

## PEMODELAN KETIBAAN PELANCONG ANTARABANGSA KE MALAYSIA

(Modelling of International Tourist Arrivals to Malaysia)

SHARIFAH HUDA SYED IBRAHIM & HUMAIDA BANU SAMSUDIN\*

### ABSTRAK

Peramalan jumlah ketibaan pelancong adalah penting bagi memberi gambaran prestasi sektor pelancongan yang tepat bagi tahun-tahun akan datang. Ramalan tersebut dapat membantu pihak berkepentingan membuat keputusan strategik seperti pelaburan, pembangunan infrastruktur dan sebagainya. Langkah strategik ini mampu memberi impak negatif atau positif bergantung kepada ketepatan nilai ramalan yang diperolehi. Kajian ini bertujuan untuk memilih model terbaik untuk peramalan jumlah ketibaan pelancong antarabangsa ke Malaysia menggunakan data bulanan dari Januari 2000 hingga Disember 2019. Dua kaedah digunakan iaitu Autoregresi Purata Bergerak Bersepadu Bermusim (SARIMA) dan Analisis Spektrum Tunggal (SSA). Ketepatan ramalan bagi data dalam dan luar sampel diuji dengan nilai RMSE, MAE dan MAPE. Model terbaik dipilih berdasarkan prestasi ramalan iaitu model yang mempunyai nilai paling kecil bagi ketiga-tiga pengukur ralat yang digunakan. Kajian menunjukkan data siri masa yang di analisis mempunyai trend linear dan komponen bermusim. Selain itu, kajian juga menunjukkan model ARIMA  $(3,1,3)(2,0,0)[12]$  menghasilkan nilai RMSE, MAE, dan MAPE yang paling kecil bagi peramalan data luar sampel. Model ARIMA  $(3,1,3)(2,0,0)[12]$  dan SSA  $(24,14)$  mempunyai nilai MAPE lebih kecil daripada 10%, maka kedua-dua model ini adalah dalam kategori ramalan yang sangat tepat dan boleh digunakan bagi peramalan.

*Kata kunci:* peramalan; pelancongan; SARIMA; SSA

### ABSTRACT

The forecast of tourist arrivals is important in showing an accurate description of the tourism industry's performance in the future. The forecast can help all stakeholders to make any strategic decisions such as investments, infrastructure development and others. These strategic moves can either give a positive or negative outcome depending on how reliable and accurate the forecast values are. Therefore, this paper aims to choose the best model for forecasting by modelling the international tourist arrivals to Malaysia using monthly data set from January 2000 to December 2019. Two methods are used, Seasonal Autoregressive Integrated Moving Average (SARIMA) and Singular Spectrum Analysis (SSA). The forecasting accuracy for in sample and out sample are tested using the value of RMSE, MAE, and MAPE. The best model is chosen from the forecasting performance that produces the minimum value for all three tests. The study shows that the time series data has a linear trend and a seasonal component. Besides that, the study also shows that ARIMA  $(3,1,3)(2,0,0)[12]$  produces the smallest value for RMSE, MAE, and MAPE based on the out sample forecasting. Both ARIMA  $(3,1,3)(2,0,0)[12]$  and SSA  $(24,14)$  model have MAPE value of less than 10%, thus both model have a very accurate forecasting results and can be used for forecasting.

*Keywords:* forecast; tourism; SARIMA; SSA

## **1. Pengenalan**

Industri pelancongan memainkan peranan yang sangat penting sebagai sumber pendapatan dan menyediakan pekerjaan kepada negara. Pihak kerajaan telah menyediakan pelbagai strategi untuk meningkatkan jumlah ketibaan pelancong antarabangsa. Pada tahun 2019, industri pelancongan telah menyumbang sekitar 2.9 trilion dollar Amerika Syarikat (AS) kepada KDNK global dan industri pelancongan di Amerika Syarikat pula menyumbang jumlah KDNK tertinggi global iaitu sebanyak 580.7 billion dolar AS (Statista 2020).

Berdasarkan laporan yang dikeluarkan oleh UNWTO (2011) iaitu *Tourism Towards 2030*, ketibaan pelancong antarabangsa di peringkat global dijangka akan meningkat dengan kadar purata 43 juta orang pada setiap tahun dan jumlah keseluruhan akan mencecah 1.8 bilion orang menjelang tahun 2030. Rantau Asia dan Pasifik juga diramal akan memperoleh majoriti daripada ketibaan pelancong baru manakala Asia Selatan akan menjadi wilayah dengan pertumbuhan paling cepat iaitu sebanyak 6% pada setiap tahun.

Permintaan ketibaan pelancong antarabangsa di Malaysia pada beberapa tahun kebelakangan ini telah mengalami pertumbuhan secara negatif dari tahun 2017 ke tahun 2018 (*Tourism Malaysia 2020*). Walaupun begitu, Malaysia tetap mengalami pertumbuhan yang stabil dari aspek jumlah perbelanjaan pelancongan yang meningkat sebanyak 2.4% untuk tahun 2018 dan 2019 yang berjumlah kepada 84.1 juta dan 86.1 juta (*Tourism Malaysia 2019*).

Berdasarkan Pelan Promosi Bersepadu 2018-2020, terdapat beberapa cabaran yang pernah dihadapi dalam sektor pelancongan Malaysia. Satu daripadanya ialah keputusan untuk merasionalisasi syarikat penerbangan tempatan, MAS. Sebelum keputusan ini dilakukan, terdapat 58 penerbangan terus mingguan dari Eropah ke Malaysia pada suku tahun pertama 2015 dengan jumlah kapasiti tempat duduk sebanyak 21,066. Namun selepas rasionalisasi dilakukan, jumlah penerbangan terus menurun kepada 38 penerbangan sahaja dan jumlah kapasiti tempat duduk dicatatkan hanya pada 13,115 pada suku tahun pertama 2017.

Selain itu, Malaysia turut mengalami persaingan yang sengit dari negara-negara jiran seperti Thailand, Filipina, dan Indonesia yang memberi kesan langsung kepada jumlah kedatangan pelancong ke Malaysia. Antara cabaran lain termasuk prestasi Ringgit yang lemah dan juga inflasi. Kerajaan berjaya mengambil langkah proaktif bagi menangani cabaran tersebut dengan strategi-strategi yang berimpak seperti melonggarkan peraturan visa, kolaborasi dengan sektor swasta serta memberi lebih penekanan kepada pemasaran digital.

Chetty (2011) menyatakan bawah sebilangan besar produk pelancongan juga adalah mudah rosak. Ini bermaksud produk yang tidak terjual tidak boleh disimpan di dalam inventori, maka ia memberi premium dalam pembentukan permintaan dalam jangka masa pendek dan juga panjang. Sebagai contoh, untuk setiap tiket penerbangan atau tiket masuk taman tema yang tidak terjual, maka terlepas sebuah peluang pendapatan yang sepatutnya boleh didapati. Beberapa strategi telah muncul untuk membantu menyelesaikan masalah ini dan ia termasuk menyelidik dan meramal permintaan ketibaan pelancong pada masa akan datang untuk memastikan perancang dan pembuat dasar dapat merumus dasar industri pelancongan dengan lebih baik.

Oleh itu, pihak kerajaan dan juga syarikat-syarikat terkemuka memerlukan ramalan ketibaan pelancong yang tepat apabila mereka ingin merangka dan merancang pertumbuhan industri seperti pembangunan infrastruktur, pengangkutan, penginapan, dan keperluan lain. Dengan itu, kajian ini bertujuan untuk menganalisis trend jumlah ketibaan pelancong antarabangsa ke Malaysia dari tahun 2000 hingga tahun 2019. Selain itu, model yang terbaik untuk peramalan dicari dengan membandingkan antara dua buah model yang digunakan iaitu

Autoregresi Purata Bergerak Bersepadu Bermusim (SARIMA) dan Analisis Spektrum Tunggal (SSA).

Terdapat tiga objektif berbeza bagi kajian ini. Pertama, kajian ini menganalisis trend jangka masa panjang bagi ketibaan pelancong ke Malaysia dari tahun 2000 hingga 2019. Kedua, membina model untuk jumlah ketibaan pelancong ke Malaysia menggunakan kaedah SARIMA dan SSA. Akhir sekali, membanding ketepatan prestasi model SARIMA dan SSA bagi data dalam sampel dan data luar sampel dan memilih model yang terbaik untuk peramalan.

## 2. Bahan dan Kaedah

### 2.1. Data kajian

Data yang dicerap adalah data sekunder jumlah ketibaan pelancong antarabangsa ke Malaysia yang diukur secara bulanan. Sumber data adalah daripada MYTourismData sejumlah 240 sampel dari Januari 2000 hingga Disember 2019. Data ini dibahagikan kepada 2 bahagian utama iaitu 228 sampel pertama dari Januari 2000 hingga Disember 2018 dikenali sebagai data dalam sampel yang digunakan untuk penganggaran parameter model dan juga pemilihan model terbaik bagi SARIMA dan SSA. 12 sampel yang seterusnya dikenali sebagai data luar sampel dan digunakan untuk menguji prestasi model yang telah dipilih.

### 2.2. Tatakaedah kajian

Dalam kajian ini, data dikaji menggunakan perisian R untuk menganalisis trend serta beberapa statistik perihalan seperti nilai minimum, nilai maksimum dan nilai purata untuk memenuhi keperluan objektif pertama. Kemudian, model SARIMA dan SSA akan digunakan untuk memodelkan jumlah ketibaan pelancong antarabangsa ke Malaysia dan model terbaik bagi kedua-dua kaedah akan dipilih. Prestasi model yang dipilih akan diuji dengan melakukan peramalan data luar sampel jumlah ketibaan pelancong antarabangsa ke Malaysia bagi tahun 2019 dan ketepatan ramalan diukur menggunakan nilai RMSE, MAE, dan MAPE.

Model SARIMA adalah berdasarkan kaedah Box-Jenkins dan penerangan bagi tatakaedah ini adalah merujuk kepada Choden & Unhapiat (2018) dan Norbaizura & Zainudin (2014). Kaedah ini mempunyai empat langkah utama iaitu pengecaman model, penganggaran parameter, penyemakan diagnostik, dan pemilihan model. Sebelum proses pemodelan dilakukan, ujian kepegunan dilakukan menggunakan Ujian *Dickey-Fuller* bagi mengenal pasti data siri masa yang digunakan memenuhi andaian kepegunan atau tidak.

SARIMA merangkumi komponen bermusim dan tidak bermusim iaitu ARIMA  $(p, d, q) (P, D, Q) [s]$ . Komponen  $p$  dan  $P$  merupakan komponen tidak bermusim dan bermusim bagi AR,  $d$  dan  $D$  adalah komponen tidak bermusim dan bermusim bagi faktor pembeza,  $q$  dan  $Q$  merupakan faktor tidak bermusim dan bermusim bagi MA, manakala  $s$  merupakan tempoh musim. Persamaan (1) adalah persamaan umum bagi model ini iaitu:

$$\Phi_P(B^s)\phi_p(B)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B)^s a_t \quad (1)$$

dengan,

$$\begin{aligned} \phi_p(B) &= \text{Komponen autoregresi tidak bermusim bagi peringkat } p \\ \Phi_P(B^s) &= \text{Komponen autoregresi bermusim bagi peringkat } P \\ X_t &= \text{Nilai semasa bagi siri masa yang diuji} \end{aligned}$$

|                 |   |
|-----------------|---|
| $B$             | = Anjakan ke belakang pengoperasi $X_t B_i = X_{t-i}$ |
| $(1 - B)^d$     | = Komponen pembeza tidak bermusim                     |
| $(1 - B^s)^D$   | = Komponen pembeza bermusim                           |
| $\theta_q(B)$   | = Purata bergerak tidak bermusim bagi peringkat $q$   |
| $\theta_Q(B)^s$ | = Purata bergerak bermusim bagi peringkat $Q$         |
| $\alpha_t$      | = Proses hingar putih                                 |

Kaedah SSA merangkumi dua peringkat utama iaitu penguraian dan pembinaan semula. Bagi peringkat penguraian, terdapat dua langkah iaitu pembenaman dan penguraian nilai tunggal manakala bagi peringkat pembinaan semula, terdapat dua langkah juga iaitu pengelompokan dan pemurataan pepenjuru. Kaedah ini melibatkan dua parameter yang perlu dianggar iaitu  $L$ , panjang tingkap ( $2 \leq L \leq \frac{N}{2}$ ) dan  $r$ , bilangan nilai eigen. Akhir sekali, siri yang diperoleh daripada peringkat pembinaan semula digunakan untuk peramalan. Penerangan bagi tatakaedah SSA ini adalah berdasarkan rujukan utama daripada Hassani & Mahmoudvand (2018), Hassani (2007) dan Silva & Alonso (2020).

### 3. Hasil dan Perbincangan

#### 3.1. Analisis Deskriptif

Data yang digunakan mempunyai trend linear dan ia tidak turun naik pada sekitar nilai min yang sama, ini menunjukkan data adalah tidak pegun. Tanda variasi bermusim secara malar juga dapat dilihat dengan jelas berbanding variasi bermusim secara meningkat. Nilai maksimum bagi jumlah kedatangan pelancong dalam tempoh sampel adalah sebanyak 2,806,565 orang pelancong yang dicatatkan pada Disember 2013.

Jadual 1 menunjukkan beberapa statistik perihalan bagi data yang digunakan. Nilai purata jumlah kedatangan pelancong secara bulanan ke Malaysia bagi tempoh 20 tahun adalah 1,754,878 orang. Nilai kepencongan dan kurtosis turut dicari untuk memberi maklumat tentang bentuk taburan bagi siri masa yang dianalisis. Nilai kepencongan menunjukkan bahawa siri masa ini bertaburan secara tidak normal dan mempunyai taburan terpencong negatif. Analisis bagi nilai kurtosis menerangkan bahawa siri masa ini mempunyai taburan platikurtosis iaitu kebarangkalian yang lebih rendah untuk nilai ekstrim berbanding dengan taburan normal dan ia terdiri daripada nilai yang mempunyai serakan yang meluas sekitar nilai min.

Bagi mengesahkan maklumat yang telah didapati melalui nilai kepencongan dan kurtosis, ujian kenormalan telah dilakukan menggunakan ujian *Shapiro-Wilk*. Nilai  $p$  yang diperolehi adalah lebih kecil daripada 0.05 menunjukkan data ini tidak normal, selari dengan hasil nilai kepencongan yang telah didapati. Selain itu, ujian *Dickey-Fuller* turut dilakukan bagi mengesahkan kepegungan data siri masa ini. Nilai yang didapati adalah 0.3716 iaitu lebih besar daripada aras keyakinan 0.05 yang menunjukkan data siri masa ini adalah tidak pegun. Akhir sekali, ujian *Rosner* dilakukan untuk melihat kewujudan data terpenencil. Hasil yang didapati menunjukkan bahawa data terpenencil adalah tidak wujud dan semua nilai bertaburan dalam julat yang ada.

Jadual 1. Statistik perihalan dan ujian tambahan

| Siri                         | Purata  | Kepencongan | Kurtosis | Ujian<br><i>Shapiro-<br/>Wilk</i> | Ujian<br><i>Dickey-<br/>Fuller</i> | Ujian<br><i>Rosner</i> |
|------------------------------|---------|-------------|----------|-----------------------------------|------------------------------------|------------------------|
| Jumlah Ketibaan<br>Pelancong | 1754878 | -0.5789     | 2.2982   | 8.425e-09                         | 0.3716                             | PALSU                  |

### 3.2. Pemodelan SARIMA

Bagi mendapatkan model SARIMA terbaik, fungsi *auto.arima* telah digunakan dalam perisian R. Fungsi ini memilih model terbaik secara automatik menggunakan data dalam sampel dengan mencari nilai AIC dan BIC yang paling kecil. Model yang dipilih adalah model ARIMA (3,1,3)(2,0,0)[12] dan Jadual 2 pula menunjukkan prestasi model dalam sampel yang diukur dengan beberapa ujian pengukur ralat.

Jadual 2. Prestasi model ARIMA (3,1,3)(2,0,0)[12] bagi data dalam sampel

| Model                    | RMSE     | MAE      | MAPE     | AIC     | BIC     |
|--------------------------|----------|----------|----------|---------|---------|
| ARIMA (3,1,3)(2,0,0)[12] | 126674.6 | 91622.53 | 6.257035 | 6004.47 | 6035.29 |

Langkah seterusnya, penyemakan diagnostik dilakukan bagi model yang telah dipilih. Langkah ini dilakukan dengan dua cara iaitu secara pengiraan menggunakan ujian *Ljung-Box* dan secara gambaran visual dengan melihat ACF, histogram, dan plot QQ. Ujian *Ljung-Box* memberikan nilai *p* sebanyak 0.7512 iaitu lebih besar daripada aras keyakinan 0.05. Oleh itu, reja yang wujud adalah hingar putih.

Setelah model terbaik dipilih menggunakan data dalam sampel, model tersebut digunakan untuk proses peramalan bagi data luar sampel. Ramalan dilakukan untuk tempoh dari Januari 2019 hingga Disember 2019 dengan aras keyakinan 95%.

Prestasi ketepatan ramalan model ini untuk data luar sampel dapat dilihat pada Jadual 3. RMSE, MAE, dan MAPE telah dihitung untuk menganggar ralat berdasarkan nilai ramalan dan nilai sebenar.

Jadual 3. Prestasi model ARIMA (3,1,3)(2,0,0)[12] bagi data luar sampel

| Model                    | RMSE       | MAE        | MAPE    |
|--------------------------|------------|------------|---------|
| ARIMA (3,1,3)(2,0,0)[12] | 154728.393 | 133250.417 | 6.08566 |

Model ini merupakan model ARIMA bermusim dan nilai *s* yang didapati adalah 12 menunjukkan ia mempunyai faktor musim secara bulanan.

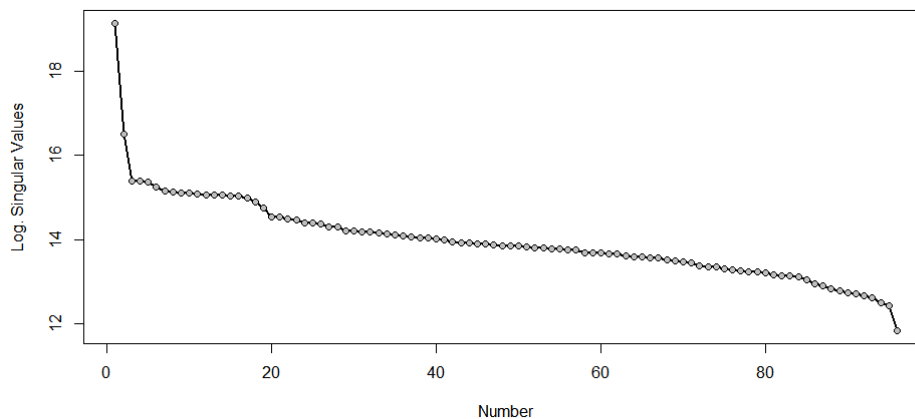
### 3.3. Pemodelan SSA

Pemodelan SSA dilakukan dengan dua cara iaitu manual dan automatik dengan kod dalam perisian R yang digunakan merujuk kepada Hassani & Mahmoudvand (2018). Dua model yang diperolehi daripada kedua-dua cara ini dibandingkan ramalannya untuk memilih satu model terbaik bagi mewakili metodologi SSA.

Bagi metodologi SSA, terdapat dua peringkat yang perlu dilakukan untuk mengasingkan komponen trend, musim dan hingar daripada siri masa yang asal iaitu penguraian dan pembinaan semula. Dalam peringkat penguraian, terdapat satu parameter yang perlu dianggar iaitu *L*, panjang tingkap. Nilai *L* disarankan perlu cukup besar tetapi tidak melebihi  $N/2$  iaitu

bagi kajian ini adalah 114. Disebabkan data ini mempunyai faktor musim secara bulanan, nilai  $L$  juga perlu berkadar dengan kala yang wujud iaitu 12 untuk memperoleh kebolehasingan yang lebih baik bagi komponen berkala. Oleh itu, nilai  $L$  yang dipilih adalah 96. Dalam peringkat penguraian, matriks trajektori  $X$  telah diurai kepada 96 eigen ganda tiga yang merangkumi 96 nilai eigen, 96 vektor eigen, dan 0 faktor vektor.

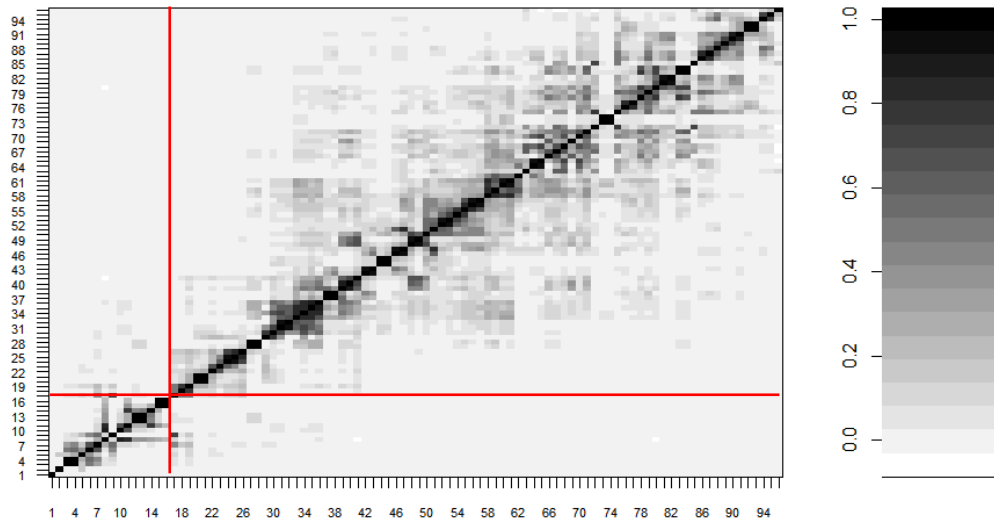
Peringkat pembinaan semula pula adalah untuk memperoleh komponen trend, musim dan hingar daripada eigen ganda tiga yang didapati dalam peringkat penguraian. Terdapat satu parameter yang perlu dipilih iaitu nilai  $r$ , bilangan nilai eigen dalam proses pengelompokan. Nilai  $r$  ini digunakan untuk menghad bilangan eigen ganda tiga yang digunakan untuk mengenal pasti komponen trend dan musim. Nilai  $r$  ditentukan berdasarkan analisis plot nilai tunggal dengan mencari bilangan eigen ganda tiga yang tidak menggambarkan hingar pada plot tersebut. Urutan nilai tunggal yang menurun secara perlahan menggambarkan komponen hingar manakala nilai tunggal bagi dua eigen ganda tiga yang harmoni adalah selalunya saling rapat antara satu sama lain (kecuali pada frekuensi 0.5 yang mempunyai satu nilai eigen ganda tiga dengan vektor tunggal yang tajam). Plot bagi nilai tunggal dengan  $L = 96$  akan menghasilkan 96 eigen ganda tiga yang disusun mengikut magnitud sumbangan nilai tersebut ketika proses penguraian. Rajah 1 menunjukkan plot bagi logaritma 96 nilai-nilai tunggal.



Rajah 1. Plot nilai tunggal bagi  $L=96$

Daripada plot di atas, ia didapati bahawa terdapat satu penurunan yang agak ketara pada komponen 17 dan ia boleh di interpretasi sebagai permulaan lantai hingar. Maka, 17 nilai tunggal yang pertama adalah mencukupi untuk membina semula siri masa utama dan menjadikan nilai  $r$  yang dipilih adalah 17. Selain itu, boleh dilihat bahawa bermula dari nilai eigen ganda tiga ke-18 hingga ke-96, nilai-nilai tunggal mula menurun secara perlahan dan menunjukkan bahawa nilai eigen ganda tiga tersebut sebagai komponen hingar.

Seterusnya, untuk mengesahkan hasil yang telah didapati, plot korelasi- $w$  telah dikenalpasti untuk menilai kesesuaian proses pengelompokan yang telah dilakukan. Rajah 2 menunjukkan korelasi- $w$  bagi 96 komponen yang dibina semula pada skala kelabu gred-20 dari putih ke hitam selari dengan nilai mutlak bagi korelasi dari 0 hingga 1. Daripada maklumat ini, 17 eigen ganda tiga yang pertama dipilih untuk pembinaan semula siri masa yang asal dan selebihnya dianggap sebagai hingar.



Rajah 2. Plot korelasi- $w$

Daripada analisis manual ini, didapati bahawa nilai  $L$  yang dipilih adalah 96 manakala nilai  $r$  adalah 17.

Pemilihan model SSA secara automatik dilakukan dengan mencari nilai  $L$  dan  $r$  yang optimum dengan mengurangkan nilai fungsi kerugian iaitu RMSE. Nilai  $L$  dan  $r$  yang memberi nilai RMSE terkecil akan memberikan nilai  $L$  dan  $r$  yang optimum bagi memperoleh ramalan bagi sesuatu data. Hasil pasangan  $L$  dan  $r$  yang optimum adalah 24 dan 14 yang menghasilkan nilai RMSE dalam sampel bersamaan dengan 58155.25.

Setelah tamat proses pemodelan, langkah seterusnya adalah untuk membandingkan hasil ramalan data luar sampel dari Januari 2019 (cerapan ke-229) hingga Disember 2019 (cerapan ke-240) bagi kedua-dua model SSA. Nilai ramalan bagi kedua-dua model telah dihitung untuk membandingkan nilai RMSE, MAE dan MAPE masing-masing. Jadual 4 menunjukkan nilai RMSE, MAE, dan MAPE yang diperoleh. Daripada jadual tersebut, dapat dirumuskan bahawa model SSA (24,14) mempunyai prestasi yang lebih baik disebabkan oleh nilai RMSE, MAE, dan MAPE yang ketiga-tiganya lebih rendah daripada SSA (96,17).

Jadual 4. Prestasi ramalan SSA (24,14) dan SSA (96,17) bagi data luar sampel

| Model       | RMSE       | MAE        | MAPE    |
|-------------|------------|------------|---------|
| SSA (24,14) | 196283.872 | 163174.250 | 7.49217 |
| SSA (96,17) | 203656.460 | 168899.667 | 7.85437 |

### 3.4. Perbandingan model SARIMA dan SSA

Langkah terakhir adalah membandingkan prestasi model ARIMA (3,1,3)(2,0,0)[12] dan SSA (24,14) secara menyeluruh iaitu prestasi dalam sampel dan luar sampel. Menggunakan kedua-dua model yang telah terpilih dalam peringkat pemodelan iaitu model SSA dan model ARIMA, prestasi bagi data dalam sampel bagi model-model tersebut dapat dilihat pada Jadual 5. Dengan membandingkan nilai RMSE, didapati model SSA (24,14) mempunyai prestasi dalam sampel yang lebih baik kerana nilai RMSE model tersebut lebih kecil berbanding ARIMA (3,1,3)(2,0,0)[12].

Jadual 5. Perbandingan prestasi data dalam sampel menggunakan RMSE

| Model                    | RMSE     |
|--------------------------|----------|
| SSA (24,14)              | 58155.25 |
| ARIMA (3,1,3)(2,0,0)[12] | 126674.6 |

Bagi prestasi model untuk data luar sampel iaitu data dari Januari 2019 hingga Disember 2019, nilai RMSE, MAE dan MAPE telah dihitung bagi kedua-dua model ARIMA dan SSA, dan nilai tersebut dapat dilihat pada Jadual 6.

Jadual 6. Prestasi ramalan bagi data luar sampel

| Model                    | RMSE       | MAE        | MAPE    |
|--------------------------|------------|------------|---------|
| ARIMA (3,1,3)(2,0,0)[12] | 154728.393 | 133250.417 | 6.08566 |
| SSA (24,14)              | 196283.872 | 163174.250 | 7.49217 |

Berdasarkan Jadual 6, model ARIMA mempunyai nilai yang paling kecil bagi kesemua pengukur ralat yang digunakan dan ini merumuskan bahawa ARIMA (3,1,3)(2,0,0)[12] merupakan model yang lebih baik bagi proses peramalan.

#### 4. Kesimpulan

Oleh itu, dapat dirumuskan bahawa model yang memberi prestasi dalam sampel yang lebih baik tidak semestinya akan memberikan prestasi luar sampel yang konsisten dan juga lebih baik daripada model yang lain. Hasil ini selari dengan kajian Oh & Morzuch (2005) yang turut mendapat hasil yang berbeza bagi data dalam sampel dan data luar sampel. Disebabkan ARIMA (3,1,3)(2,0,0)[12] mempunyai prestasi data luar sampel yang lebih baik berbanding SSA, maka hasil ini turut konsisten dengan kajian yang dilakukan oleh Abdul Halim & Muda (2020) yang mengkaji pemodelan dan peramalan jumlah permintaan pelancong bagi setiap negara ke Malaysia menggunakan model SARIMA dan SSA.

#### References

- Abdul Halim S.S. & Muda N. 2020. Modeling and forecasting of tourism demand in Malaysia. *International Journal of Current Science Research Review* 3(1): 230-244.
- Chetty. 2011. Advantages of demand forecast for the tourism industry. <https://www.projectguru.in/demand-forecast-tourism/> (10 Oktober 2020).
- Choden & Unhapiat S. 2018. ARIMA model to forecast international tourist visit in Bumthang, Bhutan. *Journal of Physics: Conference Series* 1039: 1-10.
- Hassani H. 2007. Singular spectrum analysis: methodology and comparison. *Journal of Data Science* 5(1): 239-257.
- Hassani H. & Mahmoudvand R. 2018. *Singular Spectrum Analysis Using R*. Tehran: Palgrave Advanced Text in Econometrics.
- Norbaizura B. & Zainuddin A. 2014. Forecasting international tourist demand from the US, Japan and South Korea to Malaysia: A SARIMA approach. *AIP Conference Proceedings* 1605: 955-960.
- Oh C.O. & Morzuch B.J. 2005. Evaluating time-series models to forecast the demand for tourism in Singapore. *Journal of Travel Research* 43(1): 404-413.
- Silva I. & Alonso H. 2020. New developments in the forecasting of monthly overnight stays in the North Region of Portugal. *Journal of Applied Statistics* 1(1): 1-14.
- Statista. 2020. Global tourism industry – statistics & facts. <https://www.statista.com/topics/962/global-tourism/> (7 Oktober 2020).
- Tourism Malaysia. 2019. *Malaysia Tourists Profile 2018 by Selected Markets*. Putrajaya: Tourism Malaysia.
- Tourism Malaysia. 2020. *Malaysia Tourism Performance January – December 2019*.
- UNWTO. 2011. *Tourism Towards 2030 / Global Overview*. Madrid: World Tourism Organization.



*Program Sains Aktuari,  
Jabatan Sains Matematik,  
Fakulti Sains dan Teknologi,  
Universiti Kebangsaan Malaysia,  
43600 Bangi, Selangor, Malaysia  
Mel-e: humaida@ukm.edu.my\**

Received: 22 October 2021

Accepted: 29 November 2021

---

\*Corresponding author