strategy as an enhanced variant of NSGA known as fast non-dominated sort which called NSGA-II. Necessity to specify the value of sharing parameter ($\sigma_{share}$), high computational complexity of non-dominated sorting, and lack of elitism are the main criticisms of NSGA approach, which have been solved in NSGA-II. The two novel methods for non-dominated sorting known as deductive sort and climbing sort were presented in (McClymont & Keedwell, 2012). The results demonstrate that deductive sort outperforms the fast non-dominated sort of NSGA-II. Another method for non-dominated sorting named efficient non-dominated sorting (ENS) strategy was proposed in (X. Zhang, Tian, Cheng, & Jin, 2015) that for searching within ENS, two different strategies including binary search (BS) and sequential search (SS) are employed. Although the results show that both ENS-based non-dominated sorting strategies are more efficient than other sorting approaches, the efficiency of the ENS-based sorting strategies decreases as the number of objectives increases (Modiri-Delshad & Rahim, 2016).

Using an elitism strategy to store the non-dominated solutions found within the optimization process is another mechanism proposed in (Zitzler & Thiele, 1999) to update and retain the non-dominated solutions. The main motivation for using elitist reservation

mechanism is that a result that is non-dominated to other results in its current generation is not necessarily non-dominated with respect to all other results which optimization algorithm has found so far (Coello, 2006).

An elitist reservation mechanism is adopted in this study as an external elitist archive to update and retain the non-dominated solutions in each generation of BSA. Initially the external elitist archive is empty; within the optimization progresses, it stores the non-dominated solutions according to the following 'if-then' rules:

i-    If the trial pattern (a new generated solution) dominates some of the archived elitist then all dominated members of the external elitist archive are replaced by non-dominated trial pattern;

ii-   If the trial pattern is dominated by at least one member of the external elitist archive then the trial pattern is disregarded for elitist archive;

iii-  If the archived members of the external elitist archive are not dominated by the trial pattern and trial pattern is not dominated by the archived members then the external elitist archive retains the trial pattern as a new elitist member (non-dominated solution)

As the optimization progresses, the members of external elitist archive increases. Therefore, to prevent overpopulation of the external elitist archive the crowding distance of all members are measured and the extra members of the archive are removed according to their crowding distance value.

- **Crowding distance**

To keep the external elitist archive to its maximum capacity the crowding distances (CD) of all solutions in Pareto-front (external elitist archive) are computed and the solution with the lowest CD value is subject to deletion when the archive is overloaded.

The crowding distance is a factor to evaluate the distribution of the solutions in Pareto -front by measuring the density around a solution. The crowding distances computes the distance of two neighbor points around the solution. The crowding distances of $i^{th}$ solution in the Pareto -front is calculated by Eq. (3.103).

$$CD_i = \sum_{j=1}^{m} \frac{f_j(i+1) - f_j(i-1)}{f_j^{max} - f_j^{min}} \tag{3.103}$$

where, $CD_i$ is crowding distance of solution $i$. $f_j$ is the $j^{th}$ objective function. $f_j^{max}$ and $f_j^{min}$ are maximum and minimum values of the $j^{th}$ objective function respectively. $m$ is number of objectives.

For the boundaries solutions ($f_j^{max}$ and $f_j^{min}$) the crowding distance is set to infinite, as there is only one neighbor point for those solutions.

- **Procedure of multi-objective BSA (non-dominated approach)**

In the context of multi-objective optimization, instead of unique solution, there is a Pareto optimal set corresponding to the optimal value of each objective. Thus, to extend BSA to a multi-objective optimization approach, the replacement mechanism is adapted according to the concept of Pareto dominance (Modiri-Delshad & Rahim, 2016). Similar to BSA, in multi-objective BSA mutation and crossover operators are first applied to produce the offspring ($T$). Then the comparison in the final step of BSA (export the global minimum) is modified according to the concept of Pareto dominance to replace $i^{th}$ individual of population ($P_i$) by $i^{th}$ individual in offspring ($T_i$) if the $P_i$ is dominated by $T_i$. The consecutive steps in BSA algorithm are followed to form the multi-objective BSA, expect the last step (export the global minimum). Instead of exporting a global minimum, a Pareto optimal set is stored. The following steps represent the sequential procedure of multi-objective BSA with an external elitist archive and crowding distance measure.

**Step 1**: The initial population (P) equal to the dimension of the optimization variables is randomly generated according to Eq. (3.94).

**Step 2**: Evaluate the fitness function of each individual of initial population and store the non-dominated solutions among the population into the external elitist archive.

**Step 3**: The historical population is determined randomly according to Eq. (3.95).

**Step 4**: The historical population is updated at each iteration through the 'if-then' decision rule according to Eq. (3.96)

**Step 5**: Apply the mutation operator to generate an initial form of trial population (Mutant) according to Eq. (3.98).

**Step 6:** Apply the crossover operator over the initial form of trial population obtained in pervious step to generate the final form of offspring ($T$) according to Eq. (3.99).

**Step 7:** At the end of the crossover process, if an individual in the generated offspring ($T$) violates the boundary condition of the optimization problem, the related individual in offspring is updated according to the boundary control developed by Eq. (3.100).

**Step 8**: If the i$^{th}$ element of generated offspring ($T_i$) dominates the i$^{th}$ element of population ($P_i$), then $T_i$ is replaced by $P_i$.

**Step 9:** Update the external elitist archive according to aforementioned 'if-then' rules.

**Step 10**: If the external elitist archive exceeds its maximum capacity, the crowing distance for all members of elitist archive are computed and the less crowded solutions are removed from the archive one after another.

**Step 11:** If the stopping criteria are not satisfied, set $g=g+1$ and return to the step 4.

# CHAPTER 4: LONG-TERM ELECTRICAL ENERGY CONSUMPTION FORMULATING AND FORECASTING

## 4.1    Introduction

In long-term energy modeling process, energy consumption of a country is considered as a function of diverse socio-economic indicators (SEI). As demonstrated in Table 2.1, the energy consumed in a specific year ($t$) depends not only on the values of socio-economic indicators in that year but also on the energy consumed in the previous years. Thus, two types of input historical data from 1971 to 2011 inclusive, which are expressed as follows, are considered in this study to mathematically formulate the long-term EEC of ASEAN-5 countries via optimized GEP and metaheuristic methods.

> i-      Socio-economic indicators (SEI) that electric energy consumption of a country is mostly affected/reflected by them, namely, gross domestic product (GDP), population (POP), import (IMP), and export of goods and services (EXP), total energy consumption (TEC), price of energy (POE), carbon dioxide emissions (CDE), urbanization rate (UR), industrialization rate (IR) and stock index (SI).

> ii-     EEC in preceding ten years (EEC ($t$-$1$), EEC ($t$-$2$), EEC ($t$-$3$), …, EEC ($t$-$10$)) where it is assumed as a time series with annual interval.

## 4.2    Long-term Electrical Energy Consumption Formulating

The input historical data sets of ASEAN-5 countries in the last 41 years (1971-2011) are adapted from the WB data bank. Since the range of historical data varies widely, both dependent and independent variables are normalized according to Eq. (4.1). The main feature of data normalization is adjusting raw data observed on different scales to a notionally common scale, often prior to data processing.

$$\overline{Z}(t) = \frac{Z(t) - min(Z)}{max(Z) - min(Z)} + 1 \qquad (4.1)$$

where $\overline{Z}$ is the normalized value, $Z$ is the value to be normalized, and $t$ is annual interval.

If all the aforementioned exogenous variables are applied in the modeling process, this will not only slowing down the learning process but also giving rise to a poor performance and overfitting the training data. In other words, although these factors are of vital importance to long-term energy modeling process, only those features exhibiting significant influence on the output should be picked.

The total possible combinations of $k$ objects taken from a list of $S$ at a time without repetition can be calculated by:

$$\text{The total possible combinations} = \sum_{k=1}^{S} \binom{S}{k} = \sum_{k=1}^{S} \frac{S!}{(S-k)!k!} \qquad (4.2)$$

Assuming all exogenous variables are used to mathematically formulate the EEC of ASEAN-5 countries ($S=10$), according to Eq. (4.2) there are 1023 different combinations of variables for each historical input data set. Though all subsets of variables have causal relation with EEC, it is neither efficient nor feasible to employ all of them as inputs and assess their performances. Instead, an efficient feature selection technique is applied to select the most effective subset of variables in model construction.

In the context of statistics and machine learning, feature (variables, predictors) selection, also known as attribute selection, variable selection or variable subset selection is a method for selecting a subset of relevant features in model construction. The objective of using feature selection techniques is three-fold:

i-      Improving the prediction performance of the predictors by reducing overfitting (formally, reduction of variance).

ii-      Providing faster and more cost-effective process to construct the model (facilitate learning process).

iii-      Providing a simplified model that make it easier to interpret (improving the generalization ability).

The main objective of a feature selection technique is to evaluate relevancy and redundancy of the input features for selecting the best subset of features representing the most important information of the original feature set. A large number of predictors could result in inferior performance of extracted models due to the course of dimensionality principle. Removing either redundant or irrelevant features from a data set contains many features without incurring much loss of predictive accuracy is the central premise for using a feature selection technique.

A search strategy along with an evaluation metric is applied in a feature selection algorithm to search candidate subsets of feature and score the performance of these candidates respectively.

Testing all possible feature subsets to find one that minimizes the error rate is the simplest feature selection algorithm that provides an exhaustive search of the feature space, but it is computationally intractable (Unler & Murat, 2010). Thus, the combination of a search strategy to explore the space of all possible combination of features along with an evaluation measure to assess the quality of features heavily influences on the effectiveness of a feature selection algorithm.

According to how a search technique is combined with a learning algorithm (evaluation metric) for construction of the models forms three classes of feature selection methods, namely wrappers, filters, and embedded (Renani, Elias, & Rahim, 2016).

Wrapper (search guided by accuracy) methods use a predictive model to score feature subsets. In wrapper methods, each candidate feature subsets is used to train a model that is tested on a holdout set. The score for each candidate subset is provided by evaluating the error rate of the model on testing set. As wrapper methods train a new predictive model for each candidate subsets, they often provide the best performing feature set for that particular type of model at the expense of computationally intensive tasks.

In filter (information gain) methods, instead of the error rate to score a candidate subset of feature a proxy measure is used. The proxy measures such as the pointwise mutual information, Pearson product-moment correlation coefficient and mutual information are chosen to provide fast computation for capturing the effectiveness of the feature set.

The wrapper methods are higher computationally intensive than filter methods, but wrapper methods provide a subset of feature that its performance is evaluated by specific type of learning algorithm. Due to the lack of learning algorithm in filters, a feature set from the filter methods is usually more general and giving lower prediction performance than the set from wrapper methods. The filter methods are widely used to expose the relationships between the variables and provide a feature ranking rather than an explicit best feature subset. So, a filter can be used as a preprocessing step for a wrapper to form a hybrid feature selection method. In this hybrid method, filter works as dimensionality reduction method allowing a wrapper method to be used for appropriate selection of the most relevant features on larger data sets.

Embedded methods learn which subset of features has best contribute to the accuracy of the model while the model is constructed. In embedded methods, the feature selection part and training process cannot be distinguished, as selection of the features and model construction procedures are performed simultaneously. Although, the embedded methods are less computationally intensive than wrapper methods, the main drawback of these

methods is that the selected features are sensitive to the structure of the underlying model so embedded methods are usually specific to their learning algorithms. Different types of embedded method are classification trees, random forests, and regularization approaches. Regularization approaches also known as penalization approaches are the most common type of embedded feature selection methods. The penalization approaches add additional constraints into the model construction process, which bias the model toward simplicity by penalizing the model for higher complexity.

Wrapper methods often provide the most relevant features for particular type of model, but needing a systematic searching algorithm in their evolutionary training process. Since, the exhaustive search is generally impractical, the sequential search as a heuristic technique to feature selection has been proposed to add or remove features sequentially until there is no improvement in the accuracy of model. The sequential search has a tendency to become stagnated in local optima. Thus, the randomized search algorithm known as metaheuristic algorithm is proposed to explore the space of different feature subsets. Metaheuristic algorithms incorporating randomness into search procedure of feature selection to escape local optima. GA, SA, ACO and PSO are different metaheuristic algorithms that have been applied in the context of feature selection (Unler & Murat, 2010).

These metaheuristic algorithms deal with feature selection as a single objective optimization problem, so the number of relevant features should be predefined and always find the subset of features with fixed number of features. Generally, feature selection has two main conflicting objectives, which are minimizing simultaneously both the estimation error and the number of features. Therefore, feature selection problem can be expressed as a multi-objective problem that has two main objectives, maximizing the accuracy of model and minimizing the number of features whereby the decision is a

tradeoff between these two objectives. Treating feature selection as a multi-objective problem leads to a set of non-dominated feature subsets to meet different requirements in real-world applications.

Multi-objective particle swarm optimization (MOPSO) and NSGA-II have been investigated in (Xue, Zhang, & Browne, 2013, 2014) to generate a Pareto front of feature subsets, but still more efficient searching strategy is essential to better address feature selection problems (C. Zhang, Zhou, Li, Fu, & Peng, 2017). Existing multi-objective feature selection algorithms suffer from the problems of high computational cost, too many control parameters, and over sensitivity to initial value of these parameters, BSA with only one control parameter is argued computationally less expensive than other metaheuristic algorithms (Modiri-Delshad & Rahim, 2016). A binary-valued BSA (BBSA) has been proposed by (M. S. Ahmed et al., 2017) to solve the optimization of discrete parameters. In BBSA, the individuals in the population are encoded as a binary vector, and the population value is converted to zero or one according to Eq. (4.3).

$$S_{i,j} = \frac{1}{1+e^{-w}}$$

$$BP_{i,j} = \begin{cases} 0 & if \ s_{i,j} < 0.5 \\ 1 & if \ s_{i,j} \geq 0.5 \end{cases}$$

(4.3)

where $S$ is the sigmoid function, $PB$ is its binary value and $w$ is the population value.

However, BBSA has been exploited as an efficient searching algorithm for feature selection in (C. Zhang et al., 2017). It treats the task as a single objective problem and it could not directly be used to address multi-objective feature selection problems. Different versions of multi-objective BSA (MOBSA) have been developed in (Modiri-Delshad & Rahim, 2016; Zou, Chen, Li, Lu, & Lin, 2017). Statistical analyses in (Modiri-Delshad & Rahim, 2016) confirm that MOBSA is a promising optimization method for solving high dimensional multi-objective problems over different well-known multi-objective

evolutionary algorithms (e.g. MOPSO and NSGA-II). Therefore, in this study BBSA-based multi-objective feature selection algorithm is developed as a promising technique to generate a Pareto front of non-dominated feature subsets.

To assess the performance of each candidate feature subsets, any learning algorithm (e.g. ANN, SVM, ANFIS) can be used in the evolutionary training process of BSA-based multi-objective feature selection algorithm. ANFIS is considered as a universal estimator due to its fast learning capability to approximate nonlinear functions (Tavana et al., 2016); hence, it is adopted as an evaluation metric method in proposed multi-objective wrapper-based feature selection method. In particular, ANFIS employs an efficient hybrid learning method that combines the least squares method and gradient descent. The least square method is main factor for quick training (Alizadeh, Jolai, Aminnayeri, & Rada, 2012). Thus, after only few epoch of training, ANFIS is able to construct the predictive model. Since the least square method is computationally efficient, the models are constructed for various combinations of features selected by multi-objective BBSA (MOBBSA), train them with single/few running of the least-squares method, and then a non-dominated subset of features with the best performance is chosen for constructing the model.

Before applying the feature selection to extract the most influential subsets of input variables with maximum relevancy and minimum redundancy for long-term EEC modeling, both dependent and independent variables are randomly divided into two sets: 70% as the training set and 30% as the test set. The training set is used to construct the ANFIS models with different subsets of input variables, while the test set is used to access the strength and utility of generated models.

In the developed multi-objective feature selection method, MOBBSA is used to search within 1023 different combinations of input variables and selects the non-dominated feature subsets, while ANFIS is applied as evaluation metric to determine the

performance of each feature subset. During the training process of applied learning method (ANFIS), each individual of MOBBSA represents one input variable. The developed feature selection strategy uses an elitist external archive to store non-dominated feature subsets, which simultaneously minimize both the root mean square error (RMSE) on the test set and the number of input variables as the optimal solution set according to the concept of Pareto dominance. MOBBSA as an extension of BSA has only one control parameter named "mixrate", which controls the number of individuals to be engaged in the crossover process. The population of individuals is set to 50 and the maximum value of mixrate (i.e. 100% of the population size) is considered in the developed feature selection strategy to engage all the individuals in the crossover process.

To form the structure of ANFIS for feature selection, the Sugeno-type FIS is used as a promising alternative to Mamdani-type, since the Sugeno-type is well suited for modeling nonlinear systems by interpolating between multiple linear models. Based on the work conducted in (J.-S. Jang, 1996), the scatter partitioning can be used to facilitate the training process of ANFIS in feature selection. Since subtractive clustering is a realization of scatter partitioning, it is used to set up the ANFIS for feature selection.

The optimal subsets of input variables selected by developed multi-objective feature selection for modeling the EEC of ASEAN-5 countries are illustrated in Figures 4.1- 4.5. According to the obtained results, regardless of which country is studied, a subset of SEI with four variables and a subset of EEC with three variables dominate all other optimal subsets. This implies that these obtained subsets as non-dominated solutions provide least RMSE with minimum number of input variables. The obtained non-dominated SEI and EEC subsets consist of GDP, POP, IMP and EXP, and EEC in preceding three years (EEC *(t-1)*, EEC *(t-2)*, and EEC *(t-3)*) as input variables respectively.
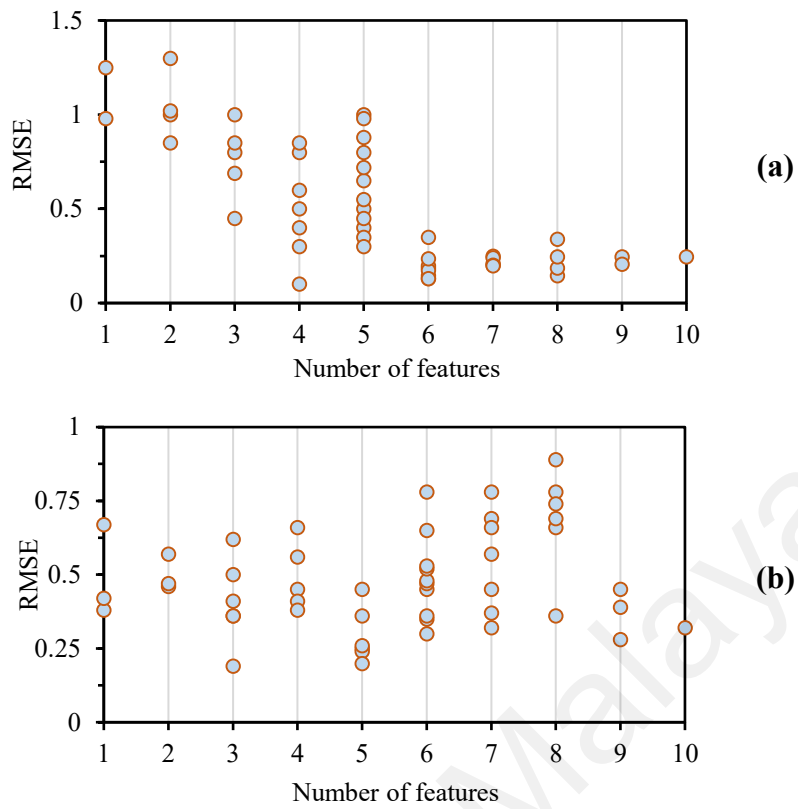
Figure 4.1: The optimal subsets of input variables for EEC modeling selected via MOBBSA feature selection from two different input historical data sets: (a) Malaysia's SEI, (b) Malaysia's EEC
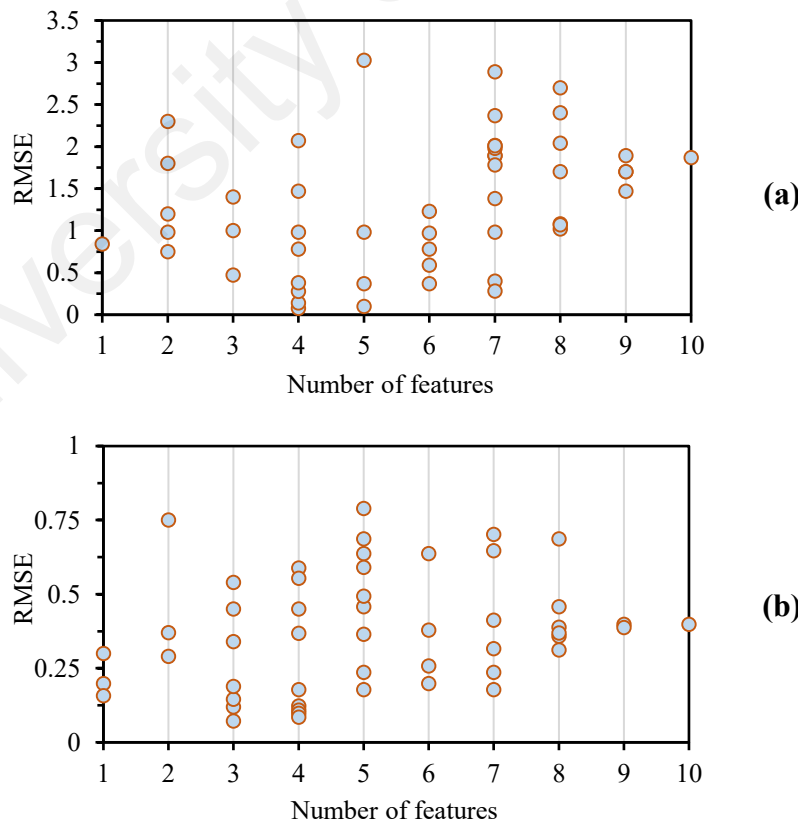


Figure 4.2: The optimal subsets of input variables for EEC modeling selected via MOBBSA feature selection from two different input historical data sets: (a) Indonesia's SEI, (b) Indonesia's EEC
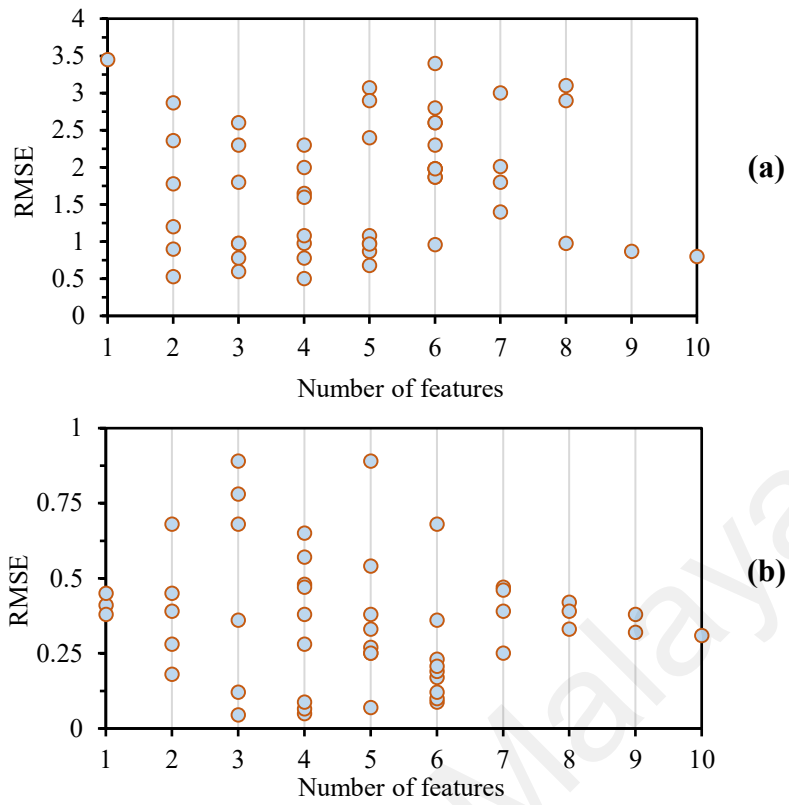
**Figure 4.3: The optimal subsets of input variables for EEC modeling selected via MOBBSA feature selection from two different input historical data sets: (a) Singapore's SEI, (b) Singapore's EEC**
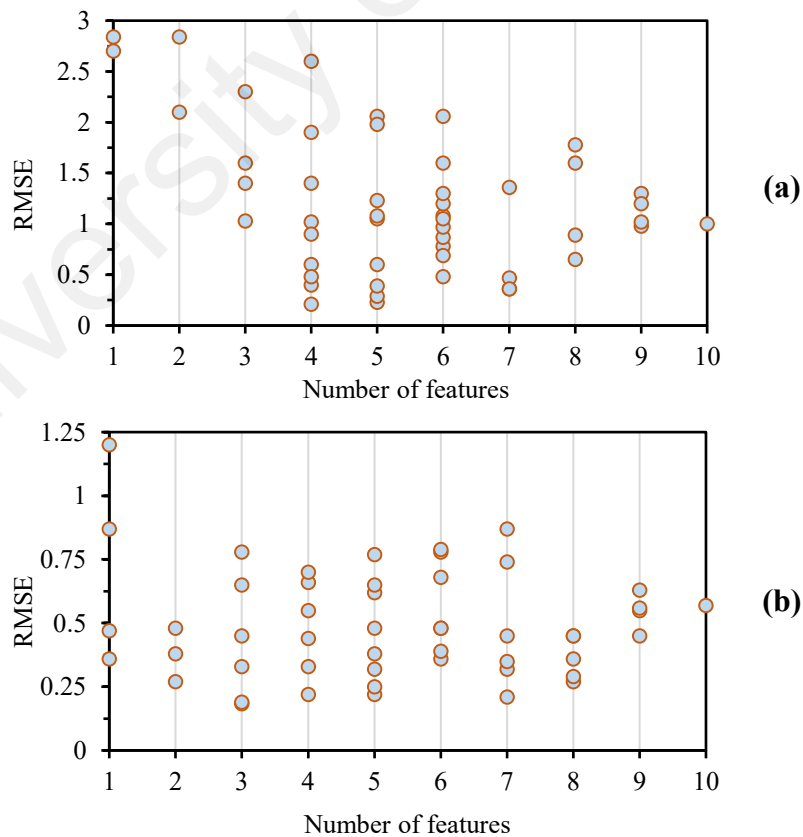


**Figure 4.4: The optimal subsets of input variables for EEC modeling selected via MOBBSA feature selection from two different input historical data sets: (a) Thailand's SEI, (b) Thailand's EEC**
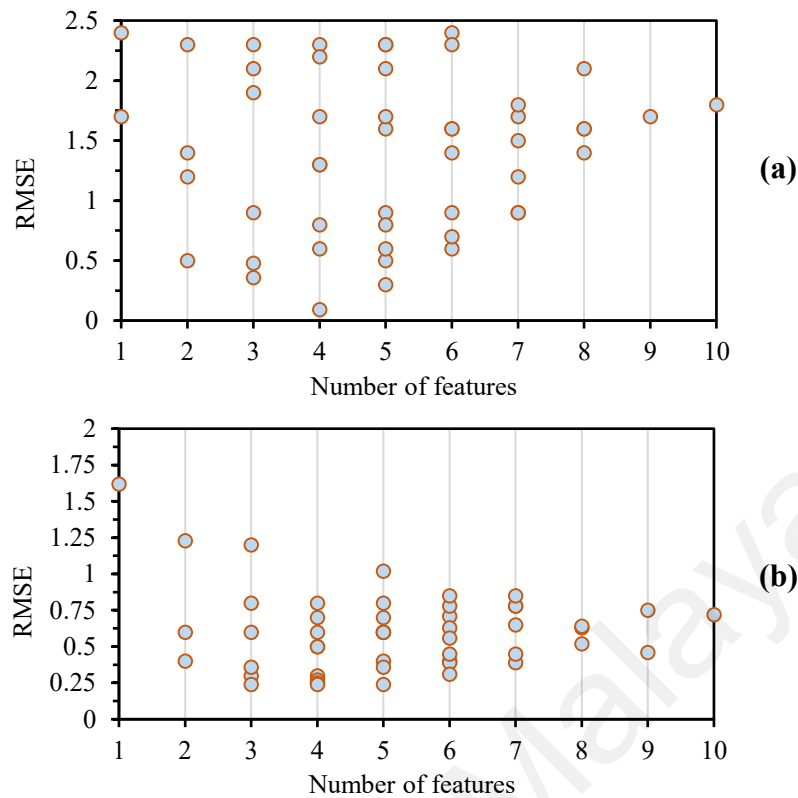
**Figure 4.5: The optimal subsets of input variables for EEC modeling selected via MOBBSA feature selection from two different input historical data sets: (a) Philippines's SEI, (b) Philippines's EEC**

The effects of two input historical data types (i.e. SEI and EEC in preceding three years) on electricity consumption of ASEAN-5 countries are more visible by the correlation coefficients. A correlation coefficient is a coefficient that elucidates a quantitative measure of some type of correlation and dependence, meaning a predictive relationship between two observed data that can be exploited in practice.

The Pearson product-moment correlation coefficient (PPMCC) denoted by 'r' and Spearman's rank correlation coefficient denoted by '$r_s$' are two different correlation analyses that are widely used in science to assess how well the relationship between two variables can be described by a linear or a monotonic function respectively (Zahedi et al., 2013). A correlation between two variables, however, does not imply causation mean that the change in one variable is the cause of the change in the values of the other variable.
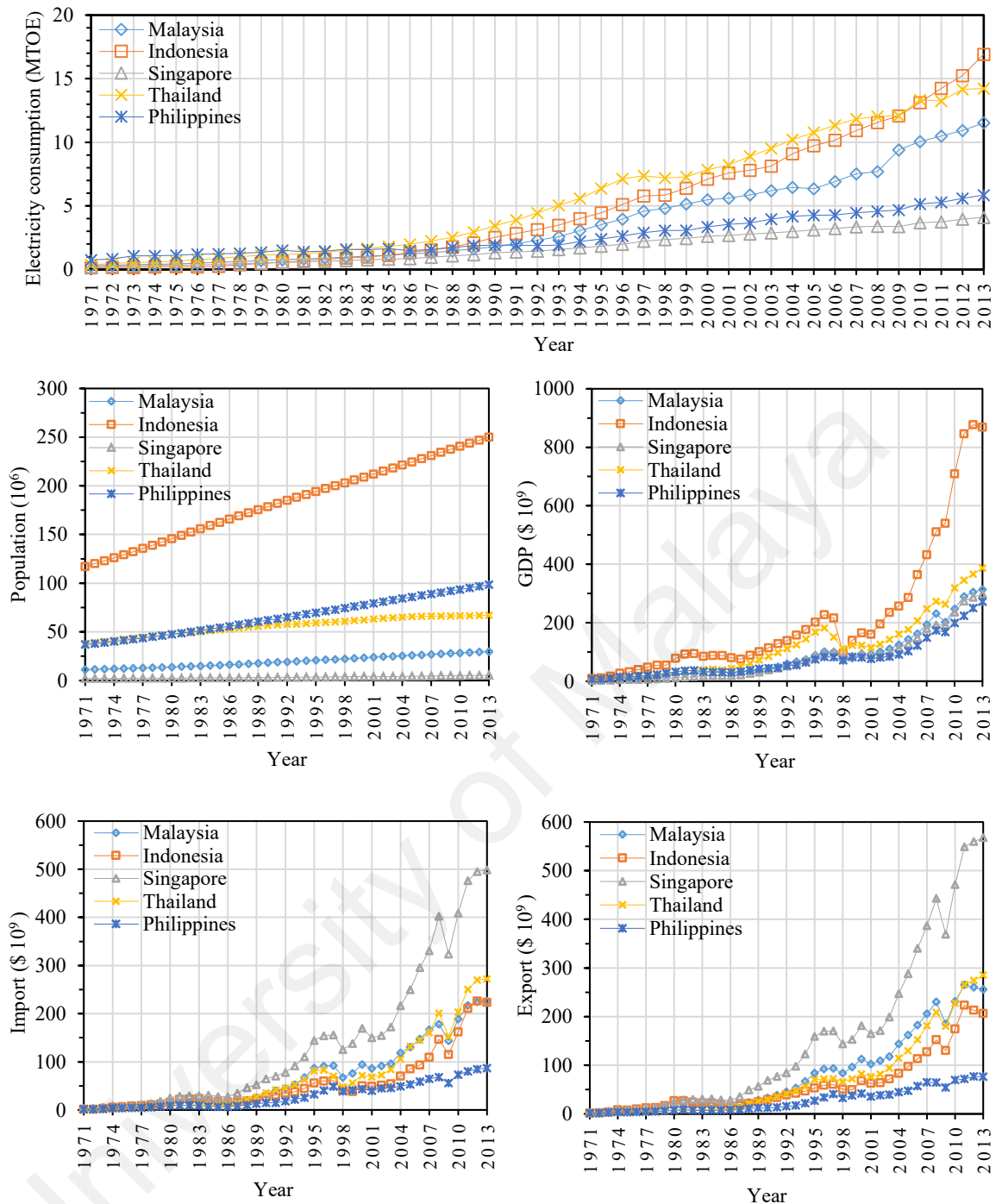
**Figure 4.6: Historical EEC and socio-economic indicators data of ASEAN-5 countries for 1971-2013**

The PPMCC describes both the strength and the direction of the linear association between two variables. It takes a range of values from +1 to -1. Where $r=+1$ indicates that there is total positive correlation, $r=0$ indicates that there is no correlation, and $r=-1$ indicates that there is total negative correlation. Pearson correlation coefficient between two variables $x$ and $y$ is calculated by using Eq. (4.4) where $N$ is total number of observations.

$$r = r_{xy} = \frac{N\sum_{t=1}^{N} x(t)y(t) - \left(\sum_{t=1}^{N} x(t) \sum_{t=1}^{N} y(t)\right)}{\sqrt{N\sum_{t=1}^{N} x(t)^2 - \left(\sum_{t=1}^{N} x(t)\right)^2} \sqrt{N\sum_{t=1}^{N} y(t)^2 - \left(\sum_{t=1}^{N} y(t)\right)^2}} \qquad (4.4)$$

As PPMCC measures only a linear relationship, still a meaningful relationship can exist even if the correlation coefficient is close to zero. So Spearman's rank correlation coefficient is used to measure a monotonic relationship between two observed data. In the monotonic relationships, the variables tend to change together, but not necessarily at a constant rate (linear relationship). The Spearman's coefficient is defined as the PPMCC between the ranked values for each variable rather than the raw data.

The measure of these two coefficients between input historical data and EEC of ASEAN-5 countries are tabulated in Table 4.1. As the results show, although there are strong linear relationships between independent and dependent variables, the monotonic functions can describe the relationships between historical data and EEC of ASEAN-5 countries in a more appropriate way.

**Table 4.1: PPMCC and Spearman's rank correlation coefficient between EEC of ASEAN-5 countries and two types of input historical data**

| Input historical data | Correlation coefficient | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $r$ | $r_s$ | $r$ | $r_s$ | $r$ | $r_s$ | $r$ | $r_s$ | $r$ | $r_s$ |
| | Malaysia | | Indonesia | | Singapore | | Thailand | | Philippines | |
| POP | 0.9646 | 0.9998 | 0.9458 | 0.9998 | 0.9939 | 0.9987 | 0.9217 | 0.9993 | 0.9681 | 0.9944 |
| GDP | 0.9632 | 0.9902 | 0.9037 | 0.9655 | 0.9413 | 0.9904 | 0.9509 | 0.9705 | 0.9441 | 0.9796 |
| IMP | 0.9758 | 0.9862 | 0.9284 | 0.9672 | 0.9365 | 0.9897 | 0.9436 | 0.9850 | 0.9798 | 0.9789 |
| EXP | 0.9789 | 0.9963 | 0.9516 | 0.9722 | 0.9346 | 0.9907 | 0.9499 | 0.9973 | 0.9820 | 0.9796 |
| EEC$(t-1)$ | 0.9967 | 0.9996 | 0.9992 | 0.9996 | 0.9988 | 0.9998 | 0.9980 | 0.9986 | 0.9883 | 0.9964 |
| EEC$(t-2)$ | 0.9936 | 0.9994 | 0.9982 | 0.9994 | 0.9978 | 0.9996 | 0.9959 | 0.9984 | 0.9867 | 0.9927 |
| EEC$(t-3)$ | 0.9900 | 0.9994 | 0.9976 | 0.9994 | 0.9963 | 0.9994 | 0.9924 | 0.9983 | 0.9865 | 0.9900 |

Table 4.1 confirms, social-economic indicators of ASEAN-5 countries (i.e. POP, GDP, IMP, and EXP) have high impacts on EEC, and POP is the most effective one as expected. As the correlation coefficients corresponding to POP demonstrate, there is a strong relationship between population and electricity consumption. The population growth leads to higher demand for social services, health care, and education so higher EEC is required to provide them.

Moreover, it is obvious that for increasing the GDP of ASEAN-5 counties higher EEC is required. Electricity is the backbone of a nation's progress. A country's GDP is the sum of its industrial, agricultural, and services output. The services and agriculture sectors are light to moderate electricity users. The industrial sector is very dependent on electricity. When a business flourishes, the EEC increases. As a result, the relationship between GDP and EEC is higher for the countries that have high exposure to the industrial sector in the GDP. In addition, the rate of IMP and EXP of goods and services in any country typically indicate the industrial activities. Thus, these two indicators are determined as effective factors on GDP (see Tables 4.2 - 4.6) and consequently on EEC of ASEAN-5 counties. The measures of PPMCC and Spearman's rank correlation coefficient between SEI of ASEAN-5 countries are reported in Table 4.2 – 4.6.

**Table 4.2: PPMCC and Spearman's rank correlation coefficient between SEI of Malaysia**

| Correlation coefficient | | $r$ | $r_s$ | $r$ | $r_s$ | $r$ | $r_s$ | $r$ | $r_s$ |
|---|---|---|---|---|---|---|---|---|---|
| | | POP | | GDP | | IMP | | EXP | |
| $r$ | POP | 1 | | 0.9151 | | 0.9478 | | 0.9411 | |
| $r_s$ | | | 1 | | 0.9904 | | 0.9864 | | 0.9965 |
| $r$ | GDP | 0.9151 | | 1 | | 0.9862 | | 0.9879 | |
| $r_s$ | | | 0.9904 | | 1 | | 0.9951 | | 0.9942 |
| $r$ | IMP | 0.9478 | | 0.9862 | | 1 | | 0.9958 | |
| $r_s$ | | | 0.9864 | | 0.9951 | | 1 | | 0.9918 |
| $r$ | EXP | 0.9411 | | 0.9879 | | 0.9958 | | 1 | |
| $r_s$ | | | 0.9965 | | 0.9942 | | 0.9918 | | 1 |

**Table 4.3: PPMCC and Spearman's rank correlation coefficient between SEI of Indonesia**

| Correlation coefficient | | r | $r_s$ | r | $r_s$ | r | $r_s$ | r | $r_s$ |
|---|---|---|---|---|---|---|---|---|---|
| | | POP | | GDP | | IMP | | EXP | |
| r | POP | 1 | | 0.8096 | | 0.8415 | | 0.8645 | |
| $r_s$ | | | 1 | | 0.9656 | | 0.9674 | | 0.9726 |
| r | GDP | 0.8096 | | 1 | | 0.9889 | | 0.9832 | |
| $r_s$ | | | 0.9656 | | 1 | | 0.9888 | | 0.9785 |
| r | IMP | 0.8415 | | 0.9889 | | 1 | | 0.9946 | |
| $r_s$ | | | 0.9674 | | 0.9888 | | 1 | | 0.9832 |
| r | EXP | 0.8645 | | 0.9832 | | 0.9946 | | 1 | |
| $r_s$ | | | 0.9726 | | 0.9785 | | 0.9832 | | 1 |

**Table 4.4: PPMCC and Spearman's rank correlation coefficient between SEI of Singapore**

| Correlation coefficient | | r | $r_s$ | r | $r_s$ | r | $r_s$ | r | $r_s$ |
|---|---|---|---|---|---|---|---|---|---|
| | | POP | | GDP | | IMP | | EXP | |
| r | POP | 1 | | 0.9567 | | 0.9465 | | 0.9438 | |
| $r_s$ | | | 1 | | 0.9883 | | 0.9874 | | 0.9888 |
| r | GDP | 0.9567 | | 1 | | 0.9929 | | 0.9931 | |
| $r_s$ | | | 0.9883 | | 1 | | 0.9952 | | 0.9947 |
| r | IMP | 0.9465 | | 0.9929 | | 1 | | 0.9995 | |
| $r_s$ | | | 0.9874 | | 0.9952 | | 1 | | 0.9996 |
| r | EXP | 0.9438 | | 0.9931 | | 0.9995 | | 1 | |
| $r_s$ | | | 0.9888 | | 0.9947 | | 0.9996 | | 1 |

**Table 4.5: PPMCC and Spearman's rank correlation coefficient between SEI of Thailand**

| Correlation coefficient | | r | $r_s$ | r | $r_s$ | r | $r_s$ | r | $r_s$ |
|---|---|---|---|---|---|---|---|---|---|
| | | POP | | GDP | | IMP | | EXP | |
| r | POP | 1 | | 0.8640 | | 0.8126 | | 0.8108 | |
| $r_s$ | | | 1 | | 0.9682 | | 0.9837 | | 0.9966 |
| r | GDP | 0.8640 | | 1 | | 0.9793 | | 0.9726 | |
| $r_s$ | | | 0.9682 | | 1 | | 0.9897 | | 0.9738 |
| r | IMP | 0.8126 | | 0.9793 | | 1 | | 0.9947 | |
| $r_s$ | | | 0.9837 | | 0.9897 | | 1 | | 0.9883 |
| r | EXP | 0.8108 | | 0.9726 | | 0.9947 | | 1 | |
| $r_s$ | | | 0.9966 | | 0.9738 | | 0.9883 | | 1 |

**Table 4.6: PPMCC and Spearman's rank correlation coefficient between SEI of Philippines**

| Correlation coefficient | | r | $r_s$ | r | $r_s$ | r | $r_s$ | r | $r_s$ |
|---|---|---|---|---|---|---|---|---|---|
| | | POP | | GDP | | IMP | | EXP | |
| r | POP | 1 | | 0.9073 | | 0.9559 | | 0.9519 | |
| $r_s$ | | | 1 | | 0.9815 | | 0.9789 | | 0.9871 |
| r | GDP | 0.9073 | | 1 | | 0.9536 | | 0.9619 | |
| $r_s$ | | | 0.9815 | | 1 | | 0.9954 | | 0.9865 |
| r | IMP | 0.9559 | | 0.9536 | | 1 | | 0.9957 | |
| $r_s$ | | | 0.9789 | | 0.9954 | | 1 | | 0.9895 |
| r | EXP | 0.9519 | | 0.9619 | | 0.9957 | | 1 | |
| $r_s$ | | | 0.9871 | | 0.9865 | | 0.9895 | | 1 |

By comparing the correlation coefficients of ASEAN-5 countries in Table 4.1, it is concluded that the Spearman's rank correlation coefficients corresponding to POP in Malaysia and Indonesia are almost equal and higher than that in other countries, which imply that EEC in these countries are more sensitive to population growth as compared with other countries. The sensitivity of EEC to population growth in ASEAN-5 countries is ranked as Malaysia = Indonesia > Thailand > Singapore> Philippines. In addition, the impact of GDP on EEC in Singapore is more than that in other countries, which highlights that Singapore has highest exposure to the industrial sector in the GDP. The effect of economic growth in terms of GDP on EEC growth is ranked as Singapore > Malaysia > Philippines > Thailand > Indonesia. Further, the impacts of IMP and EXP on EEC are ranked as Singapore > Malaysia > Thailand > Philippines > Indonesia, and Thailand > Malaysia > Singapore > Philippines > Indonesia respectively. According to Table 4.1, while the EEC is considered as a time series, the sensitivity of EEC to electricity consumption in preceding years is reduced as time passes. So the strength of relationships between EEC and EEC in preceding three years is ranked as EEC *(t-1)* > EEC *(t-2)* > EEC *(t-3)*.

As shown in Figures 4.7 – 4.11, despite almost linear trend in population growth of ASEAN-5 countries, the electricity consumption of ASEAN-5 countries have not been continuously growing during last decades. Due to numerous perturbations on economy of ASEAN-5 countries, the growth rate of electricity consumption in these countries have been volatile. Revival of the economic activities and the positive economic growth are directly reflected in GDP, IMP and EXP hence, they are considered as barometers of the economy. Since the growth rate of main economy indicators (GDP, IMP and EXP) have been highly volatile, their violation have been reflected on electricity consumption in the meanwhile.
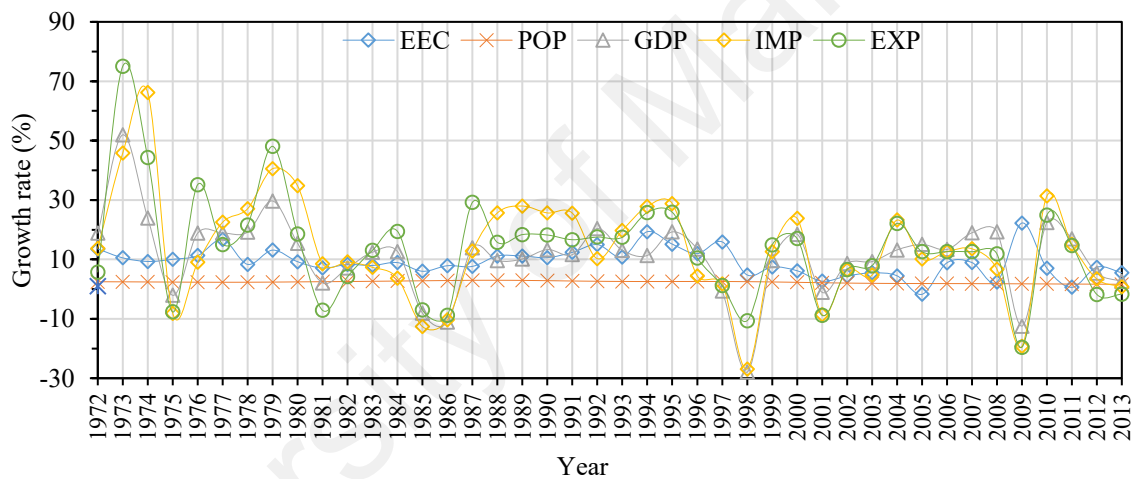


**Figure 4.7: Growth rate of Malaysia's EEC and SEI**
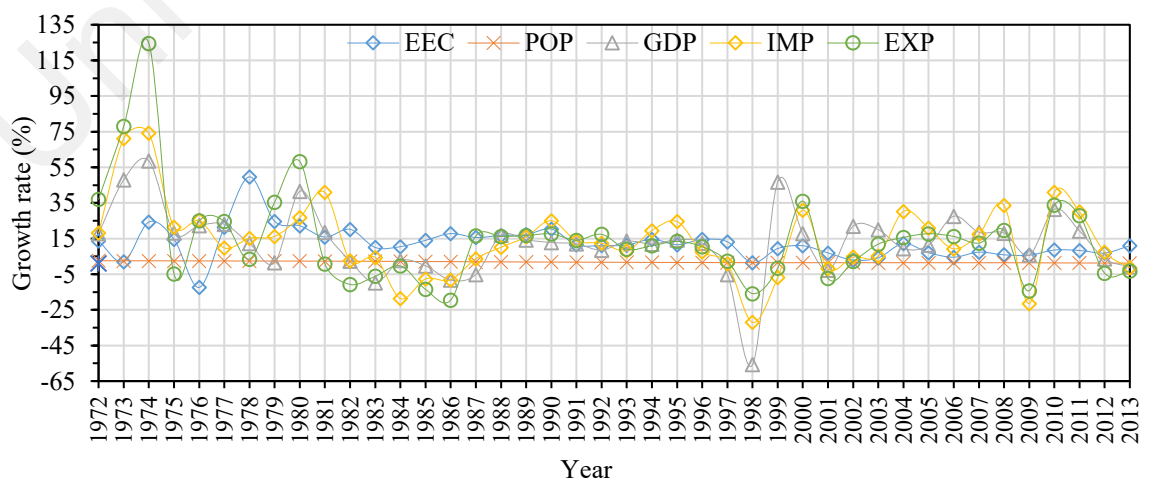


**Figure 4.8: Growth rate of Indonesia's EEC and SEI**
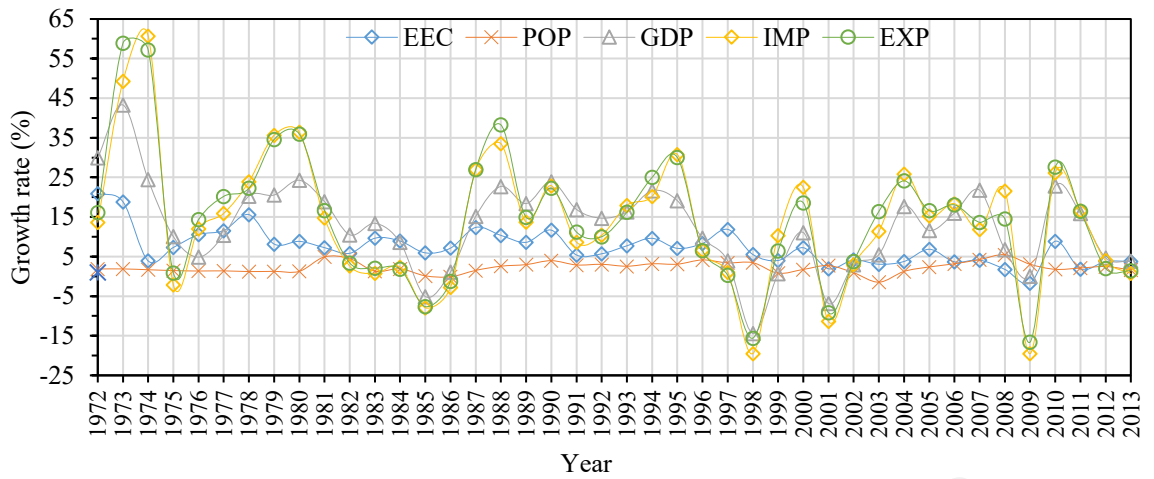
**Figure 4.9: Growth rate of Singapore's EEC and SEI**
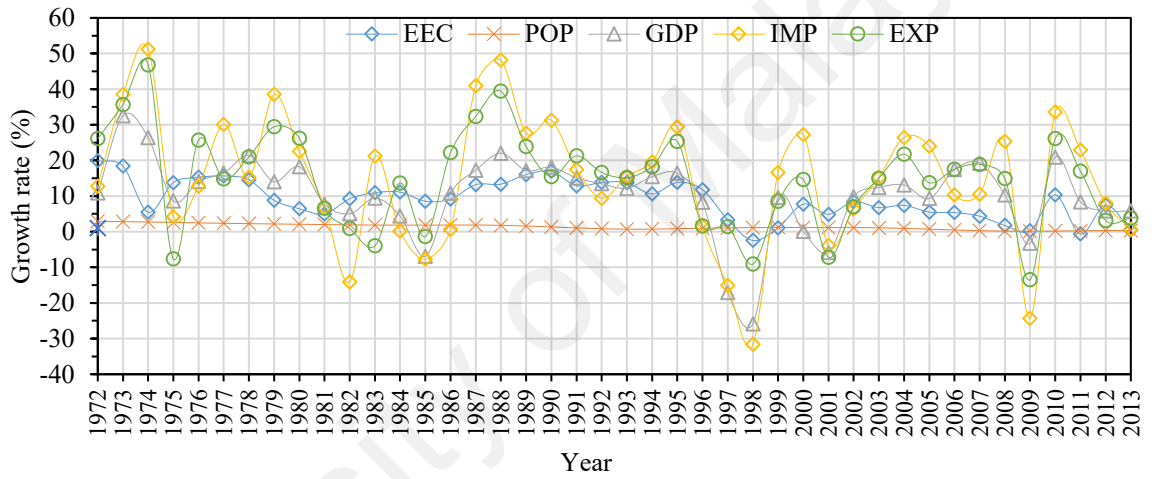


**Figure 4.10: Growth rate of Thailand's EEC and SEI**
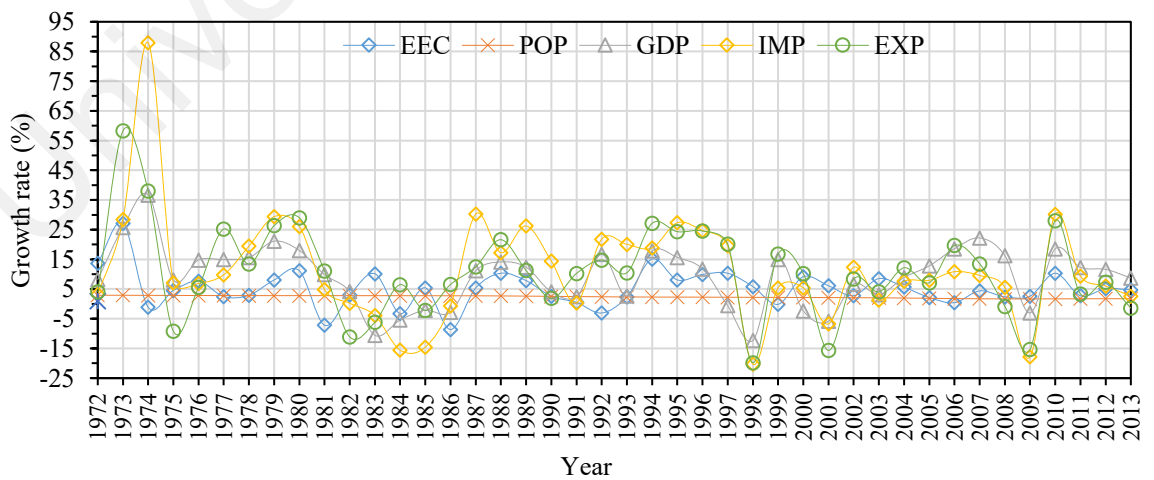


**Figure 4.11: Growth rate of Philippines's EEC and SEI**

The average annual growth rate (AAGR) of EEC and SEI in ASEAN-5 countries from 1971 until 2013 are tabulated in Table 4.7. According to this table, the highest growth rate of EEC, GDP and EXP belong to Indonesia, while Philippines and Thailand have the highest growth rate of POP and IMP respectively.

**Table 4.7: Average annual growth rate of EEC and SEI in ASEAN-5 countries**

| SEI | Average annual growth rate (AAGR) | | | | |
| --- | --- | --- | --- | --- | --- |
| | Malaysia | Indonesia | Singapore | Thailand | Philippines |
| EEC | 9.1956% | 12.3583% | 7.4240% | 9.0598% | 5.1950% |
| POP | 2.3551% | 1.8239% | 2.2679% | 1.3654% | 2.3666% |
| GDP | 11.3659% | 13.0150% | 12.8156% | 10.4486% | 9.4165% |
| IMP | 13.7575% | 14.2830% | 13.9332% | 14.9076% | 11.4034% |
| EXP | 13.8491% | 14.8257% | 14.6587% | 14.7486% | 10.7592% |

### 4.2.1 Formulation of EEC by Metaheuristic Methods

In general, the annual EEC base on factors affecting the consumption is modeled mathematically by:

$$\overline{EEC}(t) = \sum_{i=1}^{T} w_i \, \overline{X(t)}_i^{w_i^m + T} + k \left( \sum_{i=1}^{T} \sum_{j=i}^{T} w_{ij} \, \overline{X(t)}_i \, \overline{X(t)}_j \right) + w_0 \tag{4.5}$$

where $\overline{EEC}$ is the normalized electrical energy consumption, $\overline{X}_i$ and $\overline{X}_j$ are the ith and j[th] normalized independent variables respectively, $w_i$ and $w_{ij}$ are the corresponding weighting factors, $w_0$ is the constant value ,and $T$ is the total number of independent variables. In this general model $m$ and $k$, determine the exact form of mathematical model.

As indicated in Tables 2.1 and 4.1, the multiple linear ($m = 0$ and $k=0$), quadratic ($m = 0$ and $k=1$) and exponential ($m = 1$ and $k=0$) forms of equations are commonly used to model relationships between EEC and input historical data. In this study, multiple linear and quadratic models represent linear and monotonic relationships between EEC and input historical data respectively.

While the EEC is assumed as a time series and is defined as a function of annual EEC in preceding three years, the multiple linear, quadratic, and exponential models are expressed as follows:

$$\overline{EEC}(t)_{Linear} = w_1 \overline{EEC}(t-1) + w_2 \overline{EEC}(t-2) + w_3 \overline{EEC}(t-3) + w_0 \tag{4.6}$$

$$
\begin{aligned}
\overline{EEC}(t)_{Quadratic} &= w_1 \overline{EEC}(t-1) + w_2 \overline{EEC}(t-2) + w_3 \overline{EEC}(t-3) \\
&+ w_{12} \overline{EEC}(t-1)\overline{EEC}(t-2) + w_{13} \overline{EEC}(t-1)\overline{EEC}(t-3) \\
&+ w_{23} \overline{EEC}(t-2)\overline{EEC}(t-3) \\
&+ w_{11} \overline{EEC}(t-1)^2 + w_{22} \overline{EEC}(t-2)^2 + w_{33} \overline{EEC}(t-3)^2 + w_0
\end{aligned}
\tag{4.7}
$$

$$
\begin{aligned}
\overline{EEC}(t)_{Exponential} &= w_1 \overline{EEC}(t-1)^{w_4} + w_2 \overline{EEC}(t-2)^{w_5} \\
&+ w_3 \overline{EEC}(t-3)^{w_6} + w_0
\end{aligned}
\tag{4.8}
$$

For considering the structure of socio-economic indicators, GDP, POP, IMP and EXP are used as input historical data sets and the multiple linear, quadratic, and exponential models for the annual EEC are rewritten as follows:

$$\overline{EEC}(t)_{Linear} = w_1 \overline{GDP}(t) + w_2 \overline{POP}(t) + w_3 \overline{IMP}(t) + w_4 \overline{EXP}(t) + w_0 \tag{4.9}$$

$$
\begin{aligned}
\overline{EEC}(t)_{Quadratic} &= w_1 \overline{GDP}(t) + w_2 \overline{POP}(t) + w_3 \overline{IMP}(t) + w_4 \overline{EXP}(t) \\
&+ w_{12} \overline{GDP}(t)\overline{POP}(t) + w_{13} \overline{GDP}(t)\overline{IMP}(t) + w_{14} \overline{GDP}(t)\overline{EXP}(t) \\
&+ w_{23} \overline{POP}(t)\overline{IMP}(t) + w_{24} \overline{POP}(t)\overline{EXP}(t) + w_{34} \overline{IMP}(t)\overline{EXP}(t) \\
&+ w_{11} \overline{GDP}(t)^2 + w_{22} \overline{POP}(t)^2 + w_{33} \overline{IMP}(t)^2 + w_{44} \overline{EXP}(t)^2 + w_0
\end{aligned}
\tag{4.10}
$$

$$
\begin{aligned}
\overline{EEC}(t)_{Exponential} &= w_1 \overline{GDP}(t)^{w_5} + w_2 \overline{POP}(t)^{w_6} + w_3 \overline{IMP}(t)^{w_7} \\
&+ w_4 \overline{EXP}(t)^{w_8} + w_0
\end{aligned}
\tag{4.11}
$$

121

The direct and indirect effects of socio-economic indicators on electricity consumption make the EEC forecasting a complex optimization problem, which needs an efficient method to provide the accurate coefficients of the EEC models. To forecast the electricity consumption accurately the metaheuristic methods are applied for seeking the optimal coefficients of EEC models through minimizing the cost function given by:

$$F = \sum_{t=1}^{N} \left| (EEC\ (t)_{observed} - EEC(t)) \right| \qquad (4.12)$$

where $EEC\ (t)_{observed}$ and $EEC(t)$ are the observed and predicted electricity consumption respectively and $N$ is total number of observations.

The coefficients of $EEC$ models $w$ are calculated by:

$$w = \textbf{arg\ min}\ F \qquad (4.13)$$

In order to provide more accurate optimal values of weighting parameters based on historical data types for EEC forecasting of ASEAN-5 countries, those optimization methods have been invoked in optimized GEP are applied as powerful tools to search for the best estimate of the vector parameters within the search space.

### 4.2.2 Formulation of EEC by Optimized GEP

Compared to metaheuristic methods, GEP strategy provides a superior alternative for long-term energy consumption forecasting as it precludes the need for predefined form of the mathematical models. Although metaheuristic methods have been applied to provide realistic estimation models, the major limitation is that they require prior knowledge about the nature of the relationships between independent and dependent variables.

In order to cope with the limitations of the existing methods for formulating EEC, optimized GEP is employed in this study. First, GEP is applied to extract the structure of the existing relationship between the historical data and transform the derived information into a mathematical model then a metaheuristic optimization method is deployed over GEP model to determine the optimal weighting factors through minimizing the same fitness evaluation function as in GEP (Eq. (4.12)).

To formulate the long-term EEC by GEP, the electricity consumption is considered as a time series that is a function of annual consumption in preceding three years as expressed by

$$\overline{EEC}(t)_{GEP} = f\left(\overline{EEC}(t-1),\ \overline{EEC}(t-2),\ \overline{EEC}(t-3)\right) \qquad (4.14)$$

To formulate the long-term EEC by GEP while the socio-economic indicators are considered as input data sets, it is defined that

$$\overline{EEC}(t)_{GEP} = f\left(\overline{GDP}(t),\ \overline{POP}(t),\ \overline{IMP}(t),\ \overline{EXP}(t)\right) \qquad (4.15)$$

## 4.3 Simulation Results and Discussions

The optimized GEP is used in this study to precisely formulate the relationships between historical data sets and EEC of ASEAN-5 countries, namely, Malaysia, Indonesia, Singapore, Thailand, and Philippines, as they inherently have different dependencies on EEC. The most effective historical data type is selected by a parallel comparison. For the purpose of parallel comparison, two different input data types (i.e. SEI and EEC) of ASEAN-5 countries from 1971 until 2011 are examined. Furthermore, to assess the effectiveness of optimized GEP for long-term EEC forecasting its estimates are compared with those obtained from ANN, SVR, ANFIS, rule-based data mining algorithm, GEP, linear and quadratic models optimized by PSO, CSA, ACS and BSA.

In general, to find the empirical models by the soft computing techniques, a similar procedure is followed. Thus, in this study, to obtain the optimal AI-based models for long-term EEC forecasting, the following sequential steps are carried out for all methods.

i- For parallel comparison, two different types of historical data for 1971-2011 are considered as independent variables and annual EEC on that period is considered as a dependent variable.

ii- Both independent and dependent variables are divided into two subsets, where annual data from 1971 to 2001 inclusive is utilized for training of design phase and that for 2002-2011 is used for testing phase of obtained models.

iii- The training phase is designed for the learning process. The computer programs that connect the input variables to the output are derived through learning process. To speed up the learning process, both input and output variables are normalized according to Eq. (4.1).

iv- The testing phase is used to measure the performance of the models obtained by AI-based methods on data that played no role in building the models. To quantify the prediction performance of models several evaluation criteria are employed; mean absolute percentage error (MAPE), root mean square error (RMSE), Thiel's inequality coefficient (U-statistic) and coefficient of determination (R-squared), given by:

$$MAPE\% = \frac{1}{N} \sum_{t=1}^{N} \frac{\left| (EEC(t)_{observed} - EEC(t)) \right|}{EEC(t)_{observed}} \times 100 \qquad (4.16)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (EEC(t)_{observed} - EEC(t))^2} \qquad (4.17)$$

$$U = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{t=1}^{N} (EEC(t)_{observed})^2} + \sqrt{\frac{1}{N} \sum_{t=1}^{N} (EEC(t))^2}} \qquad (4.18)$$

v-    U-statistic provides a measure of how well fitted a time series of forecasted values to a corresponding time series of actual data. The U-statistic is always laid between [0,1] while the value closer to zero indicating greater forecasting accuracy with a perfect fit and the value closer to one means the estimation is no better than a naive guess.

vi-   Whiteness test (Durbin-Watson test) is used to ensure that the obtained models adequately describe a given data series (AlRashidi & El-Naggar, 2010). The whiteness test is obtained through a confirmatory analysis. The objective of confirmatory analysis is to confirm the whiteness of estimated residuals (*e(t)*). The whiteness of estimated residuals implies that they are uncorrelated. This analysis is provided via the calculation of the residuals autocorrelation function (RACF) as defined by:

$$ RACF = \frac{\left| \sum_{t=2}^{N} (e(t) \ e(t-1)) \right|}{\sum_{i=1}^{N} (e(t))^2} \qquad (4.19) $$

vii-  The RACF values are in the range of [0, 1], if the RACF value is significantly different from zero, it will fall outside a confidence level. This indicates that the residuals are not white (correlated) and an important independent variable has been omitted from the tested model.

viii- The most efficient model and the most effective input historical data type are selected according to their performance on the testing phase to choose best ones with generalization capability when dealing with unseen data in the future applications.

As, the parameters setting of AI-based methods are highly problem dependent and there is no consensus about the optimum values of parameters, the methodologies similar

to that successfully used in the literature for energy consumption forecasting are considered in this study to set the control parameters of applied methods.

The back propagation MLP as a type of ANN using feed-forward architecture trained with back propagation (BP) learning algorithm is used. This network is considered as a universal estimator because of its small solution network and quick computational speed that permit supervised training over large input data sets. The effectiveness of ANN models depend on their network's structure, transfer function, and learning algorithm, thus the following characteristics: two hidden layers, logarithmic sigmoid transfer function, and Levenberg-Marquardt PB learning algorithm as determined in (Ekonomou, 2010) for long-term EEC forecasting, are used to form a MLP model in this study.

The model constructing in SVR depends on the selection of kernel, kernel's parameters, soft margin parameter, and fraction of error goal. In this study, Gaussian RBF (radial basis function) kernel has been employed for as much as it is not only easier to implement, but also has excellent overall performance compared to other kernel functions in kernelized learning algorithms. As there is no structural method on efficient tuning of SVR parameters, this parameters have been set to ($\upsilon = 0.5$, $C = 1$, $\partial = 1/6$) for the sake of comparison as recommended in (J. Wang et al., 2012), where, $\upsilon$ within its range $[0, 1]$ controls the width of the fraction of errors, $C$ as a regularization parameter controls the empirical risk degree of SVR, and the $\partial$ controls the bandwidth of the Gaussian RBF kernel.

To form the ANFIS structure the Sugeno-type FIS is used as a promising alternative to Mamdani-type, since the Sugeno-type is well suited for modeling nonlinear systems by interpolating between multiple linear models. The membership function type is Gaussian as recommended in (A. Azadeh et al., 2009; Nadimi et al., 2010) and two

clustering methods are used to set up the ANFIS, namely Fuzzy c-means (FCM) clustering and subtractive clustering (radii = 0.8).

Data mining is the process of extracting potentially useful knowledge, from the enormous data sets to provide explicit model of what is happening behind some data so that it can predict future outcomes. Among different data mining (e.g. rule-based and/or decision tree learning) algorithms, M5-Rules (M5-R) as a rule-based learning algorithm is applied for long-term energy consumption forecasting due to its ability to reflect the growth trends beyond the training period. M5-R builds a model tree using symbolic rule learning algorithms and makes the best leaf into a rule in each iteration. It creates a decision list for regression problems using separate-and-conquer (S&C) or covering algorithm. This algorithm ensures that each instance of the original training set is covered by at least one rule. The model structure is built in WEKA environment (i.e. an open-source data mining suite) with default parameter values as recommended in (Wu et al., 2013).

GEP is undertaken using a GeneXproTools (i.e. a data-driven computing package). The GEP implementation procedure involves following steps: arranging the structure of chromosomes, forming the corresponding ET for all genes within the chromosomes to evaluate the GEP performance and developing the evolutionary processes to adapt the given data into an algebraic expression. Once the head length and the set of functions have been set, the length of a gene is also decided accordingly to arrange the structure of chromosomes. Genes within one chromosome are transformed into sub-ETs that are connected by a predefined linking function so to form an ET its parameters should be further specified. Eventually, the evolutionary operators are tuned to develop an efficient learning algorithm. Table 4.8 summarizes all parameter settings of GEP models. To

develop an optimized GEP the generated models are integrated with an optimization method in MATLAB environment.

The standard form of PSO algorithm is executed to form optimized GEP. The population is set to 100 in this algorithm. The acceleration factors $c_1$ and $c_2$ are both 2.0 similar to (Kıran et al., 2012a; Rafieerad et al., 2017), a decaying inertia weight $\omega$ starting at 0.9 and ending at 0.4 with run time increasing is used as specified in (Askarzadeh, 2014) for long-term EEC forecasting.

For CSA the number of nests ($Np$) is set to 100. The two main parameters in this method that have to be predetermined are the probability of an alien egg ($p_a$) and the value of the distribution factor ($\beta$). As reported in (Nguyen, Vo, & Truong, 2014), while $p_a$ has tuned in its range [0, 1] the effect of the probability ($p_a$) is inconsiderable and different values of the probability in this range can lead to the same optimal solution. Besides, as experience in (Nguyen et al., 2014), the value of the distribution factor in the suggested range ($\beta \in [0.3, 1.99]$) does not have much effect on the final solution of optimization problems so, it is fixed at 1.5 in this study.

The control parameter of ACS determines the probability of corporation between individuals by selecting the biological interaction level within the crossover process. This parameter varies in the range of 0.05 to 0.9. However, the ACS is not sensitive to this parameter; p = 0.15 is considered in this study as recommended in (Civicioglu, 2013a) and the population of individuals in the superorganisms is set to 100 in this algorithm.

The control parameter of BSA controls the number of individuals to be engaged within the crossover process. This parameter varies from 0% to 100% of population size (Modiri-Delshad, Aghay Kaboli, Taslimi-Renani, & Rahim, 2016). However, the BSA is not sensitive to this parameter; its maximum value is considered in this study to achieve

the best solutions (Modiri-Delshad & Rahim, 2014), and the population of individuals is

set to 100 in this algorithm.

**Table 4.8: Parameter setting of studied methods**

| | Methods | Parameters | Value |
|---|---|---|---|
| **ANN** | MLP | Hidden layer | 2 |
| | | Transfer function | logarithmic sigmoid |
| | | Learning algorithm | Levenberg-Marquardt PB |
| **SVR** | RBF kernel | Kernel's parameter ($\partial$) | 1/6 |
| | | Soft margin parameter (C) | 1 |
| | | Fraction of error ( $\upsilon$) | 0.5 |
| **ANFIS** | FCMC (FCM clustering) | FIS structure | Sugeno-type |
| | | Membership function | Gaussian |
| | SC (Subtractive clustering) | Cluster radius | 0.8 |
| | | FIS structure | Sugeno-type |
| | | Membership function | Gaussian |
| **Rule-based data mining** | M5-R | Min number of instances per leaf node | 4 |
| **GEP** | Chromosome architecture | Chromosomes | 30 |
| | | Number of genes | 4 |
| | | Head size | 7 |
| | | Function set | $\times,/,+,-$, power (x, y*) , $e^{\wedge y}$, Log, EXP |
| | | Constants per gene | 2 Floating points $\in$ [-10, +10] |
| | ET-setting | Linking function | Addition |
| | | Fitness function | Eq.15 |
| | Genetic operators | Mutation rate | 0.00138 |
| | | IS & RIS transposition rate | 0.00546 |
| | | Single & double crossover rate | 0.00277 |
| | | Gene crossover rate | 0.00277 |
| | | Inversion rate | 0.00546 |
| **Metaheuristic optimization** | PSO | Swarm population | 100 |
| | | $w$ | [0.4, 0.9] |
| | | $c_1 = c_2$ | 2 |
| | CSA | Number of nests | 100 |
| | | Distribution factor ($\beta$) | 1.5 |
| | | Probability of an alien egg ($Pa$) | [0, 1] |
| | ACS | Number of individuals | 100 |
| | | $P$ | 0.15 |
| | BSA | Number of individuals | 100 |
| | | mixrate | 100% |

The performances of machine learning methods for Malaysia's EEC forecasting based on two input historical data types (SEI and EEC) are tabulated in Tables 4.9 and 4.10. The RACF values in these tables confirm the whiteness of estimated residuals at a confidence interval level for all obtained models.

According to these tables, the forecasting accuracy of methods with two different input data types based on multiple-criteria decision analysis using mean rank of methods for MAPE index is ranked as $GEP_{-BSA} > GEP_{-ACS} > GEP_{-CSA} = GEP_{-PSO} > GEP = exponential_{-BSA} > exponential_{-ACS} > M5-R = exponential_{-PSO} > quadratic_{-PSO} > quadratic_{-BSA} > quadratic_{-ACS} > ANFIS_{-SC} = ANFIS_{-FCMC} = linear_{-ACS} > ANN = linear_{-PSO} = exponential_{-CSA} > linear_{-BSA} > SVR > linear_{-CSA} > quadratic_{-CSA}$.

Regardless of the input data type, the enhanced results in Tables 4.9 and 4.10 show that the optimized GEP methods provide better-fit estimation than other studied approaches. Moreover, it is found that the most efficient optimization algorithm for training GEP is BSA, as GEP-$_{BSA}$ yields more promising results in term of MAPE, *U*-statistic, RMSE and *R*-squared indexes. Eqs. (4.20) and (4.21) present Malaysia's EEC formulations derived from GEP-$_{BSA}$ based on SEI and EEC respectively. In Figure 4.12 GEP$_{-BSA}$ performances during training of design phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types are depicted. It is found that, while the EEC is considered as the input for long-term EEC forecasting in Malaysia by GEP-$_{BSA}$ model, the lowest MAPE (1.5882%), RMSE (0.0458), *U*-statistic (0.0133) and highest *R*-squared value (0.9933) are attained.

**Table 4.9: Comparison between forecasting accuracy of studied methods on Malaysia's EEC based on SEI**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | | | Input historical data | | |
| | | | | SEI | | |
| ANN | MLP | 4.4974 | 0.1324 | 0.0390 | 0.9490 | 0.0002 |
| support vector machines | SVR | 5.7964 | 0.1238 | 0.0345 | 0.9531 | 0.0023 |
| ANFIS | FCMC | 3.4638 | 0.0716 | 0.0205 | 0.9834 | 0.0009 |
| | SC | 4.1464 | 0.0886 | 0.0251 | 0.9771 | 0.0010 |
| Rule-based data mining | M5-R | 3.3482 | 0.1078 | 0.0295 | 0.9691 | 0.0003 |
| Multiple linear | PSO | 5.6918 | 0.1194 | 0.0339 | 0.9574 | 0.0030 |
| | CSA | 8.8197 | 0.1625 | 0.0460 | 0.9071 | 0.0019 |
| | ACS | 5.6904 | 0.1194 | 0.0339 | 0.9574 | 0.0030 |
| | BSA | 5.6904 | 0.1194 | 0.0339 | 0.9574 | 0.0030 |
| Quadratic | PSO | 5.3469 | 0.1329 | 0.0330 | 0.9475 | 0.0008 |
| | CSA | 10.085 | 0.2138 | 0.0587 | 0.8665 | 0.0035 |
| | ACS | 5.4414 | 0.1129 | 0.0326 | 0.9625 | 0.0021 |
| | BSA | 5.0932 | 0.1036 | 0.0292 | 0.9684 | 0.0015 |
| Exponential | PSO | 5.2251 | 0.1413 | 0.0418 | 0.9403 | 0.0011 |
| | CSA | 5.2046 | 0.1412 | 0.0419 | 0.9338 | 0.0008 |
| | ACS | 4.8430 | 0.1004 | 0.0292 | 0.9690 | 0.0007 |
| | BSA | 3.8568 | 0.0779 | 0.0224 | 0.9812 | 0.0004 |
| GEP | RNC | 3.9818 | 0.0939 | 0.0266 | 0.9735 | 0.0009 |
| Optimized GEP | PSO | 3.3306 | 0.0737 | 0.0214 | 0.9811 | 0.0014 |
| | CSA | 3.0735 | 0.0833 | 0.0243 | 0.9762 | 0.0010 |
| | ACS | 3.0700 | 0.0673 | 0.0193 | 0.9854 | 0.0012 |
| | BSA | 2.8935 | 0.0620 | 0.0190 | 0.9871 | 0.0015 |

$$\overline{EEC(t)}_{GEP-BSA} = \tag{4.20}$$

$$(-13.76)\overline{EXP(t)}\left( Log\left( exp\left( (-0.429)\overline{POP(t)}\left(\overline{IMP(t)}\right)^{0.2} \right)\right)\right)$$

$$\times \left((0.54)\left(\overline{POP(t)} - \overline{IMP(t)}\right)\right)^{\left(7.55 - \overline{GDP(t)}\right)} + (0.84)\overline{EXP(t)} + 1.66$$

**Table 4.10: Comparison between forecasting accuracy of studied methods on Malaysia's EEC based on EEC in preceding years**

| Model | Methods | MAPE (%) | RMSE | *U*-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | EEC | | | | |
| ANN | MLP | 7.1476 | 0.1800 | 0.0541 | 0.9068 | 0.0001 |
| support vector machines | SVR | 3.1507 | 0.0818 | 0.0239 | 0.9790 | 0.0004 |
| ANFIS | FCMC | 8.3499 | 0.2172 | 0.0658 | 0.8643 | 0.0002 |
| | SC | 5.4397 | 0.1458 | 0.0434 | 0.9389 | 0.0003 |
| Rule-based data mining | M5-R | 3.9939 | 0.1324 | 0.0317 | 0.9569 | 0.0003 |
| Multiple linear | PSO | 2.4169 | 0.0574 | 0.0166 | 0.9897 | 0.0030 |
| | CSA | 3.4282 | 0.0861 | 0.0251 | 0.9670 | 0.0009 |
| | ACS | 2.3492 | 0.0561 | 0.0162 | 0.9902 | 0.0025 |
| | BSA | 2.4405 | 0.0579 | 0.0167 | 0.9895 | 0.0030 |
| Quadratic | PSO | 1.8840 | 0.0560 | 0.0162 | 0.9903 | 0.0023 |
| | CSA | 5.7457 | 0.1270 | 0.0356 | 0.9490 | 0.0032 |
| | ACS | 2.0994 | 0.0592 | 0.0171 | 0.9893 | 0.0028 |
| | BSA | 2.4365 | 0.0693 | 0.0202 | 0.9867 | 0.0009 |
| Exponential | PSO | 1.7351 | 0.0490 | 0.0142 | 0.9924 | 0.0015 |
| | CSA | 5.1652 | 0.1024 | 0.0288 | 0.9655 | 0.0027 |
| | ACS | 1.6505 | 0.0488 | 0.0141 | 0.9925 | 0.0022 |
| | BSA | 1.6773 | 0.0502 | 0.0145 | 0.9921 | 0.0027 |
| GEP | RNC | 1.6687 | 0.0492 | 0.0143 | 0.9924 | 0.0016 |
| Optimized GEP | PSO | 1.5956 | 0.0462 | 0.0134 | 0.9932 | 0.0015 |
| | CSA | 1.5962 | 0.0461 | 0.0133 | 0.9932 | 0.0015 |
| | ACS | 1.5893 | 0.0461 | 0.0133 | 0.9931 | 0.0014 |
| | BSA | 1.5882 | 0.0458 | 0.0133 | 0.9931 | 0.0014 |

$$\overline{EEC(t)}_{GEP-BSA} =$$

$$(0.87)\left(\left(\left(\overline{EEC(t-1)}\right)^{0.25} - \left(\overline{EEC(t-3)}\right)^{0.2}\right)^{(4)\left(\overline{EEC(t-1)}\,\overline{EEC(t-3)}\right)^{0.33}}\right)^{0.2} \tag{4.21}$$

$$+ (0.94)\left(\overline{EEC(t-3)}\right)^{0.032}\left(\overline{EEC(t-1)} - \overline{EEC(t-3)}\right) + (1.029)\overline{EEC(t-2)} - 0.03$$

**Figure 4.12: Malaysia's EEC actual data from 1971 to 2011 and GEP-BSA performances during training of design phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types (SEI and EEC)**

The performances of studied methods for EEC forecasting in Indonesia based on two input data types (SEI and EEC) are quantified in Tables 4.11 and 4.12. The calculated RACF values in these tables indicate that the estimated residuals of all models are uncorrelated and obtained models adequately describe the given set of data. The forecasting accuracy of methods on Indonesia's EEC with two input data types according to average rank of methods for MAPE index is ranked as GEP-BSA > GEP-ACS > GEP-PSO > GEP-CSA > exponential-PSO > ANFIS-SC  = quadratic-ACS > quadratic-PSO > GEP > exponential-BSA > exponential-ACS > linear-ACS = quadratic-BSA > linear-PSO > ANN > SVR > linear-CSA = linear-BSA > quadratic-CSA = exponential-CSA > ANFIS-FCMC > M5-R. In addition, the parallel comparison reveals that GEP-BSA is the most efficient model and EEC is more effective input historical data type than SEI, while the best reported values in Tables 4.11 and 4.12 in term of MAPE (0.8538%), RMSE (0.017), $U$-statistic (0.0048) and $R$-squared (0.9984) indexes belong to this model on the basis of EEC in preceding three years. The effects of socio-economic indicators and EEC in preceding years on Indonesia's EEC are formulated via GEP-BSA as given by Eqs. (4.22) and (4.23) respectively. The performances of GEP–BSA based on two input data types are illustrated in Figure 4.13.

133

**Table 4.11: Comparison between forecasting accuracy of studied methods on Indonesia's EEC based on SEI**

| Model | Methods | MAPE (%) | RMSE | *U*-statistic | $R^2$ | RACF |
|-------|---------|----------|------|---------------|-------|------|
| | | Input historical data | | | | |
| | | SEI | | | | |
| ANN | MLP | 2.5090 | 0.0460 | 0.0129 | 0.9940 | 0.0006 |
| support vector machines | SVR | 4.6485 | 0.1079 | 0.0304 | 0.9605 | 0.0039 |
| ANFIS | FCMC | 12.532 | 0.2680 | 0.0715 | 0.8042 | 0.0045 |
| | SC | 2.0935 | 0.0439 | 0.0125 | 0.9928 | 0.0022 |
| Rule-based data mining | M5-R | 8.7602 | 0.1728 | 0.0516 | 0.9137 | 0.0002 |
| Multiple linear | PSO | 7.1256 | 0.1558 | 0.0458 | 0.9222 | 0.0123 |
| | CSA | 4.9234 | 0.1090 | 0.0310 | 0.9525 | 0.0067 |
| | ACS | 6.2888 | 0.1216 | 0.0353 | 0.9466 | 0.0115 |
| | BSA | 8.2636 | 0.1806 | 0.0534 | 0.8992 | 0.0152 |
| Quadratic | PSO | 2.4535 | 0.0567 | 0.0160 | 0.9907 | 0.0023 |
| | CSA | 5.8577 | 0.1428 | 0.0405 | 0.9334 | 0.0131 |
| | ACS | 3.9840 | 0.1259 | 0.0351 | 0.9560 | 0.0058 |
| | BSA | 7.7384 | 0.2312 | 0.0629 | 0.8539 | 0.0104 |
| Exponential | PSO | 3.7295 | 0.0708 | 0.0204 | 0.9800 | 0.0038 |
| | CSA | 6.7837 | 0.1312 | 0.0384 | 0.8945 | 0.0030 |
| | ACS | 4.8770 | 0.0927 | 0.0269 | 0.9586 | 0.0033 |
| | BSA | 4.1775 | 0.0919 | 0.0260 | 0.9711 | 0.0038 |
| GEP | RNC | 3.1973 | 0.0635 | 0.0183 | 0.9857 | 0.0015 |
| Optimized GEP | PSO | 2.4381 | 0.0484 | 0.0138 | 0.9917 | 0.0017 |
| | CSA | 2.4720 | 0.0476 | 0.0136 | 0.9914 | 0.0018 |
| | ACS | 2.3354 | 0.0463 | 0.0132 | 0.9915 | 0.0017 |
| | BSA | 2.0545 | 0.0384 | 0.0108 | 0.9957 | 0.0008 |

$$\overline{EEC(t)}_{GEP-BSA} = $$

$$\frac{(1.285)\overline{IMP(t)}}{\left(\overline{POP(t)}\ exp\Big((0.199)\overline{POP(t)}\ \big(\overline{GDP(t)} - \overline{POP(t)} - \overline{EXP(t)}\big)\Big)\right)^{0.25}}$$

$$+ (0.846)\big(\overline{EXP(t)} - \overline{POP(t)}\big)\left(Log\left(\frac{\overline{IMP(t)}}{\big(\overline{POP(t)}\big)^2}\right)\right)^5 + 0.075 \tag{4.22}$$

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | EEC | | | | |
| ANN | MLP | 3.7065 | 0.0866 | 0.0251 | 0.9792 | 0.0003 |
| support vector machines | SVR | 2.2494 | 0.0567 | 0.0163 | 0.9902 | 0.0003 |
| ANFIS | FCMC | 2.6444 | 0.0586 | 0.0169 | 0.9903 | 0.0005 |
| | SC | 1.9537 | 0.0420 | 0.0121 | 0.9949 | 0.0005 |
| Rule-based data mining | M5-R | 12.223 | 0.2265 | 0.0689 | 0.8393 | 0.0003 |
| Multiple linear | PSO | 1.5747 | 0.0292 | 0.0082 | 0.9969 | 0.0008 |
| | CSA | 1.9776 | 0.1279 | 0.0377 | 0.9457 | 0.0008 |
| | ACS | 1.5747 | 0.0292 | 0.0082 | 0.9969 | 0.0008 |
| | BSA | 1.5747 | 0.0292 | 0.0082 | 0.9969 | 0.0008 |
| Quadratic | PSO | 1.8720 | 0.0419 | 0.0120 | 0.9948 | 0.0005 |
| | CSA | 5.1518 | 0.1027 | 0.0285 | 0.9698 | 0.0010 |
| | ACS | 1.3224 | 0.0314 | 0.0090 | 0.9969 | 0.0006 |
| | BSA | 1.1515 | 0.0263 | 0.0075 | 0.9977 | 0.0005 |
| Exponential | PSO | 1.1355 | 0.0266 | 0.0076 | 0.9976 | 0.0005 |
| | CSA | 2.9690 | 0.0552 | 0.0155 | 0.9906 | 0.0007 |
| | ACS | 1.8054 | 0.0398 | 0.0114 | 0.9953 | 0.0006 |
| | BSA | 1.8612 | 0.0415 | 0.0119 | 0.9949 | 0.0005 |
| GEP | RNC | 1.9332 | 0.0397 | 0.0114 | 0.9932 | 0.0007 |
| Optimized GEP | PSO | 0.8754 | 0.0170 | 0.0048 | 0.9984 | 0.0007 |
| | CSA | 0.8553 | 0.0172 | 0.0049 | 0.9983 | 0.0007 |
| | ACS | 0.8545 | 0.0170 | 0.0048 | 0.9984 | 0.0007 |
| | BSA | 0.8538 | 0.0170 | 0.0048 | 0.9984 | 0.0007 |

$$\overline{EEC}(t)_{GEP-BSA} =$$
$$-2\,exp\left(\left(\overline{EEC}(t-3)+\overline{EEC}(t-1)\right)^{0.1}\left(\left(\overline{EEC}(t-3)\right)^{0.25}+\overline{EEC}(t-2)-6.672\right)\right) \quad (4.23)$$
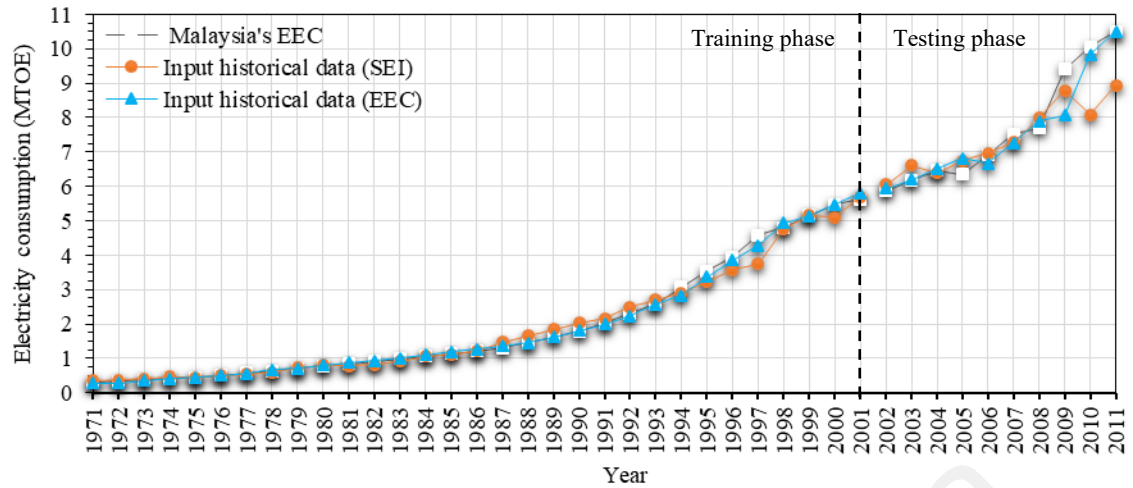$$+(1.057)\overline{EEC}(t-1)-0.035$$
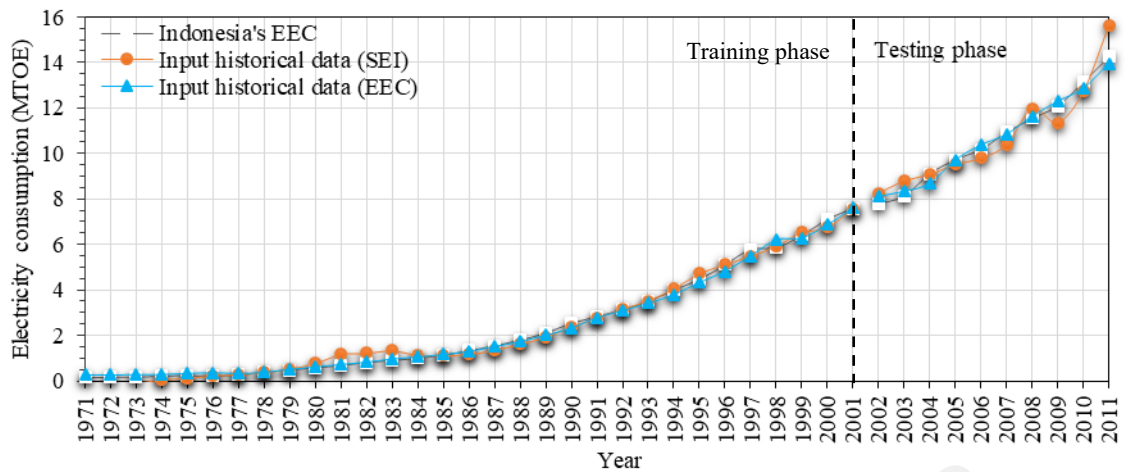
**Figure 4.13: Indonesia's EEC actual data from 1971 to 2011 and GEP-BSA performances during training of design phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types (SEI and EEC)**

For further examination of solution methodologies, the performances of optimized GEP methods for long-term EEC forecasting in Singapore are compared with those from other soft computing approaches as shown in Tables 4.13 and 4.14. The RACF values reported in these tables confirm that the estimated residuals of all obtained models are white at a confidence interval level.

According to Tables 4.13 and 4.14, the forecasting accuracy of studied methods with two different input data types based on mean rank of methods for MAPE index is ranked as GEP$_{-BSA}$ > GEP$_{-ACS}$ > GEP$_{-PSO}$ > GEP$_{-CSA}$ > exponential$_{-PSO}$ > ANFIS$_{-SC}$ = quadratic$_{-ACS}$ > quadratic$_{-PSO}$ > GEP > exponential$_{-BSA}$ > exponential$_{-ACS}$ > linear$_{-ACS}$ = quadratic$_{-BSA}$ > linear$_{-PSO}$ > ANN > SVR > linear$_{-CSA}$ = linear$_{-BSA}$ > quadratic$_{-CSA}$ = exponential$_{-CSA}$ > ANFIS$_{-FCMC}$ > M5-R.

Regardless of which input data type is considered for long-term EEC forecasting in Singapore, the enhanced results in Tables 4.13 and 4.14 show that the optimized GEP methodologies provide better-fit estimation than other studied methods. Moreover, it is determined that BSA is the most efficient optimization algorithm for training the GEP,

while in term of MAPE index, 1.7537% and 0.7438%, achieved by GEP-$_{BSA}$ on the basis of SEI and EEC respectively, that are less than the values achieved by the other optimization algorithms. Furthermore, the parallel comparison indicates that EEC in preceding three years as input for GEP-$_{BSA}$, leads to more promising results in term of MAPE (0.7438%), RMSE (0.0221), $U$-statistic (0.0059) and $R$-squared (0.9983) indexes as compared with those resulting from socio-economic indicators data. Eqs. (4.24) and (4.25) give the mathematical models of annual EEC in Singapore generated by GEP-$_{BSA}$ based on socio-economic indicators and EEC in preceding years respectively, and Figure 4.14 shows their performances.

**Table 4.13: Comparison between forecasting accuracy of studied methods on Singapore's EEC based on SEI**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | SEI | | | | |
| ANN | MLP | 5.4114 | 0.1150 | 0.0317 | 0.9675 | 0.0001 |
| support vector machines | SVR | 2.2395 | 0.0466 | 0.0125 | 0.9928 | 0.0004 |
| ANFIS | FCMC | 11.051 | 0.2301 | 0.0653 | 0.8748 | 0.0022 |
| | SC | 11.723 | 0.2433 | 0.0693 | 0.8600 | 0.0027 |
| Rule-based data mining | M5-R | 3.2925 | 0.0652 | 0.0175 | 0.9883 | 0.0006 |
| Multiple linear | PSO | 7.2170 | 0.1714 | 0.0442 | 0.9286 | 0.0017 |
| | CSA | 4.4273 | 0.1082 | 0.0284 | 0.9701 | 0.0013 |
| | ACS | 4.0048 | 0.0844 | 0.0229 | 0.9803 | 0.0011 |
| | BSA | 8.2341 | 0.1921 | 0.0493 | 0.9107 | 0.0020 |
| Quadratic | PSO | 5.0722 | 0.1400 | 0.0385 | 0.9522 | 0.0016 |
| | CSA | 3.2630 | 0.0771 | 0.0204 | 0.9842 | 0.0006 |
| | ACS | 3.0950 | 0.0836 | 0.0221 | 0.9819 | 0.0021 |
| | BSA | 3.1069 | 0.0807 | 0.0218 | 0.9829 | 0.0023 |
| Exponential | PSO | 4.1140 | 0.0864 | 0.0230 | 0.9809 | 0.0008 |
| | CSA | 6.2667 | 0.1441 | 0.0378 | 0.9464 | 0.0018 |
| | ACS | 2.9980 | 0.0682 | 0.0181 | 0.9866 | 0.0006 |
| | BSA | 2.1266 | 0.0431 | 0.0116 | 0.9936 | 0.0004 |
| GEP | RNC | 1.9352 | 0.0444 | 0.0119 | 0.9941 | 0.0005 |
| Optimized GEP | PSO | 1.7855 | 0.0437 | 0.0117 | 0.9942 | 0.0006 |
| | CSA | 1.8391 | 0.0489 | 0.0117 | 0.9929 | 0.0006 |
| | ACS | 1.8656 | 0.0494 | 0.0132 | 0.9928 | 0.0006 |
| | BSA | 1.7537 | 0.0436 | 0.0130 | 0.9943 | 0.0006 |

$$\overline{EEC(t)}_{GEP-BSA} =$$

$$\frac{\left(\overline{POP(t)}+\overline{IMP(t)}\right)\left(1.152-\overline{GDP(t)}\right)\left(\overline{GDP(t)}-\overline{POP(t)}\right)}{\left(\overline{GDP(t)}\right)^4}$$

$$+\frac{(0.477)\overline{IMP(t)}\,\overline{GDP(t)}\left(\overline{IMP(t)}-\overline{GDP(t)}\right)}{\left(\overline{GDP(t)}\right)(3)\overline{GDP(t)}}+\frac{(0.398)\left(\overline{EXP(t)}-\overline{GDP(t)}\right)}{\left(\overline{GDP(t)}\right)^{0.25}} \qquad (4.24)$$

$$+\overline{POP(t)}+0.004$$

**Table 4.14: Comparison between forecasting accuracy of studied methods on Singapore's EEC based on EEC in preceding years**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | EEC | | | | |
| ANN | MLP | 7.9625 | 0.1721 | 0.0481 | 0.9300 | 2.4E-6 |
| support vector machines | SVR | 2.1208 | 0.0542 | 0.0147 | 0.9924 | 9.6E-5 |
| ANFIS | FCMC | 1.7814 | 0.0453 | 0.0120 | 0.9950 | 0.0006 |
| | SC | 1.1823 | 0.0267 | 0.0072 | 0.9982 | 0.0002 |
| Rule-based data mining | M5-R | 6.8772 | 0.1315 | 0.0365 | 0.9426 | 0.0001 |
| Multiple linear | PSO | 2.1839 | 0.0474 | 0.0125 | 0.9938 | 0.0005 |
| | CSA | 1.2206 | 0.0307 | 0.0082 | 0.9960 | 0.0002 |
| | ACS | 2.1839 | 0.0474 | 0.0125 | 0.9938 | 0.0005 |
| | BSA | 2.1839 | 0.0474 | 0.0125 | 0.9938 | 0.0005 |
| Quadratic | PSO | 0.9837 | 0.0308 | 0.0083 | 0.9971 | 0.0004 |
| | CSA | 1.0547 | 0.0293 | 0.0078 | 0.9942 | 0.0002 |
| | ACS | 1.3785 | 0.0299 | 0.0079 | 0.9972 | 0.0005 |
| | BSA | 1.8180 | 0.0385 | 0.0102 | 0.9958 | 0.0002 |
| Exponential | PSO | 1.0078 | 0.0231 | 0.0062 | 0.9980 | 0.0001 |
| | CSA | 3.1979 | 0.0621 | 0.0170 | 0.9865 | 0.0003 |
| | ACS | 1.0155 | 0.0235 | 0.0063 | 0.9978 | 0.0002 |
| | BSA | 1.8061 | 0.0443 | 0.0120 | 0.9948 | 0.0002 |
| GEP | RNC | 0.9695 | 0.0268 | 0.0072 | 0.9977 | 0.0003 |
| Optimized GEP | PSO | 0.7442 | 0.0222 | 0.0059 | 0.9983 | 0.0003 |
| | CSA | 0.7442 | 0.0231 | 0.0062 | 0.9982 | 0.0003 |
| | ACS | 0.7441 | 0.0222 | 0.0059 | 0.9983 | 0.0003 |
| | BSA | 0.7438 | 0.0221 | 0.0059 | 0.9983 | 0.0003 |

$$\overline{EEC(t)}_{GEP-BSA} =$$

$$\frac{0.962}{\left(\left(\overline{EEC(t-1)}\right)^2-4.065\right)^4-(4.065)\left(\overline{EEC(t-1)}\right)^5\left(1-\overline{EEC(t-2)}\right)} \qquad (4.25)$$

$$+(1.032)\left(exp\left((-103.879)\left(\overline{EEC(t-3)}\right)^2-(6)\;\left(\overline{EEC(t-1)}\right)^3\right)\right)$$

$$+(1.016)\overline{EEC(t-1)}-0.017$$
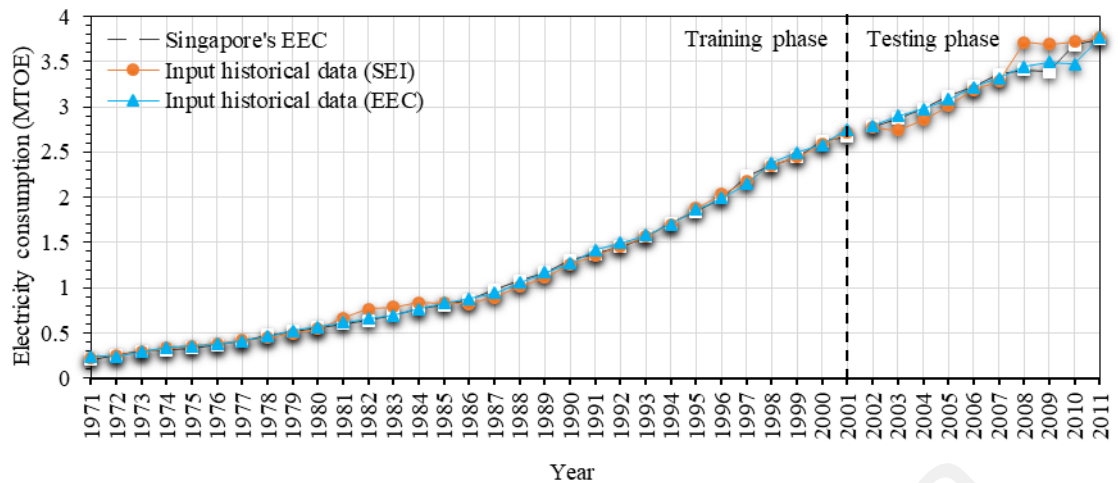
**Figure 4.14: Singapore's EEC actual data from 1971 to 2011 and GEP-BSA performances during training of design phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types (SEI and EEC)**

Tables 4.15 and 4.16 summarizes the performances of machine learning methods for Thailand's EEC forecasting based on two input historical data types (SEI and EEC). It is observed that all obtained models adequately describe the given set of data, while the RACF values reported in these tables confirm the whiteness of estimated residuals at a confidence interval level for all models.

According to Tables 4.15 and 4.16, the forecasting accuracy of methods on Thailand's EEC with two different input data types based on multiple-criteria decision analysis using average rank of methods for MAPE index is ranked as GEP$_{-BSA}$ > GEP$_{-ACS}$ > GEP$_{-PSO}$ > GEP$_{-CSA}$ > exponential$_{-ACS}$ > GEP > quadratic$_{-BSA}$ > linear$_{-CSA}$ = quadratic$_{-ACS}$ = exponential$_{-PSO}$ > exponential$_{-CSA}$ > ANFIS$_{-FCMC}$ = ANFIS$_{-SC}$ = exponential$_{-BSA}$ > linear$_{-ACS}$ > linear$_{-PSO}$ > M5-R = linear$_{-BSA}$ = quadratic$_{-PSO}$ > ANN > SVR > quadratic$_{-CSA}$. The comparison between forecasting accuracy of studied methods on Thailand's EEC reveals that optimized GEP approaches outperform the other studied methods. Whereas the superior MAPE (1.1136%), RMSE (0.0262),$U$-statistic (0.007) and $R$-squared (0.9952) values reported in Table 4.16 belong to GEP$_{-BSA}$ model based on EEC. Therefore, it is concluded that the most efficient optimization algorithm for training GEP and the most

effective historical data set for modeling annual EEC in Thailand are BSA and EEC respectively. Eqs. (4.26) and (4.27) present the mathematical models derived from GEP-<sub>BSA</sub> on the basis of SEI and EEC, respectively. Figure 4.15 depicts the GEP<sub>–BSA</sub> performances during training phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types.

**Table 4.15: Comparison between forecasting accuracy of studied methods on Thailand's EEC based on SEI**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data |
| | | SEI |
| ANN | MLP | 8.1969 | 0.1862 | 0.0482 | 0.9200 | 0.0004 |
| support vector machines | SVR | 8.7580 | 0.2041 | 0.0529 | 0.9013 | 0.0006 |
| ANFIS | FCMC | 7.2486 | 0.1717 | 0.0446 | 0.9322 | 0.0004 |
| | SC | 4.2471 | 0.1041 | 0.0275 | 0.9747 | 0.0004 |
| Rule-based data mining | M5-R | 6.2948 | 0.1393 | 0.0389 | 0.9427 | 0.0007 |
| Multiple linear | PSO | 12.791 | 0.3243 | 0.0823 | 0.7574 | 0.0004 |
| | CSA | 6.7875 | 0.1335 | 0.0372 | 0.9355 | 0.0003 |
| | ACS | 10.549 | 0.2569 | 0.0666 | 0.8322 | 0.0004 |
| | BSA | 12.813 | 0.3250 | 0.0825 | 0.7564 | 0.0004 |
| Quadratic | PSO | 8.2532 | 0.2000 | 0.0520 | 0.9070 | 0.0004 |
| | CSA | 15.347 | 0.3859 | 0.0964 | 0.6571 | 0.0005 |
| | ACS | 4.3885 | 0.1205 | 0.0332 | 0.9659 | 0.0005 |
| | BSA | 3.8569 | 0.1122 | 0.0305 | 0.9704 | 0.0004 |
| Exponential | PSO | 9.2354 | 0.2233 | 0.0580 | 0.8797 | 0.0049 |
| | CSA | 6.7774 | 0.1470 | 0.0398 | 0.9275 | 0.0042 |
| | ACS | 6.5384 | 0.1529 | 0.0402 | 0.9446 | 0.0039 |
| | BSA | 5.3402 | 0.1225 | 0.0324 | 0.9641 | 0.0019 |
| GEP | RNC | 4.3917 | 0.0827 | 0.0228 | 0.9810 | 0.0006 |
| Optimized GEP | PSO | 1.9466 | 0.0413 | 0.0112 | 0.9899 | 0.0008 |
| | CSA | 2.3729 | 0.0477 | 2.3729 | 0.9899 | 0.0007 |
| | ACS | 1.9366 | 0.0451 | 0.0121 | 0.9939 | 00007 |
| | BSA | 1.6996 | 0.0350 | 0.0095 | 0.9914 | 0.0007 |

$$
\overline{EEC(t)}_{GEP-BSA} =
$$

$$
\frac{-\overline{GDP(t)}}{\left(\overline{EXP(t)}\right)^{12} + \dfrac{6.572}{\overline{POP(t)}} - (6.572)\overline{IMP(t)} + 6.53} + (1.241)\left(\sqrt{Log\left(\overline{EXP(t)}\right)} - \overline{POP(t)}\right) \quad (4.26)
$$

$$
+ (1.33)\overline{POP(t)} + 1.22
$$

**Table 4.16: Comparison between forecasting accuracy of studied methods on Thailand's EEC based on EEC in preceding years**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | EEC | | | | |
| ANN | MLP | 4.3944 | 0.0850 | 0.0235 | 0.9833 | 0.0004 |
| support vector machines | SVR | 5.4721 | 0.1193 | 0.0332 | 0.9654 | 0.0004 |
| ANFIS | FCMC | 2.7473 | 0.0625 | 0.0167 | 0.9909 | 0.0004 |
| | SC | 7.4006 | 0.1464 | 0.0411 | 0.9508 | 0.0004 |
| Rule-based data mining | M5-R | 13.545 | 0.2562 | 0.0744 | 0.8303 | 0.0007 |
| Multiple linear | PSO | 1.9657 | 0.0542 | 0.0145 | 0.9922 | 0.0006 |
| | CSA | 1.9338 | 0.0426 | 0.0116 | 0.9928 | 0.0004 |
| | ACS | 1.9656 | 0.0542 | 0.0145 | 0.9922 | 0.0006 |
| | BSA | 1.9658 | 0.0542 | 0.0145 | 0.9922 | 0.0006 |
| Quadratic | PSO | 3.6552 | 0.0780 | 0.0214 | 0.9856 | 0.0005 |
| | CSA | 8.4025 | 0.1697 | 0.0480 | 0.9326 | 0.0003 |
| | ACS | 3.4792 | 0.0786 | 0.0216 | 0.9851 | 0.0004 |
| | BSA | 3.5277 | 0.0695 | 0.0191 | 0.9885 | 0.0005 |
| Exponential | PSO | 1.1652 | 0.0309 | 0.0083 | 0.9957 | 0.0005 |
| | CSA | 3.1172 | 0.0633 | 0.0174 | 0.9884 | 0.0003 |
| | ACS | 1.1506 | 0.0335 | 0.0090 | 0.9963 | 0.0005 |
| | BSA | 3.7468 | 0.0808 | 0.0222 | 0.9842 | 0.0006 |
| GEP | RNC | 1.6752 | 0.0432 | 0.0116 | 0.9945 | 0.0005 |
| Optimized GEP | PSO | 1.1850 | 0.0282 | 0.0076 | 0.9947 | 0.0004 |
| | CSA | 1.1870 | 0.0280 | 0.0075 | 0.9925 | 0.0004 |
| | ACS | 1.1774 | 0.0390 | 0.0106 | 0.9953 | 0.0004 |
| | BSA | 1.1136 | 0.0262 | 0.0070 | 0.9952 | 0.0004 |

$$\overline{EEC}(t)_{GEP-BSA} =$$

$$(1.112)\left( \frac{-\sqrt{\overline{EEC}(t-2)\overline{EEC}(t-1)} - exp\left((-43.31)\overline{EEC}(t-1)\right) - \frac{\left(\overline{EEC}(t-3)\right)^3}{138.438} +}{\overline{EEC}(t-2) + \overline{EEC}(t-1)} \right)$$

$$- (0.095) \qquad (4.27)$$
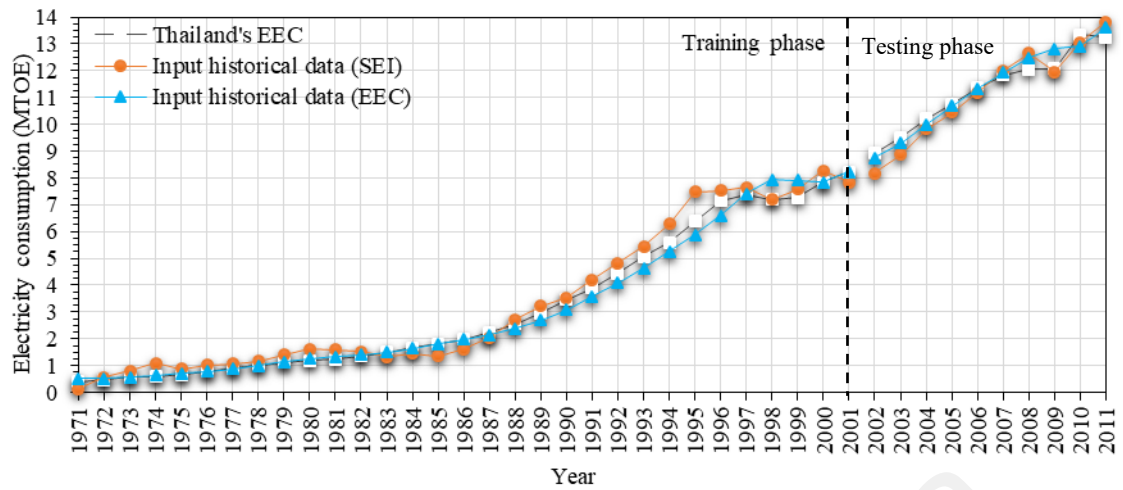
**Figure 4.15: Thailand's EEC actual data from 1971 to 2011 and GEP-BSA performances during training of design phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types (SEI and EEC)**

As shown in Tables 4.17 and 4.18, the performances of optimized GEP models for long-term EEC forecasting in Philippines are quantified in terms of MAPE, RMSE, $U$-statistic and $R$-squared indexes and the computed values are compared with those from other soft computing approaches. Furthermore, to ensure that the obtained models adequately describe the given set of data the RACF values are calculated for all models. It is observed that all obtained models adequately describe the given data sets, as the RACF values reported in these tables demonstrate that the estimated residuals for all models are uncorrelated.

In this case, the forecasting accuracy of studied methods with two different input data types based on mean rank of models for MAPE index is ranked as GEP$_{-BSA}$ > GEP$_{-ACS}$ > GEP$_{-PSO}$ > GEP$_{-CSA}$ > quadratic$_{-PSO}$ > quadratic$_{-ACS}$ > exponential$_{-ACS}$ > M5-R = exponential$_{-CSA}$ > exponential$_{-BSA}$ > ANN = ANFIS$_{-SC}$ > GEP > linear$_{-ACS}$ > linear$_{-PSO}$ > SVR > linear$_{-BSA}$ > linear$_{-CSA}$ = quadratic$_{-BSA}$ > exponential$_{-PSO}$ > ANFIS$_{-FCMC}$ > quadratic$_{-CSA}$. As the results show, the optimized GEP approaches provide better-fit estimation than other studied methods. Additionally, it is observed that the most efficient optimization algorithm for training GEP and the most effective input data set for modeling the annual EEC in Philippines are BSA and EEC respectively, as the superior reported results in

142

Tables 4.17 and 4.18 belong to GEP-BSA based on EEC in preceding three years. The mathematical models created by GEP-BSA for annual EEC in Philippines on the basis of socio-economic indicators and EEC in preceding years are given by Eqs. (4.28) and (4.29) respectively, and their performances are illustrated in Figure 4.16.

**Table 4.17: Comparison between forecasting accuracy of studied methods on Philippines's EEC based on SEI**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | SEI | | | | |
| ANN | MLP | 2.4528 | 0.0621 | 0.0172 | 0.9828 | 0.0172 |
| support vector machines | SVR | 7.9029 | 0.1508 | 0.0431 | 0.9241 | 0.0431 |
| ANFIS | FCMC | 7.1802 | 0.1787 | 0.0510 | 0.9064 | 0.0510 |
| | SC | 2.7641 | 0.0537 | 0.0145 | 0.9910 | 0.0145 |
| Rule-based data mining | M5-R | 2.8117 | 0.0555 | 0.0153 | 0.9863 | 0.0153 |
| Multiple linear | PSO | 6.9359 | 0.1392 | 0.0396 | 0.9328 | 0.0396 |
| | CSA | 5.9640 | 0.1197 | 0.0331 | 0.9455 | 0.0331 |
| | ACS | 6.5111 | 0.1232 | 0.0350 | 0.9457 | 0.0335 |
| | BSA | 6.9359 | 0.1392 | 0.0396 | 0.9328 | 0.0396 |
| Quadratic | PSO | 2.9362 | 0.0613 | 0.0168 | 0.9819 | 0.0168 |
| | CSA | 9.3903 | 0.1882 | 0.0503 | 0.8859 | 0.0503 |
| | ACS | 4.4576 | 0.0966 | 0.0270 | 0.9632 | 0.0236 |
| | BSA | 6.5858 | 0.1756 | 0.0469 | 0.9060 | 0.0469 |
| Exponential | PSO | 6.9488 | 0.1331 | 0.0364 | 0.9379 | 0.0050 |
| | CSA | 4.6084 | 0.1046 | 0.0293 | 0.9560 | 0.0018 |
| | ACS | 4.9384 | 0.1097 | 0.0306 | 0.9526 | 0.0021 |
| | BSA | 5.6783 | 0.1160 | 0.0327 | 0.9506 | 0.0016 |
| GEP | RNC | 5.3862 | 0.1062 | 0.0287 | 0.9613 | 0.0287 |
| Optimized GEP | PSO | 2.8362 | 0.0562 | 0.0155 | 0.9860 | 0.0155 |
| | CSA | 3.1573 | 0.0772 | 0.0214 | 0.9745 | 0.0214 |
| | ACS | 2.8117 | 0.0555 | 0.0153 | 0.9863 | 0.0150 |
| | BSA | 1.6177 | 0.0371 | 0.0101 | 0.9927 | 0.0101 |

$$\overline{EEC}(t)_{GEP-BSA} =$$

$$(1.502)\left( \overline{EXP}(t) + (0.349)\left( log\left( \frac{1}{\overline{IMP}(t)} \right)^3 \right) exp\left( \sqrt{\overline{IMP}(t)} \right) \right)^{0.75} \qquad (4.28)$$

$$+ (1.17)\left( \overline{POP}(t) - \left( \overline{POP}(t) \right)^{-0.5} - 0.5 \right)^3 + \frac{0.664}{\overline{GDP}(t)} - 1.016$$

**Table 4.18: Comparison between forecasting accuracy of studied methods on Philippines's EEC based on EEC in preceding years**

| Model | Methods | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
|---|---|---|---|---|---|---|
| | | Input historical data | | | | |
| | | EEC | | | | |
| ANN | MLP | 8.0189 | 0.1625 | 0.0466 | 0.9206 | 0.0466 |
| support vector machines | SVR | 1.5659 | 0.0325 | 0.0089 | 0.9921 | 0.0089 |
| ANFIS | FCMC | 3.9330 | 0.0837 | 0.0235 | 0.9764 | 0.0235 |
| | SC | 7.9228 | 0.1850 | 0.0530 | 0.8982 | 0.0530 |
| Rule-based data mining | M5-R | 2.3669 | 0.0574 | 0.0159 | 0.9604 | 0.0159 |
| Multiple linear | PSO | 2.0260 | 0.0404 | 0.0110 | 0.9904 | 0.0110 |
| | CSA | 4.3170 | 0.0885 | 0.0238 | 0.9686 | 0.0238 |
| | ACS | 2.0260 | 0.0404 | 0.0110 | 0.9904 | 0.0110 |
| | BSA | 2.0260 | 0.0404 | 0.0110 | 0.9904 | 0.0110 |
| Quadratic | PSO | 1.2871 | 0.0343 | 0.0094 | 0.9918 | 0.0094 |
| | CSA | 6.1602 | 0.1287 | 0.0342 | 0.9402 | 0.0342 |
| | ACS | 1.3965 | 0.0368 | 0.0101 | 0.9912 | 0.0102 |
| | BSA | 3.5364 | 0.0742 | 0.0200 | 0.9795 | 0.0200 |
| Exponential | PSO | 3.0411 | 0.0646 | 0.0175 | 0.9831 | 0.0008 |
| | CSA | 1.6135 | 0.0327 | 0.0089 | 0.9917 | 0.0006 |
| | ACS | 1.4638 | 0.0333 | 0.0091 | 0.9876 | 0.0005 |
| | BSA | 1.9291 | 0.0382 | 0.0104 | 0.9910 | 0.0007 |
| GEP | RNC | 2.0768 | 0.0472 | 0.0131 | 0.9856 | 0.0131 |
| Optimized GEP | PSO | 1.1239 | 0.0324 | 0.0089 | 0.9917 | 0.0089 |
| | CSA | 1.1238 | 0.0324 | 0.0089 | 0.9917 | 0.0089 |
| | ACS | 1.1237 | 0.0324 | 0.0089 | 0.9917 | 0.0089 |
| | BSA | 1.0951 | 0.0308 | 0.0084 | 0.9921 | 0.0084 |

$$\overline{EEC}(t)_{GEP-BSA} = \frac{-3.088}{\left((-16)\left(\overline{EEC}(t-1)\right)^4 - \overline{EEC}(t-1)\overline{EEC}(t-3) + \overline{EEC}(t-2) + 1.742\right)^2} + (1.006)\overline{EEC}(t-1) + 0.0151 \tag{4.29}$$
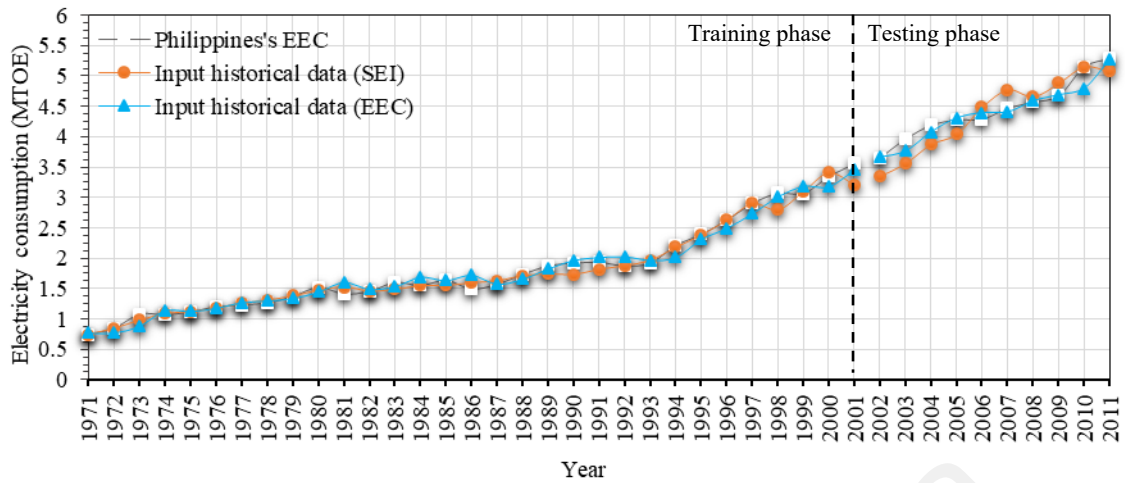
**Figure 4.16: Philippines's EEC actual data from 1971 to 2011 and GEP-BSA performances during training of design phase (1971-2001) and testing phase (2002-2011) based on two different input historical data types (SEI and EEC)**

The minimum, average, and maximum values of MAPE indicator obtained after 50 trials run of applied optimization methods on the most performant models (optimized GEP) are tabulated in Table 4.19. The robustness of the developed models is confirmed according to the obtained results in this table.

**Table 4.19: The minimum, average, and maximum values for MAPE indicator of optimized GEP methods**

| ASEAN-5 Countries | Input historical data type | MAPE Index | Methods | | | |
|---|---|---|---|---|---|---|
| | | | PSO | CSA | ACS | BSA |
| Malaysia | SEI | Max MAPE (%) | 3.5866 | 3.3560 | 3.2448 | 3.2497 |
| | | Mean MAPE (%) | 3.5698 | 3.2045 | 3.1045 | 3.0036 |
| | | Min MAPE (%) | 3.3306 | 3.0735 | 3.0700 | 2.8935 |
| | EEC | Max MAPE (%) | 1.6005 | 1.5999 | 1.5956 | 15952 |
| | | Mean MAPE (%) | 1.5987 | 1.5991 | 1.5961 | 1.5950 |
| | | Min MAPE (%) | 1.5956 | 1.5962 | 1.5893 | 1.5882 |
| Indonesia | SEI | Max MAPE (%) | 2.7456 | 2.8654 | 2.5846 | 2.5646 |
| | | Mean MAPE (%) | 2.5569 | 2.5689 | 2.4564 | 2.1546 |
| | | Min MAPE (%) | 2.4381 | 2.4720 | 2.3354 | 2.0545 |
| | EEC | Max MAPE (%) | 0.8972 | 0.9961 | 0.8551 | 0.8539 |
| | | Mean MAPE (%) | 0.8765 | 0.8958 | 0.8549 | 0.8539 |
| | | Min MAPE (%) | 0.8754 | 0.8553 | 0.8545 | 0.8538 |
| Singapore | SEI | Max MAPE (%) | 2.0134 | 2.2121 | 2.0987 | 1.9874 |
| | | Mean MAPE (%) | 1.9564 | 1.9856 | 1.9776 | 1.8236 |
| | | Min MAPE (%) | 1.7855 | 1.8391 | 1.8656 | 1.7537 |
| | EEC | Max MAPE (%) | 0.7447 | 0.8849 | 0.7441 | 0.7446 |
| | | Mean MAPE (%) | 0.7445 | 0.7947 | 0.7441 | 0.7444 |
| | | Min MAPE (%) | 0.7442 | 0.7442 | 0.7441 | 0.7438 |
| Thailand | SEI | Max MAPE (%) | 2.7654 | 3.0012 | 2.2458 | 2.4145 |
| | | Mean MAPE (%) | 2.2148 | 2.8987 | 2.0236 | 1.9754 |
| | | Min MAPE (%) | 1.9466 | 2.3729 | 1.9366 | 1.6996 |
| | EEC | Max MAPE (%) | 1.5564 | 1.5745 | 1.4698 | 1.4669 |
| | | Mean MAPE (%) | 1.3659 | 1.3398 | 1.3254 | 1.2365 |
| | | Min MAPE (%) | 1.1850 | 1.1870 | 1.1774 | 1.1136 |
| Philippines | SEI | Max MAPE (%) | 3.5569 | 3.3654 | 3.0214 | 2.1056 |
| | | Mean MAPE (%) | 3.0212 | 3.2236 | 2.9874 | 1.8803 |
| | | Min MAPE (%) | 2.8362 | 3.1573 | 2.8117 | 1.6177 |
| | EEC | Max MAPE (%) | 1.5249 | 1.1284 | 1.1244 | 1.1578 |
| | | Mean MAPE (%) | 1.3642 | 1.1269 | 1.1242 | 1.1254 |
| | | Min MAPE (%) | 1.1239 | 1.1238 | 1.1237 | 1.0951 |

### 4.3.1    Validation of the Model Using Statistical Methods

Different statistical methods are applied as external validation to verify the validity of mathematical models developed by GEP$_{-BSA}$. To evaluate the performance of the obtained model the following attributes were recommended (Mousavi et al., 2014):

- i-    If a model gives $|R| > 0.8$, a strong correlation exists between the predicted and observed values.

- ii-   If a model gives $0.2 < |R| < 0.8$, a correlation exists between the predicted and observed values.

- iii-  If a model gives $|R| < 0.2$, a weak correlation exists between the predicted and observed values.

In addition, new factors suggested by (Golbraikh & Tropsha, 2002) are checked for external validation of the obtained models on the testing phase. It is recommended that at least one slope of the regression lines ($k$ or $k'$) through the origin should be close to one. It should be noted that $k$ and $k'$ are the slopes of the regression lines between the regressions of actual output ($h_i$) against predicted output ($t_i$) or $t_i$ against $h_i$ through the origin, i.e. $h_i = k\, t_i$ and $t_i = k' h_i$, respectively. In addition, the performance indexes of $m$ and $n$ should be less than 0.1 ($m$ and $n$ are the two factors for evaluating the model performance). Recently, Roy and Roy (P. P. Roy & Roy, 2008) presented a confirmed indicator ($R_m$) for external predictability of models. For $R_m > 0.5$, the condition is satisfied. Either the squared correlation coefficient (through the origin) between predicted and experimental values ($R_o{}^2$), or the squared correlation coefficient between experimental and predicted values ($R_o'{}^2$) should be close to $R^2$ and to one (Alavi, Aminian, Gandomi, & Esmaeili, 2011; Mostafavi et al., 2013; Mostafavi, Mousavi, & Hosseinpour, 2014). In item one, $R$ should be greater than 0.8. In the second item, $k$ should be between 0.85 and 1.15. In the third item, $k'$ should be between 0.85 and 1.15. According to items four and five, $m$ and $n$ values should be smaller than 0.1. Finally $R_m$

should be greater than 0.5. The statistical factors of the GEP-BSA models for formulating the EEC of ASEAN-5 countries based on two different input historical data types (SEI and EEC) are tabulated in Tables 4.20 - 4.24. As shown, the developed models satisfy all the requisite conditions. The validation phase ensures that GEP-BSA provides precise models, which are strongly applicable for long-term EEC forecasting of ASEAN-5 countries.

**Table 4.20: Statistical factors of the GEP-BSA model for formulating the EEC of Malaysia based on two different input historical data types (SEI and EEC)**

| Item | Formula | Condition | SEI | EEC |
|---|---|---|---|---|
| 1 | $R$ | $0.8 < R$ | 0.9935 | 0.9966 |
| 2 | $k = \dfrac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} h_i{}^2}$ | $0.85 < k < 1.15$ | 0.9981 | 0.9959 |
| 3 | $k' = \dfrac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} t_i{}^2}$ | $0.85 < k' < 1.15$ | 1.0016 | 1.0038 |
| 4 | $m = \dfrac{R^2 - R_o{}^2}{R^2}$ | $|m| < 0.1$ | -0.0118 | -0.0056 |
| 5 | $n = \dfrac{R^2 - R_{o'}{}^2}{R^2}$ | $|n| < 0.1$ | -0.0120 | -0.0057 |
| 6 | $R_m = R^2 \times \left(1 - \sqrt{\left|R^2 - R_o{}^2\right|}\right)$ | $0.5 < R_m$ | 0.9763 | 0.9881 |
| Where | $R_o{}^2 = 1 - \dfrac{\sum_{i=1}^{n}(t_i - h_i{}^o)^2}{\sum_{i=1}^{n}(t_i - \bar{t_i})^2}, h_i{}^o = k \times t_i$ | $0.8 < R_o{}^2 < 1$ | 0.9998 | 0.9996 |
| | $R_{o'}{}^2 = 1 - \dfrac{\sum_{i=1}^{n}(h_i - t_i{}^o)^2}{\sum_{i=1}^{n}(h_i - \bar{h_i})^2}, t_i{}^o = k' \times h_i$ | $0.8 < R_{o'}{}^2 < 1$ | 0.9999 | 0.9997 |

**Table 4.21: Statistical factors of the GEP-BSA model for formulating the EEC of Indonesia based on two different input historical data types (SEI and EEC)**

| Item | Formula | Condition | SEI | EEC |
|---|---|---|---|---|
| 1 | $R$ | $0.8 < R$ | 0.9913 | 0.9992 |
| 2 | $k = \dfrac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} h_i{}^2}$ | $0.85 < k < 1.15$ | 0.9992 | 0.9989 |
| 3 | $k' = \dfrac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} t_i{}^2}$ | $0.85 < k' < 1.15$ | 1.0006 | 1.0010 |
| 4 | $m = \dfrac{R^2 - R_o{}^2}{R^2}$ | $|m| < 0.1$ | -0.0043 | -0.0015 |
| 5 | $n = \dfrac{R^2 - R_{o'}{}^2}{R^2}$ | $|n| < 0.1$ | -0.0043 | -0.0015 |
| 6 | $R_m = R^2 \times \left(1 - \sqrt{\left|R^2 - R_o{}^2\right|}\right)$ | $0.5 < R_m$ | 0.9914 | 0.9970 |

| Where | $R_o{}^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - h_i{}^o)^2}{\sum_{i=1}^{n}(t_i - \bar{t_i})^2}, h_i{}^o = k \times t_i$ | $0.8 < R_o{}^2 < 1$ | 1.0000 | 1.0000 |
|---|---|---|---|---|
| | $R_o{'}^2 = 1 - \frac{\sum_{i=1}^{n}(h_i - t_i{}^o)^2}{\sum_{i=1}^{n}(h_i - \bar{h_i})^2}, t_i{}^o = k' \times h_i$ | $0.8 < R_o{'}^2 < 1$ | 1.0000 | 1.0000 |

**Table 4.22: Statistical factors of the GEP-BSA model for formulating the EEC of Singapore based on two different input historical data types (SEI and EEC)**

| Item | Formula | Condition | SEI | EEC |
|---|---|---|---|---|
| 1 | $R$ | $0.8 < R$ | 0.9972 | 0.9992 |
| 2 | $k = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} h_i{}^2}$ | $0.85 < k < 1.15$ | 1.0013 | 0.9983 |
| 3 | $k' = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} t_i{}^2}$ | $0.85 < k' < 1.15$ | 0.9984 | 1.0016 |
| 4 | $m = \frac{R^2 - R_o{}^2}{R^2}$ | $|m| < 0.1$ | -0.0055 | -0.0016 |
| 5 | $n = \frac{R^2 - R_o{'}^2}{R^2}$ | $|n| < 0.1$ | -0.0055 | -0.0016 |
| 6 | $R_m = R^2 \times \left(1 - \sqrt{|R^2 - R_o{}^2|}\right)$ | $0.5 < R_m$ | 0.9890 | 0.9967 |
| Where | $R_o{}^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - h_i{}^o)^2}{\sum_{i=1}^{n}(t_i - \bar{t_i})^2}, h_i{}^o = k \times t_i$ | $0.8 < R_o{}^2 < 1$ | 1.0000 | 0.9999 |
| | $R_o{'}^2 = 1 - \frac{\sum_{i=1}^{n}(h_i - t_i{}^o)^2}{\sum_{i=1}^{n}(h_i - \bar{h_i})^2}, t_i{}^o = k' \times h_i$ | $0.8 < R_o{'}^2 < 1$ | 1.0000 | 0.9999 |

**Table 4.23: Statistical factors of the GEP-BSA model for formulating the EEC of Thailand based on two different input historical data types (SEI and EEC)**

| Item | Formula | Condition | SEI | EEC |
|---|---|---|---|---|
| 1 | $R$ | $0.8 < R$ | 0.9960 | 0.9977 |
| 2 | $k = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} h_i{}^2}$ | $0.85 < k < 1.15$ | 1.0050 | 0.9991 |
| 3 | $k' = \frac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} t_i{}^2}$ | $0.85 < k' < 1.15$ | 0.9946 | 1.0006 |
| 4 | $m = \frac{R^2 - R_o{}^2}{R^2}$ | $|m| < 0.1$ | -0.0071 | -0.0045 |
| 5 | $n = \frac{R^2 - R_o{'}^2}{R^2}$ | $|n| < 0.1$ | -0.0070 | -0.0045 |
| 6 | $R_m = R^2 \times \left(1 - \sqrt{|R^2 - R_o{}^2|}\right)$ | $0.5 < R_m$ | 0.9850 | 0.9910 |
| Where | $R_o{}^2 = 1 - \frac{\sum_{i=1}^{n}(t_i - h_i{}^o)^2}{\sum_{i=1}^{n}(t_i - \bar{t_i})^2}, h_i{}^o = k \times t_i$ | $0.8 < R_o{}^2 < 1$ | 0.9995 | 1.0000 |
| | $R_o{'}^2 = 1 - \frac{\sum_{i=1}^{n}(h_i - t_i{}^o)^2}{\sum_{i=1}^{n}(h_i - \bar{h_i})^2}, t_i{}^o = k' \times h_i$ | $0.8 < R_o{'}^2 < 1$ | 0.9995 | 1.0000 |

**Table 4.24: Statistical factors of the GEP-BSA model for formulating the EEC of Philippines based on two different input historical data types (SEI and EEC)**

| Item | Formula | Condition | SEI | EEC |
|------|---------|-----------|-----|-----|
| 1 | $R$ | $0.8 < R$ | 0.9963 | 0.9964 |
| 2 | $k = \dfrac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} h_i^2}$ | $0.85 < k < 1.15$ | 0.9934 | 0.9956 |
| 3 | $k' = \dfrac{\sum_{i=1}^{n}(h_i \times t_i)}{\sum_{i=1}^{n} t_i^2}$ | $0.85 < k' < 1.15$ | 1.0061 | 1.0041 |
| 4 | $m = \dfrac{R^2 - R_o^2}{R^2}$ | $|m| < 0.1$ | -0.0052 | -0.0063 |
| 5 | $n = \dfrac{R^2 - R_{o'}^2}{R^2}$ | $|n| < 0.1$ | -0.0056 | -0.0065 |
| 6 | $R_m = R^2 \times \left(1 - \sqrt{|R^2 - R_o^2|}\right)$ | $0.5 < R_m$ | 0.9875 | 0.9866 |
| Where | $R_o^2 = 1 - \dfrac{\sum_{i=1}^{n}(t_i - h_i^o)^2}{\sum_{i=1}^{n}(t_i - \bar{t_i})^2}, h_i^o = k \times t_i$ | $0.8 < R_o^2 < 1$ | 0.9990 | 0.9995 |
| | $R_{o'}^2 = 1 - \dfrac{\sum_{i=1}^{n}(h_i - t_i^o)^2}{\sum_{i=1}^{n}(h_i - \bar{h_i})^2}, t_i^o = k' \times h_i$ | $0.8 < R_{o'}^2 < 1$ | 0.9991 | 0.9996 |

## 4.3.2 Long-term Electrical Energy Consumption Forecasting

More importantly, it is also of interest in this study to determine the accuracy of optimized GEP approaches for long-term EEC forecasting of ASEAN-5 countries based on two different input data types and anticipate their annual growth rate up to 2030 using the most accurate methodology.

Although ASEAN-5 countries inherently have different dependencies on EEC, the reported results indicate that GEP-BSA on the basis of EEC in comparison with other studied methodologies offers sufficient accuracy to be utilized for long-term EEC forecasting. Thus, future estimations of EEC in ASEAN-5 countries are projected up to 2030 by applying the rolling forecast on mathematical models developed by GEP-BSA on the basis of EEC in preceding three years. The flow chart of optimized GEP for long-term EEC forecasting is illustrated in Figure 4.17.
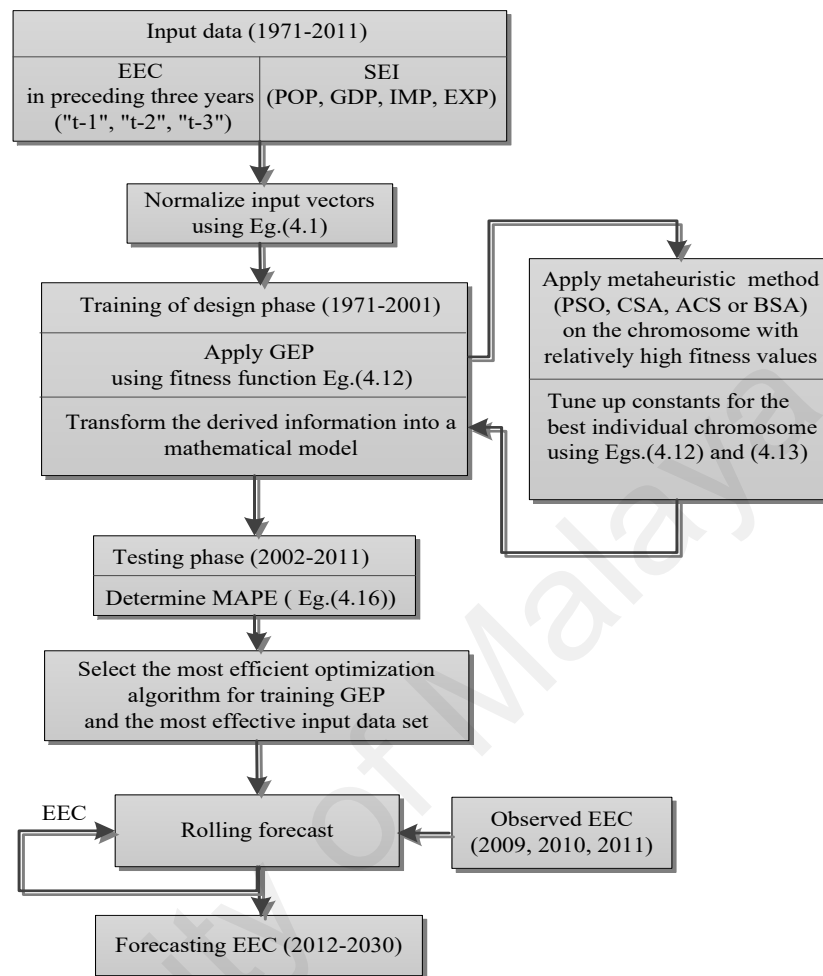
Input data (1971-2011)

| EEC in preceding three years ("t-1", "t-2", "t-3") | SEI (POP, GDP, IMP, EXP) |

Normalize input vectors using Eg.(4.1)

Training of design phase (1971-2001)

Apply GEP using fitness function Eg.(4.12)

Transform the derived information into a mathematical model

Apply metaheuristic method (PSO, CSA, ACS or BSA) on the chromosome with relatively high fitness values

Tune up constants for the best individual chromosome using Egs.(4.12) and (4.13)

Testing phase (2002-2011)

Determine MAPE ( Eg.(4.16))

Select the most efficient optimization algorithm for training GEP and the most effective input data set

EEC → Rolling forecast ← Observed EEC (2009, 2010, 2011)

Forecasting EEC (2012-2030)

**Figure 4.17: The procedure of optimized GEP for long-term EEC forecasting**

A rolling forecast is a first-in, first-out (FIFO) process for projecting the future over a time period. Rolling forecast is often used either in the short-term or in long-term forecasting as it rolls the forecast toward the next period of time (e.g. year, month, day, hour and minute) (Hong, 2009). Figure 4.18 shows the rolling-based forecasting procedure. As shown in this figure, a rolling forecast's drop/add process ensures that the number of periods in a rolling forecast window remains constant. Because a rolling forecast window requires consecutive revisions, it is also known as an iterative forecast or a continuous forecast.
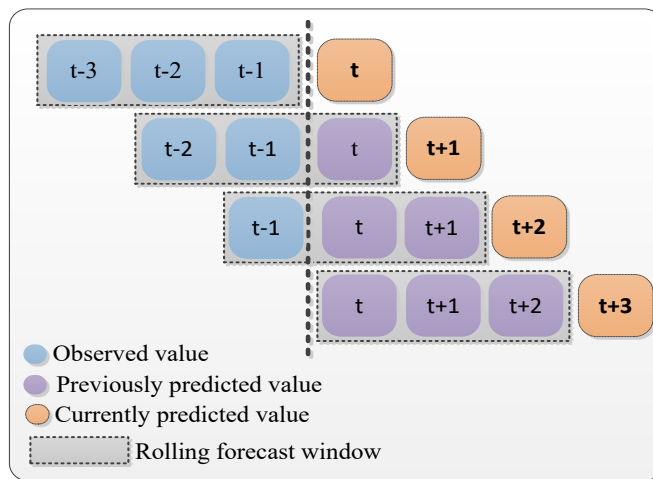
**Figure 4.18: The rolling-based forecasting procedure**

To anticipate the annual EEC in ASEAN-5 countries up to 2030, the rolling-based procedure is implemented on GEP$_{-BSA}$ mathematical models while EEC in preceding three years is considered as a rolling forecast window. Firstly, the primary observed samples (EEC in 2009, 2010, and 2011) are fed into the GEP$_{-BSA}$ model, then the EEC in the year of 2012 is forecasted. Secondly, the forecast window rolls one-step ahead and the observations from 2010 to 2011, and the forecasted value for 2012 are similarly again fed into the GEP$_{-BSA}$ model, then EEC in the year of 2013 is forecasted. The annual forecasts for EEC up to 2030 are obtained through rolling the forecast window until 2029.

Furthermore, EEC in ASEAN-5 countries is also forecasted up to 2030 by another two different time series forecasting methods, namely ARIMA and GM (1, 1), and their forecasts are compared with those obtained by proposed method. The performances of applied time series methods for long-term EEC forecasting of ASEAN-5 countries is tabulated in Table 4.25. The RACF values in this table confirm the whiteness of estimated residuals at a confidence interval level for all obtained models. Regardless of which country is studied, the enhanced results in Table 4.25 confirm that the developed model by GEP$_{-BSA}$ provides better-fit estimations in comparison with other applied time series forecasting methods.

**Table 4.25: Comparison between forecasting accuracy of ARIMA, GM (1, 1), and GEP-BSA models for EEC of ASEAN-5 countries**

| ASEAN-5 countries | Methods | Performance indexes | | | | |
|---|---|---|---|---|---|---|
| | | MAPE (%) | RMSE | $U$-statistic | $R^2$ | RACF |
| Malaysia | ARIMA | 3.1337 | 0.0931 | 0.0272 | 0.9750 | 0.0002 |
| | GM (1,1) | 3.3584 | 0.0965 | 0.0282 | 0.9731 | 0.0002 |
| | GEP-BSA | 1.5882 | 0.0458 | 0.0133 | 0.9933 | 0.0014 |
| Indonesia | ARIMA | 1.5342 | 0.0287 | 0.0081 | 0.9977 | 0.0006 |
| | GM (1,1) | 2.6985 | 0.0586 | 0.0167 | 0.9906 | 0.0002 |
| | GEP-BSA | 0.8538 | 0.0170 | 0.0048 | 0.9984 | 0.0007 |
| Singapore | ARIMA | 1.5064 | 0.0334 | 0.0090 | 0.9973 | 0.0002 |
| | GM (1,1) | 1.4341 | 0.0285 | 0.0076 | 0.9980 | 0.0003 |
| | GEP-BSA | 0.7438 | 0.0221 | 0.0059 | 0.9983 | 0.0003 |
| Thailand | ARIMA | 2.1290 | 0.0435 | 0.0119 | 0.9902 | 0.0004 |
| | GM (1,1) | 1.3575 | 0.0318 | 0.0085 | 0.9923 | 0.0003 |
| | GEP-BSA | 1.1136 | 0.0262 | 0.0070 | 0.9952 | 0.0004 |
| Philippines | ARIMA | 1.2251 | 0.0253 | 0.0069 | 0.9912 | 0.0003 |
| | GM (1,1) | 1.3258 | 0.0362 | 0.0099 | 0.9901 | 0.0002 |
| | GEP-BSA | 1.0951 | 0.0308 | 0.0084 | 0.9921 | 0.0084 |

The forecasted EEC in ASEAN-5 countries based on ARIMA, GM (1, 1) and GEP-BSA models are shown in Figures 4.19 - 4.23.

It is determined that based on ARIMA model the average annual growth rates (AAGRs) of EEC from 2011 until 2030 for Malaysia, Indonesia, Singapore, Thailand and Philippines are 6.29%, 4.39%, 1.7%, 3.71% and 3.03%, respectively.

GM (1, 1) provides the AAGRs of EEC from 2011 until 2030 for Malaysia, Indonesia, Singapore, Thailand and Philippines, respectively as: 2.67%, 2.91%, 3.98%, 3.53% and 2.5%.

By utilizing GEP-BSA the AAGRs of EEC from 2011 until 2030 for Malaysia, Indonesia, Singapore, Thailand and Philippines are obtained as 3.52%, 2.9%, 1.83%, 3.35% and 2.02%, respectively. According to the obtained results based on GEP-BSA models, it can be concluded that highest AAGRs of EEC in this region will be belong to Malaysia, Thailand, Indonesia, Philippines and Singapore respectively.
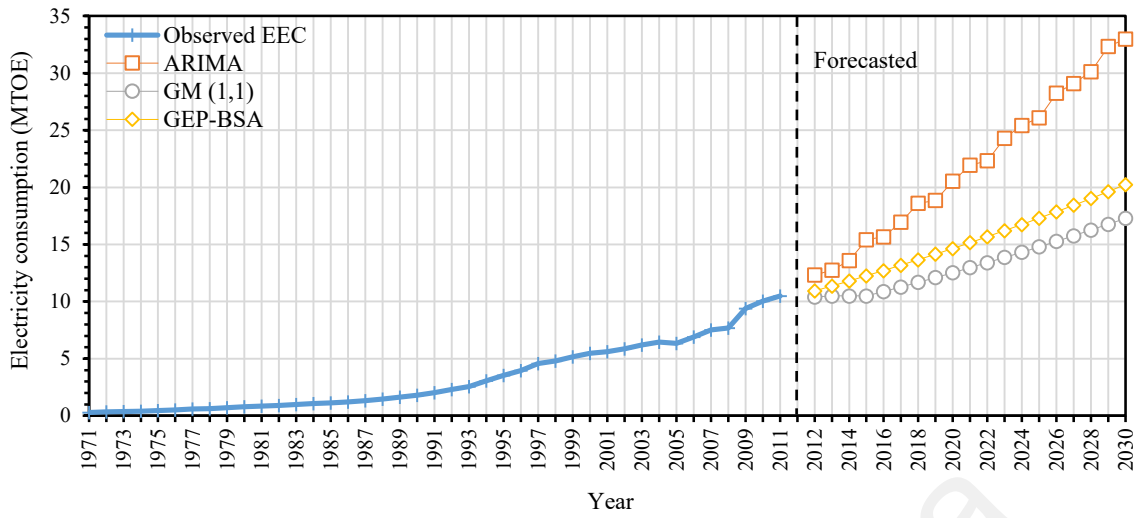
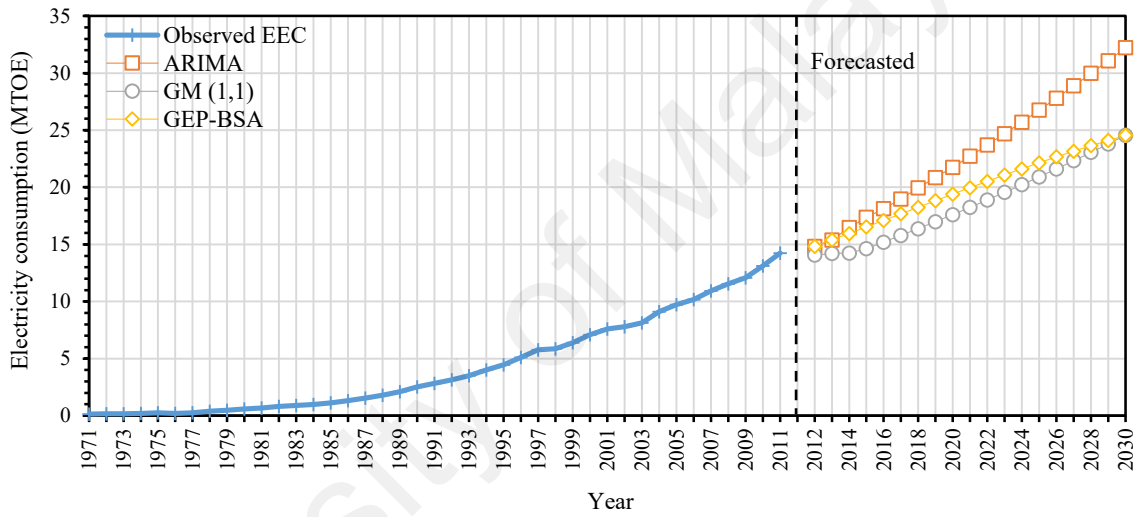**Figure 4.19: Future projection of Malaysia' EEC up to 2030**



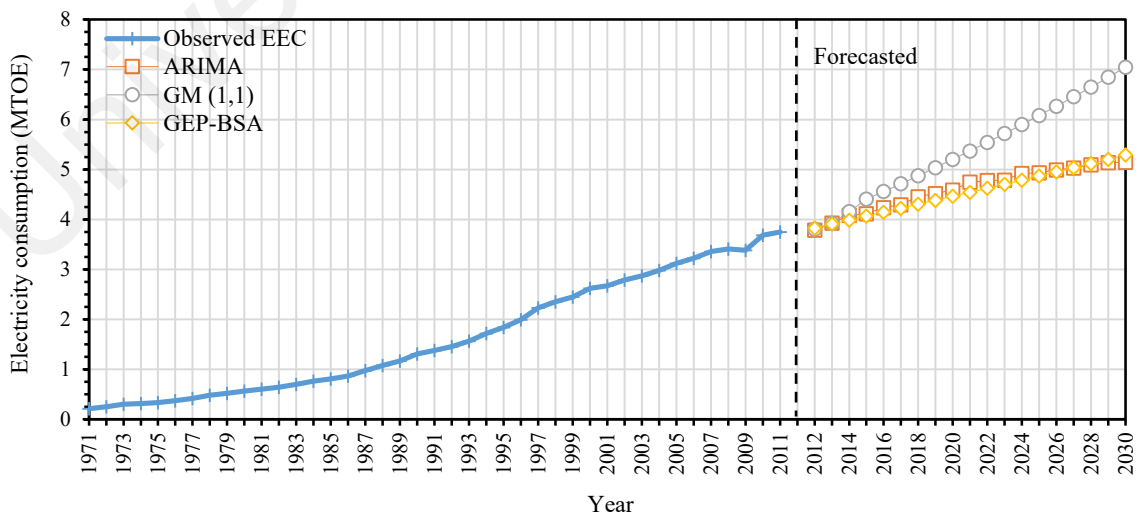**Figure 4.20: Future projection of Indonesia' EEC up to 2030**



**Figure 4.21: Future projection of Singapore's EEC up to 2030**
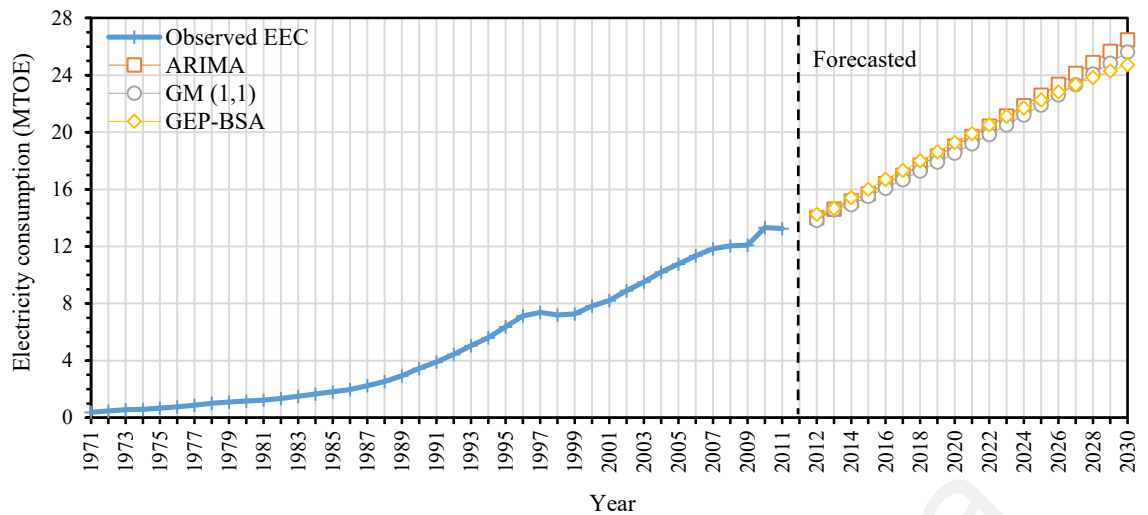
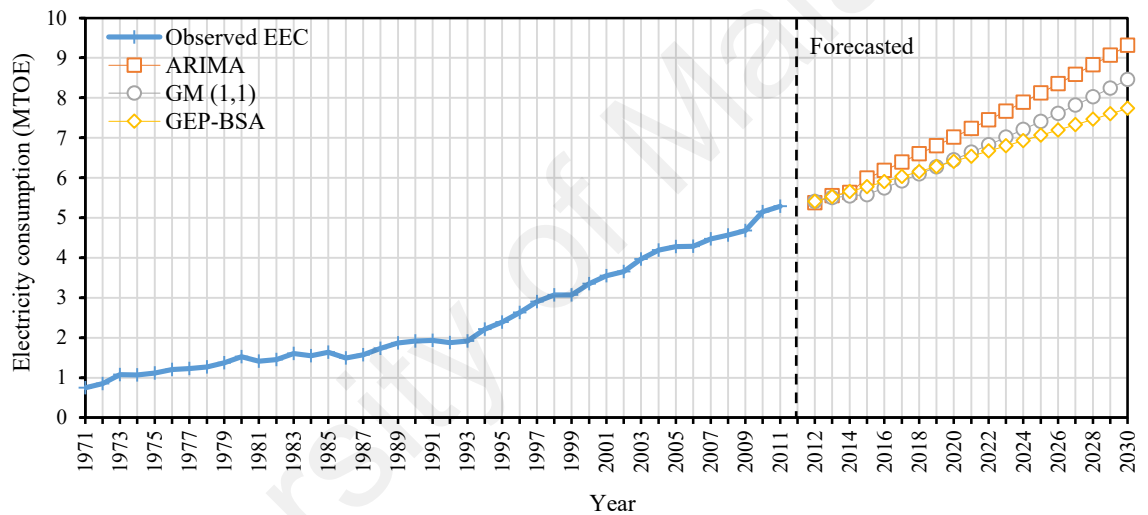**Figure 4.22: Future projection of Thailand's EEC up to 2030**



**Figure 4.23: Future projection of Philippines's EEC up to 2030**

To access the forecasting accuracy of applied time series methods the annual data from 2012 to 2014 inclusive is considered as unseen dataset (hold-out dataset). The performances of applied methods for long-term EEC forecasting of ASEAN-5 countries on unseen dataset are tabulated in Table 4.26. The collected results in this table indicate that, regardless of which country is studied, GEP-BSA in comparison with other studied methods offers sufficient accuracy to be utilized for long-term EEC forecasting.

**Table 4.26: Annual forecasted EEC in ASEAN-5 countries based on applied time series forecasting methods**

| Methods | 2012 (MTOE) | Relative error | 2013 (MTOE) | Relative error | 2014 (MTOE) | Relative error | RMSE | $U$-statistic | MAPE% |
|---|---|---|---|---|---|---|---|---|---|
| Malaysia | | | | | | | | | |
| ARIMA | 12.3351 | -0.1375 | 12.7500 | -0.1131 | 13.5975 | -0.1382 | 2.1936 | 0.0599 | 12.9625 |
| GM(1,1) | 10.3987 | 0.0410 | 10.5001 | 0.0832 | 10.5007 | 0.1210 | 1.3434 | 0.0404 | 8.17816 |
| GEP-BSA | 10.9279 | -0.0077 | 11.3619 | 0.0080 | 11.8043 | 0.0118 | 0.1492 | 0.004 | 0.9229 |
| Indonesia | | | | | | | | | |
| ARIMA | 14.506 | 0.0741 | 15.3911 | 0.0815 | 16.3034 | 0.0846 | 1.9934 | 0.0412 | 8.0102 |
| GM(1,1) | 14.0571 | 0.1027 | 14.1982 | 0.1527 | 14.2486 | 0.2000 | 3.6570 | 0.0782 | 15.1836 |
| GEP-BSA | 14.8256 | 0.0537 | 15.3877 | 0.0817 | 15.9533 | 0.1042 | 1.9328 | 0.0399 | 7.9922 |
| Singapore | | | | | | | | | |
| ARIMA | 3.7859 | 0.0444 | 3.8254 | 0.0507 | 4.0817 | 0.0187 | 0.2736 | 0.0227 | 3.7991 |
| GM(1,1) | 3.8000 | 0.0409 | 3.9100 | 0.0297 | 4.3164 | -0.037 | 0.2210 | 0.0182 | 3.6115 |
| GEP-BSA | 3.8268 | 0.0341 | 3.9071 | 0.0304 | 3.9869 | 0.0415 | 0.2081 | 0.0172 | 3.5386 |
| Thailand | | | | | | | | | |
| ARIMA | 14.0152 | 0.0166 | 14.6250 | 0.0074 | 15.2098 | -0.0180 | 0.3048 | 0.0068 | 1.4070 |
| GM(1,1) | 13.8341 | 0.0294 | 14.5815 | 0.0104 | 14.9402 | -1.76e-5 | 0.4462 | 0.0101 | 1.3277 |
| GEP-BSA | 14.2562 | -0.0002 | 14.6596 | 0.0051 | 15.4422 | -0.0336 | 0.2996 | 0.0067 | 1.2979 |
| Philippines | | | | | | | | | |
| ARIMA | 5.37745 | 0.0313 | 5.5545 | 0.0433 | 5.6311 | 0.0643 | 0.3787 | 0.0220 | 4.6314 |
| GM(1,1) | 5.41213 | 0.0250 | 5.5029 | 0.0522 | 5.5483 | 0.0780 | 0.4299 | 0.0251 | 5.1783 |
| GEP-BSA | 5.4096 | 0.0255 | 5.5324 | 0.0471 | 5.6558 | 0.0602 | 0.3724 | 0.0217 | 4.4283 |

The resulting forecasts of total EEC in ASEAN-5 countries via GEP-BSA are shown in Figure 4.24. As it shown, the total EEC of ASEAN-5 countries is predicted to continually growth from 2011 up to 2030 with average annual growth rate of 3.01%. The total EEC of ASEAN-5 countries is projected to be 82.56 MTOE in 2030, which is about 1.75 times that in 2011. Moreover, it is observed that Malaysia' EEC approaches 20.25 MTOE in

2030 which is about 1.93 times that in 2011. For the Indonesia, EEC in 2030 is forecasted as 172% of that in 2011 at 24.54 MTOE. The Singapore's EEC is predicted to continually growth from 3.75 MTOE in 2011 to 5.29 in 2030. The EEC of Thailand is expected to be 24.74 MTOE in 2030, which is about 1.87 times that in 2011. The Philippines's EEC in 2030 is anticipated as 1.46 times more than that in 2011 at 7.73 MTOE.
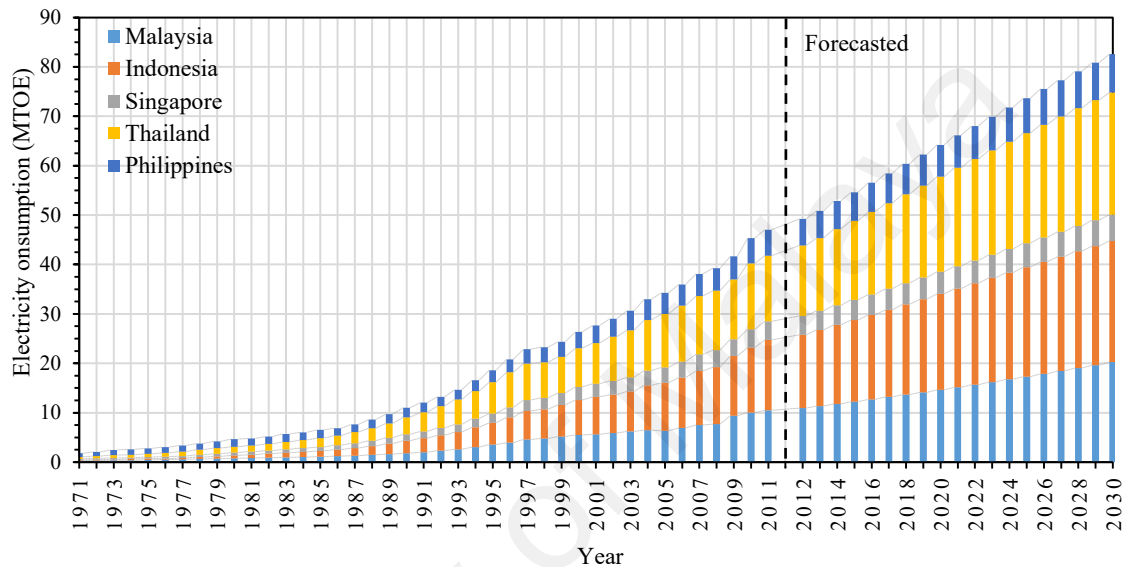


**Figure 4.24: Future projection of annual EEC of ASEAN-5 countries up to 2030 using GEP-BSA**

**CHAPTER 5: CONCLUSION**

**5.1    Conclusion**

In this study, optimized GEP models based on PSO, CSA, ACS, and BSA algorithms for long-term EEC forecasting using two different data types (i.e. SEI and EEC) are developed. The solution framework is implemented for ASEAN-5 counties (i.e. Malaysia, Indonesia, Singapore, Thailand, and Philippines), which inherently have different dependencies on EEC.

Regardless of which country is studied, the results show that the developed model by GEP-$_{BSA}$ based on EEC in preceding three years provides better-fit estimations as compared to other AI-based solution frameworks. The EEC of ASEAN-5 countries is modeled by ANN, SVR, ANFIS, M5-R, GEP, and predefined mathematical expressions (i.e. linear, quadratic and exponential models) optimized by PSO, CSA, ACS, and BSA to compare the performances of the proposed models with those obtained from other AI-based models on the same case studies.

The total EEC of ASEAN-5 countries is predicted to continually growth from 2011 up to 2030 with average growth rate of 3.01% per year. It is projected to be 82.56 MTOE in 2030, which is about 1.75 times that in 2011. The forecasting results release that EEC of ASEAN-5 countries has increasing trend, which is forecasted to continue in future, unless the demand-side management programs are properly designed and implemented to optimize the electricity consumption.

Furthermore, it is demonstrated that recent enhancements in AI-based approaches, as in GEP-$_{BSA}$, could result higher accuracy with the least complexity for long-term EEC forecasting. While in term of MAPE index the obtained results by GEP-$_{BSA}$ on the basis of socio-economic indicators are 27.33%, 35.74%, 9.38%, 61.3%, 69.97%, and based on EEC in preceding years are 4.82%, 55.83%, 23.28%, 33.52%, 47.27% higher accurate

than those obtained by GEP for Malaysia, Indonesia, Singapore, Thailand and Philippines, respectively. The proposed method is superior to other AI models in giving an optimized explicit equation, which clearly shows the relationship between input historical data and EEC in different countries without prior knowledge about the nature of the relationships between historical data and EEC.

It can be concluded that the mathematical models derived from optimized GEP can be used satisfactorily for long-term energy consumption forecasting by anyone not necessarily being expert in the field of soft computing in a spreadsheet on a personal computer even on a hand-held calculator. Thus, it can be a potential tool for policy makers and scholars to develop energy strategies plainly.

## 5.2    Future Works

The following tasks can be carried out as the future works.

- The proposed methodology can be applied in electricity market environment for mid-term and short-term electricity demand forecasting, electricity price forecasting and spinning reserve requirement forecasting.

- The proposed methodology can be utilized for wind and solar power forecasting.

- The optimized GEP approach can be applied for wind turbine power curve monitoring.

- The proposed methodology can be used to develop prediction models for energy consumption and carbon dioxide ($CO_2$) emissions in different countries.

- The proposed methodology can also be used to provide future estimations of socio-economic indicators in different countries.

# REFERENCES

Ahmed, M. S., Mohamed, A., Khatib, T., Shareef, H., Homod, R. Z., & Ali, J. A. (2017). Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm. Energy and Buildings, 138, 215-227.

Ahmed, T., Mekhilef, S., Shah, R., Mithulananthan, N., Seyedmahmoudian, M., & Horan, B. (2017). ASEAN power grid: A secure transmission infrastructure for clean and sustainable energy for South-East Asia. Renewable and Sustainable Energy Reviews, 67, 1420-1435. doi: https://doi.org/10.1016/j.rser.2016.09.055

Alavi, A. H., Aminian, P., Gandomi, A. H., & Esmaeili, M. A. (2011). Genetic-based modeling of uplift capacity of suction caissons. Expert Systems with Applications, 38(10), 12608-12618.

Alizadeh, M., Jolai, F., Aminnayeri, M., & Rada, R. (2012). Comparison of different input selection algorithms in neuro-fuzzy modeling. Expert Systems with Applications, 39(1), 1536-1544.

AlRashidi, M. R., & El-Naggar, K. M. (2010). Long term electric load forecasting based on particle swarm optimization. Applied Energy, 87(1), 320-326. doi: http://dx.doi.org/10.1016/j.apenergy.2009.04.024

Amjadi, M., Nezamabadi-Pour, H., & Farsangi, M. (2010). Estimation of electricity demand of Iran using two heuristic algorithms. Energy Conversion and Management, 51(3), 493-497.

analysis, M. G. I. (2013). International Monetary Fund (IMF). World Economic Outlook (WEO).

Ardakani, F., & Ardehali, M. (2014). Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. Energy, 65, 452-461.

ASEAN Community in Figures (ACIF), 2013 (6th ed.), Jakarta: ASEANRetrieved 9 May 2015).

Askarzadeh, A. (2014). Comparison of particle swarm optimization and other metaheuristics on electricity demand estimation: A case study of Iran. Energy, 72(0), 484-491. doi: http://dx.doi.org/10.1016/j.energy.2014.05.070

Assareh, E., Behrang, M. A., Assari, M. R., & Ghanbarzadeh, A. (2010). Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran. Energy, 35(12), 5223-5229. doi: http://dx.doi.org/10.1016/j.energy.2010.07.043

Azadeh, A., Ghaderi, S., Tarverdian, S., & Saberi, M. (2007). Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. Applied Mathematics and Computation, 186(2), 1731-1741.

Azadeh, A., Ghaderi, S. F., & Sohrabkhani, S. (2008). A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. Energy Policy, 36(7), 2637-2644. doi: http://dx.doi.org/10.1016/j.enpol.2008.02.035

Azadeh, A., Saberi, M., Ghaderi, S. F., Gitiforouz, A., & Ebrahimipour, V. (2008). Improved estimation of electricity demand function by integration of fuzzy system and data mining approach. Energy Conversion and Management, 49(8), 2165-2177. doi: http://dx.doi.org/10.1016/j.enconman.2008.02.021

Azadeh, A., Saberi, M., Gitiforouz, A., & Saberi, Z. (2009). A hybrid simulation-adaptive network based fuzzy inference system for improvement of electricity consumption estimation. Expert Systems with Applications, 36(8), 11108-11117. doi: http://dx.doi.org/10.1016/j.eswa.2009.02.081

Azadeh, A., Saberi, M., & Seraj, O. (2010). An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: A case study of Iran. Energy, 35(6), 2351-2366. doi: http://dx.doi.org/10.1016/j.energy.2009.12.023

Behrang, M. A., Assareh, E., Ghalambaz, M., Assari, M. R., & Noghrehabadi, A. R. (2011). Forecasting future oil demand in Iran using GSA (Gravitational Search Algorithm). Energy, 36(9), 5649-5654. doi: http://dx.doi.org/10.1016/j.energy.2011.07.002

Bian, X.-Q., Han, B., Du, Z.-M., Jaubert, J.-N., & Li, M.-J. (2016). Integrating support vector regression with genetic algorithm for $CO_2$-oil minimum miscibility pressure (MMP) in pure and impure $CO_2$ streams. Fuel, 182, 550-557. doi: https://doi.org/10.1016/j.fuel.2016.05.124

Canyurt, O. E., & Ozturk, H. K. (2008). Application of genetic algorithm (GA) technique on demand estimation of fossil fuels in Turkey. Energy Policy, 36(7), 2562-2569. doi: http://dx.doi.org/10.1016/j.enpol.2008.03.010

Ceylan, H., Ceylan, H., Haldenbilen, S., & Baskan, O. (2008). Transport energy modeling with meta-heuristic harmony search algorithm, an application to Turkey. Energy Policy, 36(7), 2527-2535. doi: http://dx.doi.org/10.1016/j.enpol.2008.03.019

Chang, E. H., Zhu, G. N., & Chen, J. W. (2015). A Combined Model Based on Cuckoo Search Algorithm for Electrical Load Forecasting. Paper presented at the Applied Mechanics and Materials.

Civicioglu, P. (2013a). Artificial cooperative search algorithm for numerical optimization problems. Information Sciences, 229(0), 58-76. doi: http://dx.doi.org/10.1016/j.ins.2012.11.013

Civicioglu, P. (2013b). Backtracking Search Optimization Algorithm for numerical optimization problems. Applied Mathematics and Computation, 219(15), 8121-8144. doi: https://doi.org/10.1016/j.amc.2013.02.017

Coello, C. C. (2006). Evolutionary multi-objective optimization: a historical view of the field. IEEE computational intelligence magazine, 1(1), 28-36.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE transactions on evolutionary computation, 6(2), 182-197.

Department of Energy. (2013). Key Energy Statistic Republic of the Philippines.
Duran Toksarı, M. (2007). Ant colony optimization approach to estimate energy demand of Turkey. Energy Policy, 35(8), 3984-3990. doi: http://dx.doi.org/10.1016/j.enpol.2007.01.028

Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. Energy, 35(2), 512-517. doi: http://dx.doi.org/10.1016/j.energy.2009.10.018

Elias, C. N., & Hatziargyriou, N. D. (2009). An Annual Midterm Energy Forecasting Model Using Fuzzy Logic. Power Systems, IEEE Transactions on, 24(1), 469-478. doi: 10.1109/TPWRS.2008.2009490

Energy Policy and Planning Office (EPPO). (2015). Electricity Generating Authority of Thailand (EGAT), Metropolitan Electricity Authority (MEA), Provincial Electricity Authority (PEA), . Ministry of Energy (Thailand).

Gandomi, A. H., Yang, X.-S., & Alavi, A. H. (2013). Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems. Engineering with computers, 29(1), 17-35.

Ghelardoni, L., Ghio, A., & Anguita, D. (2013). Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression. Smart Grid, IEEE Transactions on, 4(1), 549-556. doi: 10.1109/TSG.2012.2235089

Golbraikh, A., & Tropsha, A. (2002). Beware of q 2! Journal of molecular graphics and modelling, 20(4), 269-276.

Hamzacebi, C., & Es, H. A. (2014). Forecasting the annual electricity consumption of Turkey using an optimized grey model. Energy, 70(0), 165-171. doi: http://dx.doi.org/10.1016/j.energy.2014.03.105

Hernandez, L., Baladron, C., Aguiar, J. M., Carro, B., Sanchez-Esguevillas, A. J., Lloret, J., & Massana, J. (2014). a survey on electric power demand forecasting: Future trends in smart grids, microgrids and smart buildings. Communications Surveys & Tutorials, IEEE, 16(3), 1460-1495.

Holland, J., & Goldberg, D. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. ed: Addison-Wesley, Reading, MA.

Hong, W.-C. (2009). Electric load forecasting by support vector model. Applied Mathematical Modelling, 33(5), 2444-2454. doi: http://dx.doi.org/10.1016/j.apm.2008.07.010

Hong, W.-C. (2010). Application of chaotic ant swarm optimization in electric load forecasting. Energy Policy, 38(10), 5830-5839. doi: https://doi.org/10.1016/j.enpol.2010.05.033

Hsu, C. C., & Chen, C. Y. (2003). A modified Grey forecasting model for long-term prediction. Journal of the Chinese Institute of Engineers, 26(3), 301-308.

IEA, I. E. A. (2015). Development Prospects of the ASEAN Power Sector: IEA.

Ismail, A. M., Ramirez-Iniguez, R., Asif, M., Munir, A. B., & Muhammad-Sukki, F. (2015). Progress of solar photovoltaic in ASEAN countries: A review. Renewable and Sustainable Energy Reviews, 48, 399-412. doi: https://doi.org/10.1016/j.rser.2015.04.010

Jang, J.-S. (1996). Input selection for ANFIS learning. Paper presented at the Fuzzy Systems, 1996., Proceedings of the Fifth IEEE International Conference on.

Jang, J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics, 23(3), 665-685. doi: 10.1109/21.256541

Kaboli, S. H. A., Selvaraj, J., & Rahim, N. A. (2016). Long-term electric energy consumption forecasting via artificial cooperative search algorithm. Energy, 115, Part 1, 857-871. doi: http://dx.doi.org/10.1016/j.energy.2016.09.015

Kandil, M., El-Debeiky, S. M., & Hasanien, N. (2002). Long-term load forecasting for fast developing utility using a knowledge-based expert system. Power Systems, IEEE Transactions on, 17(2), 491-496.

Kankal, M., Akpınar, A., Kömürcü, M. İ., & Özşahin, T. Ş. (2011). Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. Applied Energy, 88(5), 1927-1939. doi: http://dx.doi.org/10.1016/j.apenergy.2010.12.005

Karabulut, K., Alkan, A., & Yilmaz, A. S. (2008). Long term energy consumption forecasting using genetic programming. Mathematical and Computational Applications, 13(2), 71.

Kaveh, A., Shamsapour, N., Sheikholeslami, R., & Mashhadian, M. (2012). Forecasting transport energy demand in Iran using meta-heuristic algorithms. Int. J. Optim. Civil Eng, 2(4), 533-544.

Kialashaki, A., & Reisel, J. R. (2014). Development and validation of artificial neural network models of the energy demand in the industrial sector of the United States. Energy, 76, 749-760. doi: http://dx.doi.org/10.1016/j.energy.2014.08.072

Kıran, M. S., Özceylan, E., Gündüz, M., & Paksoy, T. (2012a). A novel hybrid approach based on Particle Swarm Optimization and Ant Colony Algorithm to forecast energy demand of Turkey. Energy Conversion and Management, 53(1), 75-83. doi: http://dx.doi.org/10.1016/j.enconman.2011.08.004

Kıran, M. S., Özceylan, E., Gündüz, M., & Paksoy, T. (2012b). Swarm intelligence approaches to estimate electricity energy demand in Turkey. Knowledge-Based Systems, 36(0), 93-103. doi: http://dx.doi.org/10.1016/j.knosys.2012.06.009

KÜÇÜKDENİZ, T. (2010). Long Term Electricity Demand Forcesting: An Alternative Approach With Support Vector Machines. İÜ Mühendislik Bilimleri Dergisi, 1(1), 45-54.

Landeras, G., López, J. J., Kisi, O., & Shiri, J. (2012). Comparison of Gene Expression Programming with neuro-fuzzy and neural network computing techniques in estimating daily incoming solar radiation in the Basque Country (Northern Spain). Energy Conversion and Management, 62, 1-13. doi: https://doi.org/10.1016/j.enconman.2012.03.025

Li, H.-z., Guo, S., Li, C.-j., & Sun, J.-q. (2013). A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. Knowledge-Based Systems, 37, 378-387. doi: http://dx.doi.org/10.1016/j.knosys.2012.08.015

Li, H., Guo, S., Zhao, H., Su, C., & Wang, B. (2012). Annual electric load forecasting by a least squares support vector machine with a fruit fly optimization algorithm. Energies, 5(11), 4430-4445.

Liu, X., & Fang, J. a. (2013). Long-term load forecasting based on a time-variant ratio multiobjective optimization fuzzy time series model. Mathematical Problems in Engineering, 2013.

Malaysia Energy Information Hub Unit. (2015). Malaysia energy statistics handbook. Energy commission.

McClymont, K., & Keedwell, E. (2012). Deductive sort and climbing sort: New methods for non-dominated sorting. Evolutionary computation, 20(1), 1-26.

Ministry of Energy and Mineral Resources, I. (2015). Handbook of Energy & Economic Statistics of Indonesia

Modiri-Delshad, M., Aghay Kaboli, S. H., Taslimi-Renani, E., & Rahim, N. A. (2016). Backtracking search algorithm for solving economic dispatch problems with valve-point effects and multiple fuel options. Energy, 116, Part 1, 637-649. doi: http://dx.doi.org/10.1016/j.energy.2016.09.140

Modiri-Delshad, M., & Rahim, N. A. (2014). Solving non-convex economic dispatch problem via backtracking search algorithm. Energy, 77, 372-381.

Modiri-Delshad, M., & Rahim, N. A. (2016). Multi-objective backtracking search algorithm for economic emission dispatch problem. Applied Soft Computing, 40, 479-494.

Mostafavi, E. S., Mostafavi, S. I., Jaafari, A., & Hosseinpour, F. (2013). A novel machine learning approach for estimation of electricity demand: An empirical evidence from Thailand. Energy Conversion and Management, 74, 548-555.

Mostafavi, E. S., Mousavi, S. M., & Hosseinpour, F. (2014). Gene Expression Programming as a Basis for New Generation of Electricity Demand Prediction Models. Computers & Industrial Engineering.

Mousavi, S. M., Mostafavi, E. S., & Hosseinpour, F. (2014). Gene expression programming as a basis for new generation of electricity demand prediction models. Computers & Industrial Engineering, 74, 120-128.

Nadimi, V., Azadeh, A., Pazhoheshfar, P., & Saberi, M. (2010, 17-19 Nov. 2010). An Adaptive-Network-Based Fuzzy Inference System for Long-Term Electric Consumption Forecasting (2008-2015): A Case Study of the Group of Seven (G7) Industrialized Nations: U.S.A., Canada, Germany, United Kingdom, Japan, France and Italy. Paper presented at the Computer Modeling and Simulation (EMS), 2010 Fourth UKSim European Symposium on.

Nasab, E. H., Khezri, M., Khodamoradi, M. S., & Gargari, E. A. (2010). An application of imperialist competitive algorithm to simulation of energy demand based on economic indicators: evidence from Iran. European Journal of Scientific ResearchVol, 43, 495-506.

Nguyen, T. T., Vo, D. N., & Truong, A. V. (2014). Cuckoo search algorithm for short-term hydrothermal scheduling. Applied Energy, 132, 276-287. doi: http://dx.doi.org/10.1016/j.apenergy.2014.07.017

Oltean, M., & Grosan, C. (2003). A comparison of several linear genetic programming techniques. Complex Systems, 14(4), 285-314.

Ozcelik, Y., & Hepbasli, A. (2006). Estimating Petroleum Exergy Production and Consumption Using a Simulated Annealing Approach. Energy Sources, Part B: Economics, Planning, and Policy, 1(3), 255-265. doi: 10.1080/00908310600718809

Rafieerad, A. R., Bushroa, A. R., Nasiri-Tabrizi, B., Fallahpour, A., Vadivelu, J., Musa, S. N., & Kaboli, S. H. A. (2016). GEP-based method to formulate adhesion strength and hardness of Nb PVD coated on Ti–6Al–7Nb aimed at developing

mixed oxide nanotubular arrays. Journal of the Mechanical Behavior of Biomedical Materials, 61, 182-196. doi: http://dx.doi.org/10.1016/j.jmbbm.2016.01.028

Rafieerad, A. R., Bushroa, A. R., Nasiri-Tabrizi, B., Kaboli, S. H. A., Khanahmadi, S., Amiri, A., . . . Wasa, K. (2017). Toward improved mechanical, tribological, corrosion and in-vitro bioactivity properties of mixed oxide nanotubes on Ti–6Al–7Nb implant using multi-objective PSO. Journal of the Mechanical Behavior of Biomedical Materials, 69, 1-18. doi: http://dx.doi.org/10.1016/j.jmbbm.2016.11.019

Raza, M. Q., & Khosravi, A. (2015). A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renewable and Sustainable Energy Reviews, 50, 1352-1372. doi: https://doi.org/10.1016/j.rser.2015.04.065

Renani, E. T., Elias, M. F. M., & Rahim, N. A. (2016). Using data-driven approach for wind power prediction: A comparative study. Energy Conversion and Management, 118, 193-203.

Roy, P. P., & Roy, K. (2008). On some aspects of variable selection for partial least squares regression models. QSAR & Combinatorial Science, 27(3), 302-313.

Roy, S., Ghosh, A., Das, A. K., & Banerjee, R. (2015). Development and validation of a GEP model to predict the performance and exhaust emission parameters of a CRDI assisted single cylinder diesel engine coupled with EGR. Applied Energy, 140, 52-64. doi: http://dx.doi.org/10.1016/j.apenergy.2014.11.065

Sebtahmadi, S. S., Azad, H. B., Kaboli, S. H. A., Islam, M. D., & Mekhilef, S. (2017). A PSO-DQ Current Control Scheme for Performance Enhancement of Z-source Matrix Converter to Drive IM Fed by Abnormal Voltage. IEEE Transactions on Power Electronics.

Sözen, A., Arcaklioğlu, E., & Özkaymak, M. (2005). Turkey's net energy consumption. Applied Energy, 81(2), 209-221. doi: http://dx.doi.org/10.1016/j.apenergy.2004.07.001

Srinivas, N., & Deb, K. (1994). Muiltiobjective optimization using nondominated sorting in genetic algorithms. Evolutionary computation, 2(3), 221-248.

Svalina, I., Galzina, V., Lujić, R., & Šimunović, G. (2013). An adaptive network-based fuzzy inference system (ANFIS) for the forecasting: The case of close price

indices. Expert Systems with Applications, 40(15), 6055-6063. doi: https://doi.org/10.1016/j.eswa.2013.05.029

Tavana, M., Fallahpour, A., Di Caprio, D., & Santos-Arteaga, F. J. (2016). A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection. Expert Systems with Applications, 61, 129-144. doi: https://doi.org/10.1016/j.eswa.2016.05.027

The Energy Market Authority (EMA). (2016). Singapore Energy Statistics (SES). Ministry of Trade and Industry (MTI).

Ünler, A. (2008). Improvement of energy demand forecasts using swarm intelligence: The case of Turkey with projections to 2025. Energy Policy, 36(6), 1937-1944. doi: http://dx.doi.org/10.1016/j.enpol.2008.02.018

Unler, A., & Murat, A. (2010). A discrete particle swarm optimization method for feature selection in binary classification problems. European Journal of Operational Research, 206(3), 528-539.

Uzlu, E., Kankal, M., Akpınar, A., & Dede, T. (2014). Estimates of energy consumption in Turkey using neural networks with the teaching–learning-based optimization algorithm. Energy, 75(0), 295-303. doi: http://dx.doi.org/10.1016/j.energy.2014.07.078

Wang, J., Li, L., Niu, D., & Tan, Z. (2012). An annual load forecasting model based on support vector regression with differential evolution algorithm. Applied Energy, 94, 65-70.

Wang, Y., Wang, J., Zhao, G., & Dong, Y. (2012). Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: a case study of China. Energy Policy, 48, 284-294.

Whigham, P. A., & Dick, G. (2010). Implicitly Controlling Bloat in Genetic Programming. Evolutionary Computation, IEEE Transactions on, 14(2), 173-190. doi: 10.1109/TEVC.2009.2027314

Wu, G., Kechavarzi, C., Li, X., Wu, S., Pollard, S. J. T., Sui, H., & Coulon, F. (2013). Machine learning models for predicting PAHs bioavailability in compost amended soils. Chemical Engineering Journal, 223, 747-754. doi: http://dx.doi.org/10.1016/j.cej.2013.02.122

Xue, B., Zhang, M., & Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. IEEE transactions on cybernetics, 43(6), 1656-1671.

Xue, B., Zhang, M., & Browne, W. N. (2014). Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. Applied Soft Computing, 18, 261-276.

Yu, S., Zhu, K., & Zhang, X. (2012). Energy demand projection of China using a path-coefficient analysis and PSO–GA approach. Energy Conversion and Management, 53(1), 142-153. doi: http://dx.doi.org/10.1016/j.enconman.2011.08.015

Yuan, C., Liu, S., & Fang, Z. (2016). Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. Energy, 100, 384-390. doi: https://doi.org/10.1016/j.energy.2016.02.001

Zahedi, G., Azizi, S., Bahadori, A., Elkamel, A., & Wan Alwi, S. R. (2013). Electricity demand estimation using an adaptive neuro-fuzzy network: A case study from the Ontario province – Canada. Energy, 49, 323-328. doi: http://dx.doi.org/10.1016/j.energy.2012.10.019

Zhang, C., Zhou, J., Li, C., Fu, W., & Peng, T. (2017). A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting. Energy Conversion and Management, 143, 360-376.

Zhang, X., Tian, Y., Cheng, R., & Jin, Y. (2015). An efficient approach to nondominated sorting for evolutionary multiobjective optimization. IEEE Transactions on Evolutionary Computation, 19(2), 201-213.

Zhao, H., & Guo, S. (2016). An optimized grey model for annual power load forecasting. Energy, 107, 272-286. doi: https://doi.org/10.1016/j.energy.2016.04.009

Zhong, J., Ong, Y.-S., & Cai, W. (2015). Self-Learning Gene Expression Programming. Evolutionary Computation, IEEE Transactions on, PP(99), 1-1. doi: 10.1109/TEVC.2015.2424410

Zitzler, E., & Thiele, L. (1999). Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE transactions on Evolutionary Computation, 3(4), 257-271.

Zou, F., Chen, D., Li, S., Lu, R., & Lin, M. (2017). Community Detection in Complex Networks: Multi-objective Discrete Backtracking Search Optimization Algorithm with Decomposition. Applied Soft Computing.

# LIST OF PUBLICATIONS AND PAPERS PRESENTED

## JOURNAL PAPERS

S. Hr. Aghay Kaboli, J. Selvaraj, and N. A. Rahim. (2016). Long-term electric energy consumption forecasting via artificial cooperative search algorithm. *Energy,* 115, 857-871.

S. Hr. Aghay Kaboli, A. Fallahpoura, J. Selvaraja and N. A. Rahim. (2017). Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. *Energy*, 126, 144-164.

S. Hr. Aghay Kaboli, A. Fallahpour, N. Kazemi, J. Selvaraj and N. A. Rahim. (2016). An expression-driven approach for long-term electric power consumption forecasting. *American Journal of Data Mining and Knowledge Discovery,* 1, 16-28.

M. Modiri-Delshad, S. Hr. Aghay Kaboli, E. Taslimi-Renani, and N. A. Rahim. (2016). Backtracking search algorithm for solving economic dispatch problems with valve-point effects and multiple fuel options. *Energy*, 116, 637-649.

## CONFERENCE PAPERS

S. Hr. Aghay Kaboli, A. Fallahpour, N. Kazemi, J. Selvaraj and N. A. Rahim. (2016). Electric energy consumption forecasting via expression-driven approach. *Clean Energy and Technology (CEAT), 4<sup>th</sup> IET Conference*.