Forecasting Total Visitor Arrival in NZ, from 2019 to 2024: An Application of Singular Spectrum Analysis

Surafel Tilahun
AUT
Department of Mathematical Science
surafeltilahun@hotmail.com

Abstract—Singular Spectrum Analysis (SSA) is becoming one of the most influential time series analysis techniques. It has been implemented in a variety of areas to detect essential components of time series as well as to forecast. SSA is a non-parametric technique used to identify seasonal and trend components by decomposing a series into small feasible components. We considered a time series of total visitors arrival in NZ, from 1921 to 2019 to illustrate the ability of SSA to gain a better forecasting accuracy compared to other statistical models. We forecasted five years of tourism demand in NZ using Recurrent SSA (Rssa), Autoregressive Intergrated Moving Average (ARIMA) and State Space (ETS) models. The outcome confirms the strength of SSA over other frequently used econometric models, specifically ARIMA and ETS, to forecast total visitor arrival in NZ. Furthermore, the five years of forecast using Rssa indicates that, under the same scenario, the total visitor arrival in NZ will keep following a yearly seasonal pattern with an increase of variance and an upward trend. In general, we advise the implementation of this technique in NZ to secure more accurate forecasting results.

Index Terms—Forecast, SSA, Visitors, Arrival, Tourism, Demand, Rssa, Time Series

1 Introduction

Increasingly, tourism is becoming the most important export industry regarding foreign exchange earnings in New Zealand (NZ), shedding light on this activity as one of the primary sources of income in the country [1]. Recently, different agencies have reported approximately 39.1 billion dollars of profit at the end of 2018, that is, an increase of nearly 8% compared to 2017, where indeed, activities related to tourism in NZ have been reported to have contributed $15.9 billion towards Gross Domestic Product (GDP), out of which 6.1% was a direct contribution and $11.1 billion an indirect one [1]. Overall, in 2018, a profit of around $27 billion was estimated from visitors.

Numerous results from different studies have been reported, forecasting the tourism demand in NZ . Nevertheless, they have mainly applied parametric statistical frameworks, such as ARIMA or Holt Winter. For instance, [2], conducted forecasting based on econometric models and [3], adopted univariate Autoregressive Moving Average (ARMA) to forecast tourism flow to NZ. However, these models may serve as well only up to a certain extent, as economic situations can easily influence them. Tourism demand hugely relies on the volatility of the economic situation, which in turn results in no stationerity in mean and variance. Thus, employing these models on tourism demand time series could lead to violation of the parametric assumption of stationarity, where indeed transforming the data into a stationary time series could result in loss of information.

The aim of this project is to employ a non-parametric forecasting technique, known as Singular Spectrum Analysis (SSA), to forecast the tourism demand in NZ from 2019 to 2024. The result from this primary framework will be compared with those from other parametric statistical forecasting models, such as ARIMA and State Space (EST). Additionally:

- A descriptive analysis of the time series between 1921 and 2019 will be performed.
- A residual analysis and an assessment of the goodness of fit of the model will be conducted.

These new goals are proposed to assess the pivotal components of time series (seasonality and trend). Furthermore, before predicting future activities of tourism demand in NZ, the model's prediction quality will be assessed by two types of statistical accuracy measurements to obtain the most accurate forecast: Mean Absolute Percentage Error (MAPE) and Mean Absolute Square Error (MASE).

Specifically, this study addresses the following questions:

- Does SSA outperform other statistical parametric forecasting models, namely ARIMA and EST?
- Will tourism demand in NZ increase or decrease in the next five years?

The selected model will be tested with our data to predict the tourism demand for the desired years. Hopefully, results from this project will benefit government agencies and stakeholders with regard to enlightening them to foresee the prospective number of visitors that will arrive in NZ. Generally speaking, the results will serve to cause these entities to think ahead on what can be done to either retain the success or improve the industry implicitly. In addition to that, we expect that this research paper will illuminate the effectiveness of the SSA in terms of forecasting total visitor arrival (TVA) in NZ over the candidate models (ARIMA and ETS).

2 Background

There exists a considerable body of research on employing various forecasting models and comparing their performance with other candidate models [4]. However, it is an arduous task to decide which forecasting model outperforms at all times despite the nature of time series. A research conducted by [5] shows that the prediction accuracy of different forecasting models varies based on data frequencies implemented in the model estimation, the country of destination, explanatory variables incorporated in the model and length of prediction horizon concerned. Besides, [6] concluded that a consistent forecasting model had not been built for tourism demand. [7] suggests that the ARMA model outperforms the other two statistical forecasting models at any point of time. On the other hand, [3] proves that ARIMA does not even outperform Naive 1(no-charge) model. Previous research by [2] used three different forecasting models: ARIMA, a simple linear regression model, and forecast model produced by Delphi method. However, none of these models seemed to accurately measure the forecast of tourism demand because most of these models rely on the nature of the time series. Therefore, this project will forecast tourism demand in NZ based on SSA (non-parametric), and the result will be compared with those of the other models, previously used to forecast tourism demand in NZ. As far as we know, no previous research has employed SSA to forecast tourism demand in NZ. Hence, this work intends to be the first work in this direction in the country.

3 Limitation

While this powerful forecasting technique is implemented to forecast tourism demand in NZ, there are still a few shortcomings with this paper. Indeed, the forecast prediction accuracy counts upon the number of determinants. For example, [8] justified that people from different regions have their own tendencies to select tourism destination. In addition, [9] reported that there are five stages visitors undergo. The first one is deciding the time of travelling. The second is , extrapolating their expenses and organizing their budgets. The third is, the amount of time spent at the destination. The fourth is deciding the destination place, and the fifth is the mode of transportation. Hence, the author identified that these five stages determine the tourism flow to a particular country. Not only that, but also irregular events that occur in the country can negatively impact the number tourists arriving in NZ. For instance, the tragic unprecedented Christchurch terrorist attack could possibly have an evanescent negative impact on the number of visitors coming into NZ. Notwithstanding, the potential limitation of this analysis is that the data do not incorporate other features that might affect the number of visitors' arrivals in NZ. The data merely consist of time and number of visitors count. In the real world, according to the studies mentioned above, tourism flow is subject to manifold factors. Despite this fact, this research will not consider the deterministic variables that might influence the tourism demand in NZ.

4 Data Acquisition

This paper analyzes monthly data compiled by stats NZ. The data accommodates the TVA from April 1921 to January 2019. The data is available at [10]. The original data set will be employed to build the ARIMA, EST and SSA, as training set. Their performance will be compared according to their error on training set, then the preferred one will be used to forecast TVA for the interested years.

5 Methodology

5.1 Singular Spectrum Analysis (SSA)

SSA is one of the contemporary time series analysis and forecasting techniques that encompasses classical components of multivariate geometry, linear algebra, dynamics system and signal processing. SSA performs analysis by splitting the original time series into small feasible and explicable components, namely trend, oscillatory component (periodic wave) and noise [11]. SSA is an innovative non-parametric method that illustrates its efficiency of forecasting in a wide range of areas, from economics to physics [12]. Many applications of this methodology can be found in the literature that have shown it to be advantageous over other methods due to its ability to cope with uncommon assumptions, including the following:

- The evolution of Big Data might intensify noise, which impacts the accuracy of overall prediction due to misrepresentation of the signal [12].
- The sunny side of SSA is that, unlike other most models, it performs well for small size of data, [13].
- SSA is a model-free technique that is used to decompose the original time series into small interpretable components, trend, oscillation and noise. [14].
- The main benefit of SSA is its ability to perform a time series forecast, without taking into account the complexity of seasonal components, non-stationarity of trend and not-normality of the series [15].
- SSA is mainly essential for long time series analysis, in which seasonal and cycles are challenging to visualize [16].

5.2 Algorithm of SSA

The algorithm of SSA encloses two critical stages: Decomposition (First stage), and Reconstruction (Second stage).

The first stage engages the course of decomposition and the second stage involves reconstruction of the series.

5.2.1 Stage 1 : Decomposition

This stage comprises three different steps:

1. **Computing Trajectory Matrix**

This is an initial stage of SSA's algorithm where the original (univariate) time series is turned into a multidimensional series, which is known as trajectory matrix, built from L sequence of lagged vectors by forming K=L-N+1 lagged vectors, where $2 \leq L \leq \frac{N}{2}$. See [17] for determination of L size.

We assume a time series

$$M_t = (m_1, m_2, m_3, ..., m_N)^T \tag{1}$$

Then, the trajectory matrix of the univariate series (M_t) is presented as follows:

$$Y = \begin{bmatrix} \mathrm{m}_1 & m_2 & m_3 & . & . & . & m_K \\ m_2 & m_3 & m_4 & . & . & . & m_{K+1} \\ m_3 & m_4 & m_5 & . & . & . & m_{K+2} \\ . & . & . & . & & & . \\ . & . & . & & . & & . \\ . & . & . & & & . & . \\ m_L & m_{L+1} & m_{L+2} & . & . & . & m_N \end{bmatrix}$$

The vectors of Y shall be denoted as L lagged vectors [18]. The trajectory matrix above is known as Hankel matrix (a matrix with constant diagonal elements)

2. **Arranging the matrix to apply SVD**

In this step, YY^T is computed.

3. **Applying SVD to matrix YY^T**

Here, eigenvalues and eigenvectors $X = YY^T$ are computed and presented in the form of $U\Lambda V^T$, where U is orthonormal matrix that consists of eigenvectors of matrix X and Λ is a diagonal matrix that contains singular values of X corresponding to the eigenvectors of matrix U and can be denoted as follow:

$$\Lambda = diag(\sigma_1, \sigma_2, \sigma_3 ... \sigma_L) \tag{2}$$

The collection of these values is called singular spectrum [12] and they are arranged in descending order of magnitude,

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 ... \geq \sigma_L \geq 0 \tag{3}$$

Singular values, role is in separating the seasonal component from noise. They have a potential capacity to determine the frequency of the oscillation of the original series. The other two matrices, U and V^T or left and right eigenvectors, respectively, preserve the original matrices.

Following on from this stage, the SVD of the trajectory matrix Y can be written as

$$Y = Y_1 + Y_2 + Y_3 + Y_4 + ... + Y_L, \tag{4}$$

where $Y_i = \sqrt{\sigma}_i U_i V_i^T$ and $V_i = Y^T U_i / \sqrt{(}\sigma_i)$, if $\sigma_i = 0$, then set $Y_i = 0$. In the context of SSA, Y_i matrices are referred to as elementary matrices, while the collection of $(\sqrt{\sigma_i}, U_i, V_i^T)$ is known as i^{th} eigentriple of the SVD [12].

5.2.2 Stage 2: Reconstruction

This stage consists of two steps: Grouping and Diagonal averaging.

1. **Grouping**

In this step, the elementary matrices (X_i) are broken down into several groups, and the matrices within each group are summed up. Our primary goal here is to make the signal and structure-less components recognizable.

Breaking down the set of indices {1,2,3...L} into $I_1, ..., I_n$ disjoint subsets results in the following equation:

$$Y_{I1} + ... + Y_{In}, \quad (5)$$

where $Y_{Ij} = \sum_{l \in I_j} Y_l, j = 1, 2, 3, ...n$. The procedure of selecting these values $(I1, ..., I_n)$ is known as grouping.

This is one of the essential phases of SSA's algorithm to obtain an optimal separation between noise and signal. grouping step enables us to analyze the scatterplot of right eigenvectors, periodogram or eigenvalue function to distinguish between signal and noise. Once the appropriate eigenvalues are selected, the effectiveness of separability can be evaluated by weighted correlation (w-correlation), which is a useful method to evaluate the relationship between the reconstructed components. A smaller or close to zero w-correlation between reconstructed components signifies that these components should be grouped together.

2. **Diagonal Averaging**

The primary intention of this stage is Hankelizing the trajectory matrix, which allows us to convert reconstructed matrices back into a time series. We assume $k_{ij} \in K$, then averaging k_{ij} gives the $m - th$ element of the series, so that i+j equals 1+m. Through diagonal averaging of the components of Y_{Ii}, expansion of Y obtained from the grouping step leads to further expansion, which is $X = \hat{X}_{I1} + ... + \hat{X}_{In}$, where X_{Ii} is diagonalized X_{Ij}, which in turn produces an equivalent expression of original time series X_i into a sum of m series

$$m_t = \sum_{j=1}^{n} \hat{m}_t^{(t)}, \quad (6)$$

where $\hat{m}_N^{(j)} = (\hat{m}_1^{(j)}, ..., \hat{m}_N^{(j)})$, inferring the matrix $\hat{m}_{Tj}$. The final stage of this pipeline is applying the diagonal averaging formula:

$$\hat{m}_t^j = \frac{1}{s_2 - s_1 + 1} \sum_{i=s_1}^{s_2} \hat{x}_{i,t+1-i}, \quad (7)$$

where $s_1 = max(1, t - N + L), s2 = min(L, t)$.

[12] suggests three different approaches to achieve this stage. First, analyzing the scree plot of eigenvalues, the second approach is scatter plot of the eigenfunctions component, while the third one is w-correlation among possible groups. However, the author assumes these methods are not sufficient, and further research is required to establish a more general rule for achieving an optimal grouping.

5.3 Recurrent SSA (Rssa)

The main criterion to forecast a time series with SSA is that the original series M_N must satisfy a linear recurrent formula (LRF), with order of L-1. That is to say,

$$m_j = \sum\nolimits_{i=1}^{L-1} c_i m_{j-i}, L \leq j \leq N \quad (8)$$

, where $(c_1, c_2, ...c_d)^T$ are obtained from the eigenvectors (U_i). Like the Autoregressive model, the recurrent forecasting method relies on the weight of previous observations, although the weights are expressed through U_i. If the original time series meet the LRF (8), then at most d nonzero singular values exist in SVD of the trajectory matrix Y; hence, even if the number of window length L and K=N-L+1 is larger than d, we only take into account at most d matrices X_i so as to reconstruct the series [19].

Let $u^2 = \alpha_1^2, \alpha_2^2, \alpha_3^2, ..., \alpha_r^2$, where α_i is the last component of eigenvector $U_i (i = 1, 2, 3, ..., r)$. In addition, assume for any vector U^L represented by $U^G \in R^{L-1}$ ibe the vectors incorporating the first L-1 components of the vector U. Let $M_{N+h} = (m_1, m_2, ..., m_N + h)$ is h terms of recurrent forecast. Then the h step ahead forecast by Rssa can be defined by the following formula.

$$m_i = \begin{cases} \hat{m}, & \text{for i=1,2,...,N.} \\ \sum_{j=1}^{L-1} c_j m_{i-j}, & \text{for i=N+1+2,...,N+h.} \end{cases} \quad (9)$$

Where $\hat{m}_i (i = 1, 2, 3, ..., N)$ creates the reconstructed series (with out the structure less component) and vector $A = (c_{L-1}, ..., c_1)$ is computed by the following formula.

$$A = \frac{1}{1 - v^2} \sum_{i=1}^{r} \alpha_i U_i^G \quad (10)$$

6 Descriptive Analysis

Figure 1 shows a series of TVA in NZ for the last 98 years. An upward trend exists in the series, and the amplitude of the seasonal shift increases following the general trend. The trend is almost consistent from 1920 to 1960, but then it shows a dramatic rise up until present.

7 Findings

This section gives a brief description of results that are obtained from SSA. Here, we also forecast TVA by Rssa and compare its performance against ARIMA and ETS.

7.1 Results

Figure 2 exhibits the first 50 eigenvalues of the SVD. From this figure, identifying the eigenvalues that embrace principal components of the series (trend and seasonality) is conceivable. The single eigenvalue or the one that does not have any pair contributes to the trend. This aspect of the plot is also apparent in Figure 7 that the leading eigentriple contributes the most to the trend component of the series, or 93.01% of the trend is well represented by the leading eigentriple.

It is important to bear in mind that the eigenvalues that are associated with the respective seasonal components are usually close to each other in the scree plot. From the periodogram plot in figure 3, it is explicit that there is yearly seasonality (period of 12) that is in consists of sine waves. Table 3 extends to the frequency and period of the corresponding seasonality in the series.

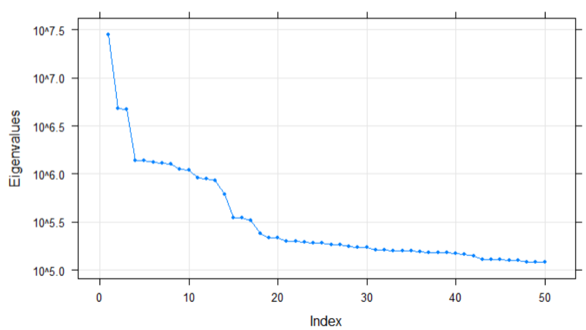

Fig. 1. Scree plot of SSA for total visitors, arrival in NZ (L=82).

With the guidance of Figure 2, Figure 4 and Figure 5, we can identify the required sine waves disguised in the SVD. Figure 4 illustrates the sine\cosine waves that are created by the respective pairs of eigenvectors. Note that the scatter plot of sine waves with same period and zero phase will have p-vertex polygon shape. Hence, the scatterplot of eigenvectors (Figure 4) shows that there are almost six polygons that contribute to the seasonality of the series with corresponding periods given in Table 4. The leading 16 eigentriples with a 99.93% share can be related to the six periodicities in ET2-3, ET4-5, ET7-8, ET9-10, ET11-12 and ET15-16, which are well represented in the periodogram (figure 3). These pairs of eigenvectors describe the annual period modulated harmonic (with-out the trend component). This is also specifically described by W-correlation matrix (figure 5). The first block (ET1) contributes to the trend component. The weight correlation plot also depicts that the first 16 eigentriples have less noise compared to the eigentriples greater than 16, and the considered pairs of components have a strong correlation within and almost zero correlation between, which is an implication of good separability between signal and noise.

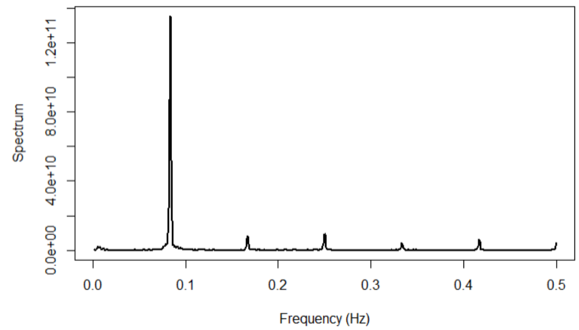

Fig. 2. Periodogram of the series.

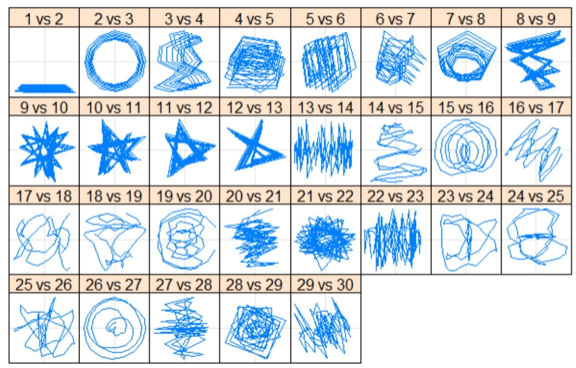

Fig. 3. Scatterplot of eigenvector pairs or left eigenvector pairs (L=82).

By taking into account the pairs of components that have been mentioned above, we reconstruct here the decomposed series. Figure 6 shows the reconstructed time series using the regarded components. F1 is a trend that is reconstructed by 1^{st} eigentriple. F2, F3 , F4, F5, F6 and F7 are the seasonalities of series that are reconstructed by ET2-3, ET4-5, ET7-8, ET9-10, ET11-12 and ET15-16.

7.2 Forecast by SSA and comparison against ARIMA and ETS

Our original time series Y_N satisfies the criteria of linear recurrent formula:

$$Y_N = c_1 m_{n-1} + ... + c_d m_{n-d}, \qquad (11)$$

for n = $L+1, ..., N$. Therefore, it is all systems go to forecast TVA by using recurrent forecasting method.

We have implemented Rssa to forecast TVA in NZ, from 2019 to 2024. Two accuracy metrics are applied to compare SSA's performance against other statistical frameworks (ARIMA and ETS), in particular, mean

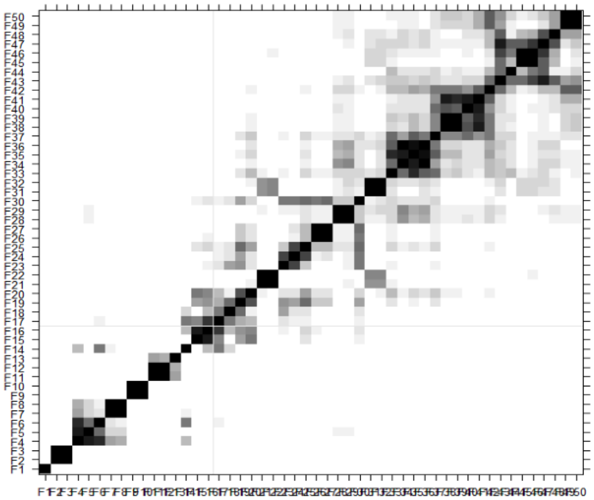

Fig. 4. W-correlation matrix (L=82).

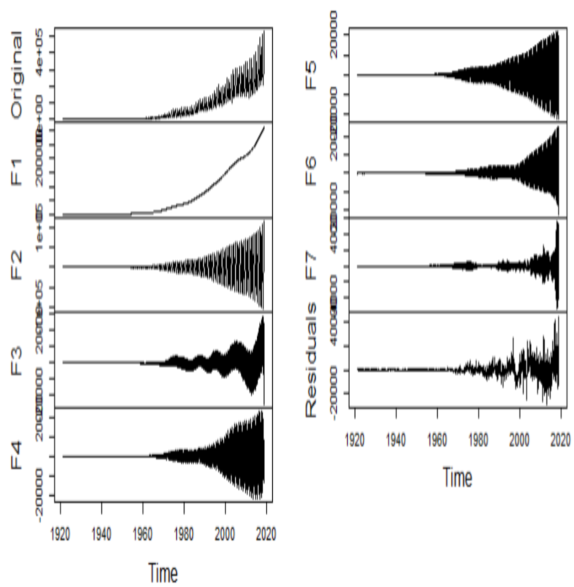

Fig. 5. Reconstructed series by the corresponding eigentriples (L=82).

absolute percentage error (MAPE) and mean absolute scaled error (MASE) [20], which can be calculated as follows:

$$MAPE = (\sum_{t=1}^{n} \| \frac{O_t - F_t}{O_t} \|) \frac{100\%}{n}, \quad (12)$$

where O_t denotes the original time series at time t and F_t denotes the forecast at time t.

Mean absolute scaled/square error (MASE) is another forecasting accuracy measurement that is adopted here to measure forecast accuracy by the

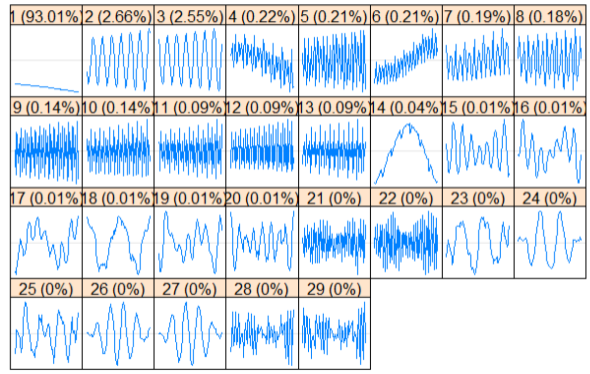

Fig. 6. Right Eigenvectors (L=82) for total visitors, arrival in NZ data.

three proposed methods. According to [21], MASE is calculated as follows:

$$MASE = MEAN\{ \| \frac{e_t}{\frac{1}{n-1}\sum_{t=2}^{n} |O_t - O_{t-1}|} \| \}, \quad (13)$$

where O_t is the original observation at time t and e is the forecast error.

The eigentriples that are considered to forecast the tourism demand span from 1 to 16.

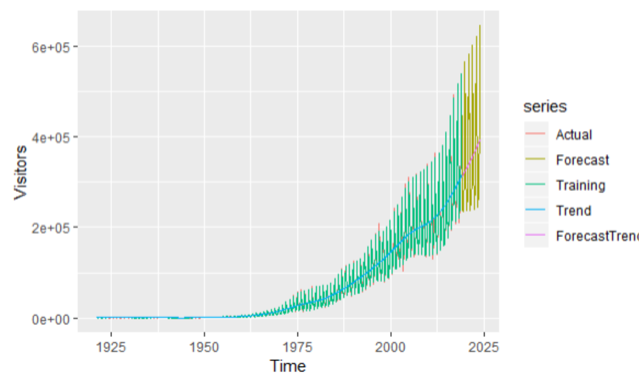

Fig. 7. Forecast of TVA in NZ by SSA, from 2019 to 2021.

8 Discussion

Here, we compare the results of the proposed method with those of the traditional methods. As it is shown in table 4, SSA gives an MAPE of 7.9311 on training set , while ARIMA and ETS give 12.4749 and 11.2724 on training set. SSA also has the smallest MASE (0.3894) for training set compared to ARIMA and ETS (0.6009 and 0.6546, respectively), which is consistent with MAPE. From these results, it is clear that SSA is fitting the training set better than the other two models.

Finally, we performed forecast of the TVA by recurrent forecasting method, and the result is shown in figure 8. The trend and seasonal components of the series was forecasted individually, using the the corresponding essential components we discovered above in result section and recombined them to form one series as shown in figure 8. The result suggests that the five years of TVA will keep following a yearly seasonal pattern, while the trend and variance will keep growing steadily. The numerical values of TVA forecast for the next three years is given in table 5. The numerical values of TVA trend forecast is also shown in table 6

9 Conclusion

This paper considers a time series of TVA in NZ observed between April 1921 and January 2019. Our aim was to forecast the TVA in NZ, from December 2019 to January 2024, using one of the well-known forecasting techniques, namely SSA. We also showed the benefits of utilizing this technique. This is to say, it enables the user to decompose a time series into small different features and gain an overall impression of the principal components of a series. Other forecasting frameworks were also employed on the time series, and their performance was benchmarked against SSA. Based on the results from this research, we conclude that SSA provides a better forecasting precision over other statistical forecasting models. Moreover, our finding shows that TVA in NZ for the next five years will have an upward trend and increase of variance, while the seasonality remains the same.

One limitation of SSA, however, is that it requires the analyst's intervention to decide the appropriate window size (L). In other words, the user must select the window size with a trial and error approach, aiming to minimize the error of fit, which is expensive in time. It will be important that future research investigates a method to automat the selection of window size. Besides, this paper considers a univariate time series to explore the performance of SSA. As we mentioned earlier in the limitation section, tourism demand is influenced by a number of deterministic variables. Hence, future research could build a predictive model that takes the most relevant deterministic variables into account and examine the performance of SSA to forecast TVA in NZ. Some of the predictive variables might include weather, dollar exchange, fuel price, anomaly events in the country, etc.

SSA is not just limited to decomposing time series into multiple components and forecasting. It is also a type of technique that allows us to detect the change points. The authors would like to harness SSA so that one may detect the signal change point in total visitors, arrival in NZ time series as well as a time series that is compiled from different domains in NZ.

SSA has several applications in different domains and has been widely implemented worldwide to gain an insight into time series and to forecast. Many researchers have shown the advantages of SSA in terms of prediction accuracy. However, this method has rarely been applied here in NZ. This paper has made the first contribution to illustrate the robustness of this technique to gain a better TVA forecast in NZ. Ideally, this method should be replicated in a study where a detailed time series analysis is required, and therefore the authors advise implementation of SSA in NZ in order to get a better forecast result.

ACKNOWLEDGMENT

I would like to express my gratitude to Dr Victor Miranda, one of my lecturers at AUT who has been supporting and guiding me throughout this paper. Discussion in regards to this paper with him was an illuminating and I appreciate his insightful feedback and suggestions.

REFERENCES

[1] StatsNZ, "Tourism satellite account: 2018." Retrived from www.stats.govt.nz., 2018.

[2] A. B. M. W. A. Stroombergen and B. Patterson, "Understanding current and forecast visitors flows to the south island," 2018.

[3] E. SMERAL and M. WÜGER, "Does complexity matter? methods for improving forecasting accuracy in tourism: The case of austria.," *Journal of Travel Research*, vol. 44, no. 1, pp. 100–110, 2005.

[4] H. Hassani, A. Webster, E. S. Silva, and S. Heravi, "Forecasting us tourist arrivals using optimal singular spectrum analysis.," *TOURISM MANAGEMENT*, vol. 46, pp. 322 – 335, n.d.

[5] H. Song and G. Li, "Tourism demand modelling and forecasting—a review of recent research," *Tourism Management*, vol. 29, no. 2, pp. 203 – 220, 2008.

[6] H. Song and G. Li, "Tourism demand modelling and forecasting - a review of recent research.," *TOURISM MANAGEMENT*, vol. 29, no. 2, pp. 203 – 220, n.d.

[7] V. Cho, "Tourism forecasting and its relationship with leading economic indicators.," *Journal of Hospitality and Tourism Research*, vol. 25, no. 4, pp. 399–420, 2001.

[8] V. Cho, "A study of the non-economic determinants in tourism demand.," *International Journal of Tourism Research*, vol. 12, no. 4, pp. 307 – 320, 2010.

[9] J. L. Eugenio-Martin, "Modelling determinants of tourism demand as a five-stage process:a discrete choice methodological approach.," *Tourism Hospitality Research*, vol. 4, no. 4, p. 341, 2003.

[10] StatsNZ, "Visitor arrival totals (monthly)." Data retrieved from Infoshare, http://archive.stats.govt.nz/infoshare/. Browth-Infoshare-Tourism-International Travel and Migration-Visitor arrival totals (Monthly) .

[11] N. Golyandina and A. Zhigljavsky, *Singular spectrum analysis for time series.* SpringerBriefs in statistics, Springer, 2013.

[12] H. Hassani and R. Mahmoudvand, *Singular spectrum analysis : using R.* Palgrave Advanced Texts in Econometrics, Palgrave Macmillan, 2018.

[13] H. Hassani, "A brief introduction to singular spectrum analysis," 05 2019.

[14] M. Dettinger, M. Ghil, C. Strong, W. Weibel, and P. Yiou, "Software expedites singular-spectrum analysis of noisy time series.," *Eos, Transactions American Geophysical Union*, vol. 76, no. 2, pp. 12–21, 1995.

[15] H. Hassani, S. Heravi, G. Brown, and D. Ayoubkhani, "Forecasting before, during, and after recession with singular spectrum analysis.," *JOURNAL OF APPLIED STATISTICS*, vol. 40, no. 10, pp. 2290 – 2302, n.d.

[16] L. Michael and E. Bruce, "Automatic singular spectrum analysis and forecasting.," *JOURNAL OF APPLIED STATISTICS*, 2019.

[17] M. Atikur Rahman Khan and D. S. Poskitt, "A note on window length selection in singular spectrum analysis.," *Australian New Zealand Journal of Statistics*, vol. 55, no. 2, pp. 87 – 108, 2013.

[18] N. Golyandina and A. Zhigljavsky, *Singular spectrum analysis for time series.* SpringerBriefs in statistics, Springer, 2013.

[19] S. Sanei and H. Hassani, *Singular spectrum analysis of biomedical signals.* CRC Press, 2016.

[20] J. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, no. 1, pp. 69 – 80, 1992.

[21] T. Perera, D. Higgins, and W. Wong, "The evaluation of the australian office market forecast accuracy.," *Journal of Property Investment Finance*, no. 3, p. 259, 2018.

Table 1 shows a summary of the TVA in NZ. The average number of visitors, arrival in NZ for the last 98 years is 63,445. The maximum number of visitors within this period is 529,255, while 40 is the minimum.

TABLE 1
Summary of TVA in NZ.

Minimum	1st Qu.	Median	Mean	3rd Qu	Max
40.00	1161.25	13876	63445.15	99179.00	529255.00

TABLE 2
Frequency and period in seasonality of TVA in NZ.

Frequency	Spectrum	Period (Months)
0.085	135580812654	11.9761
0.0826	87259790383	12.1065
0.0843	52398703056	11.8624
0.0818	24193765223	12.2249
0.0852	17120884859	11.7371
0.0809	10039153367	12.3609

TABLE 3
Accuracy metrics of SSA and other two statistical frameworks.

	Data	MAPE (%)	MASE
ARIMA(4,1,2)(2,1,0)[12]	Training Set	12.47497	0.6009397
ETS(M,Ad,M)	Training Set	11.27242	0.65469
SSA	Training Set	7.931123	0.3894523

TABLE 4
TVA in NZ forecast by SSA, between February 2019 and January 2024.

Month	2019	2020	2021	2022	2023	2024
Jan		428748.7	439280.3	448970.3	460668.3	474878.1
Feb	445902.4	464490.3	479378.2	494439.3	512378.8	.
Mar	399828.9	414845.0	426906.6	439568.0	454910.2	.
Apr	321302.2	330131.3	336086.0	342752.2	351992.2	.
May	237439.9	241324.1	242927.0	245515.2	250699.1	.
Jun	234474.3	237129.6	237848.8	239595.5	243638.1	.
Jul	259876.5	260889.6	259924.8	259908.2	262067.6	.
Aug	256266.0	257828.5	257645.4	258520.4	261970.4	.
Sep	275112.1	277567.4	278077.4	279528.9	283290.2	.
Oct	288092.4	288092.4	290581.0	291759.2	295124.0	.
Nov	382476.6	392912.7	401851.0	412646.3	426387.2	.
Dec	564552.4	583306.2	601221.3	621531.7	644986.5	.

TABLE 5
TVA in NZ, trend forecast, between February 2019 and January 2024.

Month	2019	2020	2021	2022	2023	2024
Jan		324078.5	340028.7	356915.4	374570.9	392844.7
Feb	310347.2	325368.6	341401.2	358359.2	376070.9	.
Mar	311559.2	326666.3	342780.9	359809.2	377576.6	.
Apr	312779.1	327971.7	344167.8	361265.3	379087.6	.
May	314006.5	329284.6	345561.3	362726.8	380603.3	.
Jun	315241.3	330604.6	346961.3	364193.4	382123.3	.
Jul	316483.4	331931.6	348367.5	365664.7	383647.0	.
Aug	317732.6	333265.5	349779.5	367140.3	385173.9	.
Sep	318988.6	334605.6	351196.7	368619.2	386702.6	.
Oct	320250.9	335951.8	352618.5	370101.0	388233.0	.
Nov	321519.9	337304.2	354045.5	371586.7	389766.3	.
Dec	322795.7	338663.1	355477.6	373076.5	391303.3	.