

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA INFORMÁTICA

**PREDICCIÓN DE TURISTAS EXTRANJEROS QUE
ARRIBARÁN A CHILE UTILIZANDO REDES NEURONALES
ARTIFICIALES Y SARIMA**

JOAQUÍN ESTEBAN GONZÁLEZ NÚÑEZ

INFORME FINAL DEL PROYECTO
PARA OPTAR AL TÍTULO PROFESIONAL DE
INGENIERO CIVIL EN INFORMÁTICA

Diciembre 2015

Pontificia Universidad Católica de Valparaíso
Facultad de Ingeniería
Escuela de Ingeniería Informática

**PREDICCIÓN DE TURISTAS EXTRANJEROS QUE
ARRIBARÁN A CHILE UTILIZANDO REDES NEURONALES
ARTIFICIALES Y SARIMA**

JOAQUÍN ESTEBAN GONZÁLEZ NÚÑEZ

Profesor Guía: Nivaldo Rodríguez A.

Profesor Co-referente: José Miguel Rubio L.

Carrera: Ingeniería Civil Informática

DICIEMBRE 2016

Quiero agradecer a mi familia por su apoyo y confianza durante este largo camino para convertirme en profesional. Agradecer especialmente a mi hermano Juan Marcos por toda la ayuda brindada, sobre todo en los momentos más difíciles. Finalmente agradezco a Fanny por ser una gran compañera en este proceso y por ser mi compañera de la vida.

Resumen

Una acertada predicción de la demanda turística es un problema crucial para el turismo y la industria de servicios porque puede proporcionar eficientemente la información básica para la planificación del turismo y la posterior formulación de políticas. Para lograr el mejor resultado en la predicción de la demanda turística se empleó dos modelos autoregresivos, SARIMA y Red Neuronal.

Los parámetros del modelo SARIMA fueron obtenidos utilizando el algoritmo de máxima verosimilitud, la red neuronal se entrenó utilizando el algoritmo de Levenberg-Marquardt. Ambos modelos utilizaron para esta fase el 85% de los datos disponibles. Los datos de turistas arribados a Chile fueron los utilizados. Los resultados empíricos indican que el modelo SARIMA obtuvo un mejor rendimiento que la Red Neuronal a una proyección de doce meses.

Palabras Claves: SARIMA, Redes Neuronales artificiales, Modelo Predictivo, Llegada de turistas, Chile

Abstract

An accurate prediction of tourism demand is a crucial problem for tourism and the service industry because it can efficiently provide the basic information for tourism planning and subsequent policy. To achieve the best result in predicting the tourist demands two autoregressive models was used, autoregressive SARIMA and Neural Network.

ARIMA model parameters were obtained using the maximum likelihood algorithm; the neural network was trained using the Levenberg-Marquardt algorithm. Both models used for the training phase 85% of available data. The monthly data in tourist arrivals in Chile were used. The empirical results indicate that the SARIMA model performed better than the neural network to a projected twelve months.

Keywords: SARIMA, Artificial Neural Network, Predictive Model, Tourists Arrival, Chile

Índice

Listado de Figuras	vii
Listado de Tablas	viii
Lista de Abreviaturas	ix
1. Introducción.....	1
1.1 Descripción del Problema	1
1.2 Análisis de Objetivos.....	2
1.2.1 Objetivo General	2
1.2.2 Objetivos Específicos.....	2
2. Marco Teórico	3
2.1 Estado del Arte	3
3. Caso de Estudio	6
3.1 La relevancia del turismo en Chile	6
3.2 Conjunto de Datos.....	7
3.3 Detección de estacionalidad.....	8
3.4 Detección de tendencia.....	11
4. Resolución del Caso Mediante un Modelo SARIMA	12
4.1 ARIMA.....	12
4.2 SARIMA.....	13
4.3 Aplicación del modelo SARIMA	13
4.4 Detección parámetros del modelo SARIMA con mejor ajuste	14
4.5 Estudios de los residuos del modelo.....	15
4.6 Salidas del modelo	17
4.7 Predicción.....	18
5. Resolución del Caso Mediante una Red Neuronal Autoregresiva	20
5.1 Red Neuronal	20
5.2 Redes Neuronales Propuestas	21
5.3 Algoritmo de entrenamiento Levenberg–Marquardt	22
5.3.1 Algoritmo de Gradiente Descendente.....	22
5.3.2 Algoritmo de Gauss-Newton.....	22
5.3.3 Algoritmo de Levenberg–Marquardt.....	23
5.4 Predicción realizada mediante Red Neuronal utilizando doce Neuronas en la capa de entrada	23

5.4.1 Salida Red Neuronal	24
5.4.2 Estudios de los Residuos	25
5.5 Predicción realizada mediante Red Neuronal variando las Neuronas en la capa de entrada desde 13 hasta 24 neuronas.	27
5.5.1 Predicción utilizando una Red Neuronal Autoregresiva de orden 13	27
5.5.2 Salida Red Neuronal	28
5.5.3 Estudios de los Residuos	29
6. Comparativa entre los modelos SARIMA y Redes Neuronales Artificiales Autoregresiva ..	31
7. Predicción de turistas de los 12 periodos próximos a la serie de tiempo en estudio.....	32
8. Conclusión.....	33
9. Anexo.....	35
9.1 NSE.....	35
9.2 RMSE	35
9.3 Operador Backshift.....	36
9.4 Función de Autocorrelación	36
9.5 Función de Autocorrelación Parcial.....	36
9.6 Criterio de información de Akaike (AIC).....	37
9.7 Criterio de información de Akaike corregido (AICc).....	37
9.8 Diferenciación	38
9.9 Función Sigmoide.....	38
9.10 Estimador Máxima Verosimilitud	38
9.11 Algoritmo Levenberg–Marquardt.....	39
9.12 Ejemplo numérico predicción utilizando SARIMA	39
10 Referencias.....	41

Listado de Figuras

Figura 1: Exportación de bienes y turismo (US\$ MM)	6
Figura 2: Comparación de Turismo con otros Sectores Económicos.....	7
Figura 3: Extranjeros que han Arribado a Chile.....	8
Figura 4: Función de Auto-Correlación a la Serie Original	9
Figura 5: Periodograma Suavizado Aplicado a la Serie	10
Figura 6: Periodograma sin suavizar	10
Figura 7: Tendencia de la Serie Original.....	11
Figura 8: Diferenciación de los datos en primer grado	15
Figura 9: ACF Residuos	16
Figura 10: PACF Residuos	16
Figura 11: Gráfico Cuantil-Cuantil (QQ) de los residuos	16
Figura 12: Histograma de los residuos	17
Figura 13: Predicción versus Valores Reales	19
Figura 14: Estructura Red Neuronal.....	20
Figura 15: Red Neuronal 12 Neuronas Entrada.....	24
Figura 16: Datos Predichos utilizando RNA	25
Figura 17: ACF RNA Orden 12	26
Figura 18: Histograma RNA Orden 12.....	26
Figura 19: Estructura RNA Orden 13.....	27
Figura 20: Datos Predichos Utilizando RNA Orden 13	28
Figura 21: ACF RNA Orden 13	29
Figura 22: Histograma RNA Orden 13.....	30
Figura 23: Comparativa datos Predichos.....	31
Figura 24: Predicción de los 12 Periodos Siguietes de los Datos Observados.....	32

Listado de Tablas

Tabla 1: Factores del modelo.....	18
Tabla 2: Resultados de Predicción.....	18
Tabla 3: Resultados de Predicción utilizando RNA 12 entradas.....	24
Tabla 4: Datos Observados y Estimados por RNA Orden 13	28
Tabla 5: Resultados Estadísticos	31
Tabla 6: Factores del modelo para ejemplo.....	39
Tabla 7: Residuos SARIMA.....	39

Lista de Abreviaturas

ACF: Función de autocorrelación que refleja la correlación entre un valor con sus valores pasados.

AIC: Criterio de información de Akaike, mide la calidad relativa del modelo.

AICc: Criterio corregido de información de Akaike.

ANN: Artificial Neural Network (Red Neuronal Artificial), modelo para la estimación de valores futuros en base a valores de las variables ingresadas al modelo.

ARIMA: Modelo que cuenta con una parte autoregresiva, integrada y de promedios móviles.

INE: Instituto Nacional de Estadística, institución encargada de realizar estudios estadísticos con fines públicos.

MSE: Error cuadrático Medio, mide la diferencia numérica entre valores predichos con los valores observados.

NSE: Coeficiente de eficiencia Nash–Sutcliffe. Mide la el poder predictivo del modelo.

PACF: Función parcial de autocorrelación, refleja la correlación de un valor con sus valores pasados eliminando la influencia en la correlación de los valores entre medio.

R: Coeficiente de Pearson, mide la correlación entre los datos observados y datos obtenidos por los modelos.

RMSE: Raíz cuadrática del error medio, mide la diferencia numérica entre los valores predichos con los valores observados.

SARIMA: modelo Seasonal ARIMA (ARIMA estacional).

SERNATUR: Servicio Nacional de Turismo, institución encargada de promover y difundir el desarrollo de la actividad turística de Chile.

1. Introducción

1.1 Descripción del Problema

El turismo en Chile es una actividad en crecimiento y que aporta miles de empleos directos, como también aporta un relevante porcentaje al PIB chileno. Es por aquello, que es necesario tener un pronóstico de los turistas que arribarán al país, para que exista una preparación logística y de infraestructura con el objetivo de dar un mejor servicio e invertir en diferentes tópicos si es necesario, anticipándose a la llegada de ellos. Esta inversión podrá ser por parte de los actores privados ligados a la actividad turística o por el gobierno de turno.

En el presente trabajo de investigación se expondrá una predicción del número total de turistas extranjeros que arribarán a Chile en base a los datos obtenidos en los informes anuales de turismo del Instituto Nacional de la Estadística (INE) y del Servicio Nacional del Turismo (SERNATUR). Esta predicción tendrá un alcance de un año, por lo tanto, permitirá tener una noción, con un pertinente porcentaje de certeza, sobre la cantidad de extranjeros que arribarán a Chile en las próximas temporadas. Para realizar lo anteriormente expuesto, se utilizarán modelos autoregresivos y análisis estadístico; el resultado obtenido será contrarrestado con datos reales y pasados para comprobar la precisión del modelo desarrollado.

Como primera aproximación se utilizará un modelo de predicción denominado ARIMA estacional o SARIMA por sus siglas en inglés. Posteriormente se realizará la predicción anteriormente detallada utilizando un modelo de red neuronal artificial (ANN) y así poder comparar que tipo de modelo autoregresivo tiene mejor precisión en la estimación, si un modelo paramétrico como es SARIMA o un modelo no paramétrico como son las redes neuronales artificiales.

Este trabajo de investigación se dividirá en capítulos consecutivos. Cada capítulo abordará una temática importante para la realización de éste. Se hará una breve presentación del estado del arte de la temática, en donde se detallará los trabajos previamente realizados sobre la materia del turismo, la aplicación de los modelos en otros rubros o actividades, y la aplicación de estos modelos en el rubro.

A continuación se explicará la problemática abordada y la relevancia del turismo para Chile. En los siguientes capítulos se detallará los datos utilizados para finalizar con la exposición del primer modelo utilizado para obtener una estimación de la cantidad futura de turistas que arribarán a Chile en los próximos 12 meses. Se detallará las transformaciones a la data original que se realizarán, el procedimiento de cómo se obtuvieron los parámetros del modelo, el modelo obtenido y finalmente se contrastará con datos reales y pasados para comprobar la precisión del modelo desarrollado. Para este contraste se utilizarán métricas tales como: RMSE, R y NSE.

1.2 Análisis de Objetivos

1.2.1 Objetivo General

- Implementar una Red Neuronal y SARIMA para determinar la cantidad de turistas extranjeros que arribaran a Chile en los próximos 12 meses y evidenciar cuál de los dos modelos autoregresivos obtiene un mejor desempeño en dicha tarea predictiva.

1.2.2 Objetivos Específicos

- Recolectar los datos necesarios que servirán como entrada a los modelos propuestos.
- Estimar los parámetros del modelo SARIMA mediante el algoritmo de mínimos cuadrados
- Estimar los parámetros de la Red Neuronal mediante el algoritmo de Levenberg-Marquardt.
- Comprobar la eficacia de los modelos desarrollados para verificar su concordancia con la realidad mediante herramientas estadísticas.
- Contrarrestar los resultados obtenidos mediante los modelos SARIMA y Red Neuronal

2. Marco Teórico

2.1 Estado del Arte

La existente literatura sobre la predicción de la demanda turística es amplia en términos de las diferentes técnicas empleadas y en término de los diferentes países que han sido cubiertos. . Varios autores incluidos [Chan y Lim, 2011], [Chu, 2008], [Coshall y Charlesworth, 2011] y, [Goh y Law, 2012] enfatizan en la importancia de la predicción para realizar inversión y planificar el desarrollo del turismo. Este mensaje es acentuado por autores tales como [Gounopoulos *et al.*,2012] y [Hui y Yen, 2002], quienes agregan que tales predicciones son además importantes como una consecuencia de la vulnerabilidad del turismo a gran fluctuación de la demanda. Algunos autores también enfatizan en la importancia del turismo a una economía en particular para reforzar la importancia de prever con exactitud la demanda turística. Algunos ejemplos incluyen trabajos como [Jackman y Greenidge, 2010] para Barbados y [Chu, 2011] para Macao. Aquellos lectores que busquen una completa reseña de la literatura, la publicación por [Song y Li, 2008] cubre 121 estudios desde el 2000 a la fecha de publicación. Esta reseña ofrece una mayor explicación para el sostenido y extenso interés en pronosticar la demanda turística. Ellos encontraron que ningún modelo predictivo es superior a todos los otros en todas las posibles situaciones. Una mayor reseña sobre predicción y sobre problemas estrechamente relacionados con la modelación de la demanda turística es incluida en la publicación de [Song *et al.*, 2012].

Tal vez la forma más común de estudiar esta materia es aquella que evalúa el rendimiento de una o más técnicas de predicción. [Álvarez-Díaz y Roselló-Nadal, 2010] examinan la predicción de turistas británicos que arriban a las Islas Baleares, usando variadas metodologías. Ellos compararon el rendimiento de un modelo ARIMA y de una red neuronal auto-regresiva, encontrando que la última posee un mayor rendimiento. [Cho, 2003] investiga tres diferentes técnicas (suavizado exponencial, ARIMA univariante y redes neuronales artificiales) para la predicción de turistas arribados en Hong Kong, encontrando que las redes neuronales artificiales predicen con mayor precisión.

[Chu, 2008] explora que los modelos ARMA de integración fraccionada en la predicción de turistas que arribana Singapur, observando que ellos tienen un buen desempeño en comparación a modelos ARIMA más tradicionales. [Chu, 2011] compara un modelo lineal piecewise con tendencia autoregresiva, ARIMA estacional y un modelo ARIMA de integración fraccionada en la demanda turística para Macao, concluyendo que el modelo lineal piecewise es el más preciso. Igualmente, [Gil-Alana, 2005] considera predicciones utilizando datos mensuales de turistas arribados a Estados Unidos usando un procedimiento que combina la integración entera y fraccional en la variación estacional. Él encuentra evidencia de gran capacidad de memoria y debería revertir el comportamiento. [Goh y Law, 2002] utiliza datos para los turistas arribados a Hong Kong con la finalidad de comparar predicciones de un modelo estocástico no-estacionario estacional (SARIMA) y un modelo de componente de intervención (MARIMA) con una selección de otros ocho modelos de serie de tiempo. Sus

resultados sugieren que los modelos SARIMA y MARIMA tienen la mayor precisión para predecir de todos los modelos analizados.

[Lim y McAleer, 2001] analizaron el rendimiento de varios modelos de suavizado exponencial para la predicción de turistas arribados a Australia, concluyendo que utilizando modelos expresados en primeras diferencias mejoran la precisión de la predicción. [Shareef y McAleer, 2007] evalúan la capacidad de los modelos ARMA para capturar los efectos de la volatilidad en las series tiempo de los turistas arribados a las islas Maldivas. [Wan *et al.*, 2013], utilizan datos de los turistas arribados a Hong Kong, evalúan las propiedades de las predicciones desagregadas utilizando un modelo estacional ARIMA (SARIMA), relacionada a predicciones agregadas. Ellos encuentran que la suma de predicciones desagregadas provee una mejor precisión que una predicción agregada.

Numerosas publicaciones consideran las consecuencias de las crisis económicas a uno o más modelos predictivos de demanda turística. [Gounopoulos *et al.*, 2012] considera en predicción el impacto de la crisis macro-económica en la llegada de turistas a Grecia. Ellos compararon un número de diferentes métodos de predicción, encontraron que un modelo ARIMA es el más preciso y también desarrollaron un modelo de vector autoregresivo. [Smeral, 2010] usa un modelo de catástrofe cúspide para predecir la tasa de recuperación de turistas arribados a Taiwán desde la epidemia del Síndrome respiratorio agudo y grave. Sus resultados sugieren que el turismo desde China y Estados Unidos se recuperaron rápidamente pero no lo hizo el turismo desde Japón. En una hebra similar [Page *et al.*, 2012] estiman los efectos negativos de la epidemia de la gripe porcina en la llegada de turistas de Gran Bretaña utilizando modelo paramétrico de tiempo variable. [Fourie y Santana-Gallego, 2011] utiliza un modelo de gravedad para estimar y predecir el impacto de los eventos mega-deportivos en la llegada de turistas

Los modelos volátiles están contruidos sobre un modelo ARIMA al cual se le agrega una segunda ecuación para explicar la varianza condicional. [Coshall, 2009] provee un buen panorama de estas técnicas y su aplicación en la predicción de demanda turística. La especificación comúnmente más utilizada es el modelo GARCH, desarrollado por [Bollerslev, 1986]. Éste agrega al modelo ARIMA una ecuación para explicar la varianza condicional. Estas ecuaciones modelan la varianza condicional del actual periodo en términos de los cuadrados de los residuos retardados (capturando el impacto de corto plazo de las pasadas crisis económicas) y efectos a largo plazo de los valores retardados de la varianza condicional. Extensiones del modelo GARCH incluyen el TGARCH (el cual utiliza variables ficticias para modelar choques asimétricos) y modelos EGARCH. Por ejemplo, [Kim y Wong, 2006] utiliza ambos modelos, EGARCH y TGARCH, para proveer una predicción de la demanda turística en Korea con respuestas asimétricas para nuevos choques. [Coshall, 2009], en una aplicación para los turistas que salen desde Gran Bretaña, muestra que los pronósticos utilizando el modelo EGARCH pueden ser combinados con aquellos modelos de suavizado exponencial, tal que la combinación de pronósticos es más preciso que ambos por sí solos.

[Claveria y Torra, 2014] evaluaron el rendimiento de la predicción de una red neuronal artificial con un acercamiento relativo a los modelos con diferentes series de tiempo de turismo, modelos autoregresivos integrando medias móviles y autoregresiones. La evidencia empírica de los estudios presentados anteriormente muestra que los modelos no lineales que pueden representar ambas series de tiempo estacionales y no estacionales, generalmente

superan a los métodos lineales en el modelado de comportamiento turístico, como los clásicos modelos de serie de tiempo, tales como ARIMA, modelo de regresión lineal y regresión múltiple, suavizado exponencial y medias móviles.

Los modelos de redes neuronales artificiales han sido aplicados en diversas áreas de interés para los investigadores, pero específicamente en el área del turismo han sido utilizados en varias ocasiones para estimar la demanda turística de diversas zonas geográficas y países. Por ejemplo, [Law y Au, 1999] utilizaron un modelo de red neuronal artificial feedforward para predecir la demanda turística de Japón. [Law, 2000] aplicó una red neuronal de tipo backpropagation para predecir la demanda turística. [Pai y Hong, 2005] empleó máquinas de vectores de soporte para las redes neuronales con algoritmos genéricos para pronosticar los arribos de pasajeros a Barbados.

[Chen y Wang, 2007] aplicó vectores de soporte regresivos para redes neuronales con algoritmos genéticos en la predicción de la demanda turística. [Fernandes *et al.*, 2008] investigó y destacó la utilidad de la metodología de las redes neuronales artificiales como alternativa a la metodología Box-Jenkins en el análisis de la demanda turística. [Fernandes y Teixeira, 2009] desarrollaron un nuevo acercamiento de la metodología de las redes neuronales utilizando el tiempo en su entrada en vez de los doce pasados registros, como usualmente es utilizado; los autores intentaron comparar el clásico uso de la metodología de las redes neuronales con una modulación que utiliza los años y meses en la entrada.

3. Caso de Estudio

3.1 La relevancia del turismo en Chile

En el año 2013 el país tuvo un ingreso de divisas de US\$ 2.581,1 millones, generado por el gasto que realizaron los visitantes extranjeros en nuestro país, cifra que incluye un desembolso de US\$ 2.133,6 millones realizado por turistas, US\$35,1 millones por excursionistas y US\$ 412,4 millones generados por transporte internacional. El turismo contribuye con un 3,48% al PIB al año 2012. Además esta industria generó 286.165 empleos directo durante el año 2011 lo que representa un 3,8% del total empleo nacional.

Al comparar los rubros de exportación más relevantes del país es posible señalar que los ingresos de divisas por turismo receptivo representan el 5,9% de los ingresos por exportaciones de minería; el 47,4% de los productos químicos; el 54,5% del sector frutícola, el 71,5% de celulosa, papel y otros, el 92,3% en productos metálicos, maquinaria y equipo, y el 93,1% en salmón; superando en un 30,6% a los ingresos por exportaciones de vino y en un 13,6% a los productos forestales y muebles de madera.

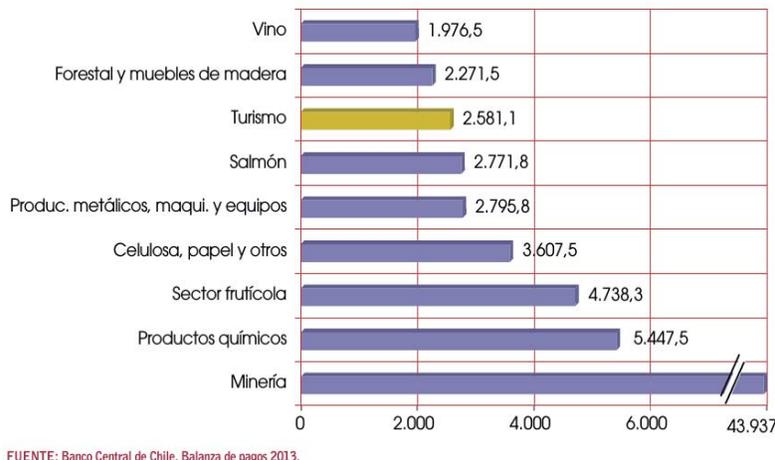


Figura 1: Exportación de bienes y turismo (US\$ MM)

En el año 2013 el aporte del turismo receptivo representó el 3,4% del total de las exportaciones de bienes, el 20,2% del total de las exportaciones de servicios y el 2,9% del total de las exportaciones de bienes y servicios de la economía. Los ingresos por turismo receptivo - contrastados con las exportaciones realizadas por sectores económicos representan los siguientes valores: 44,9% de las exportaciones en el sector agropecuario, silvícola y pesquero (US\$5.749,5 millones); 5,9% de la minería (US\$43.937,4 millones) y 9,6% de la industria (US\$26.997,4 millones).

INGRESOS POR TURISMO RECEPTIVO VERSUS EXPORTACIONES DE SECTORES ECONÓMICOS, AÑO 2013

SECTOR	Exportaciones (millones de US\$)	Turismo/Exportaciones (en porcentajes)
Agropecuario, Silvícola y Pesquero	5.749,5	44,9
Minería	43.937,4	5,9
Industria	26.997,4	9,6
Total	76.684,3	3,4

FUENTE: Balanza de pagos 2013-Banco Central de Chile. Sernatur.

Figura 2: Comparación de Turismo con otros Sectores Económicos

Según datos de Sernatur, en base a información proporcionada por la Jefatura Nacional de Extranjería y Policía Internacional, durante el año 2013 ingresaron al país 3.576.204 turistas de nacionalidad extranjera, 0,6% más en comparación al año anterior (3.554.279).

Los turistas residentes en el extranjero permanecieron en promedio 8,9 días en el país, efectuaron un gasto diario individual promedio (GPDI) de US\$ 66,9 y un gasto total por persona de US\$596,6. El turista europeo presentó una permanencia promedio de 19,2 días, un gasto diario medio por persona de US\$ 68,1 y un gasto total por persona de US\$ 1.310,4. Es importante señalar que Brasil es el país con el mayor GPDI en comparación al resto de los países, US\$ 144,5.

Entre los turistas que efectúan el mayor gasto total por persona durante su estadía en Chile se cuentan los residentes en España (US\$ 1.532,8), Asia (US\$ 1.362,6), Alemania (US\$1.263,7),Francia (US\$ 1.202,2), Canadá (US\$ 1.197,3), EE.UU. (US\$ 1.172,1), Inglaterra (US\$1.159,0), México (US\$1.154,5) y Brasil (US\$ 1.014,6), quienes en su totalidad superaron el gasto total individual de los turistas (US\$ 596,6), y en su mayoría lo doblan.

Si bien los países limítrofes generan el mayor número de llegadas de turistas a Chile, en la práctica son los que menos gasto realizan. La contribución de los distintos mercados en la generación de ingreso de divisas (excluyendo los valores por transporte aéreo internacional y por llegada de excursionistas) en Chile es la siguiente: Europa (25,0%), Países Limítrofes (27,6%), Resto de América (27,1%), América del Norte (14,0%), Resto del Mundo (3,6%) y Asia (2,7%).

3.2 Conjunto de Datos

Este trabajo de investigación se basó en los datos publicados por el Instituto Nacional de Estadísticas, en relación a la llegada total mensual de turistas extranjeros al país desde Enero de 1997 hasta enero de 2015, es decir, se cuenta con 18 años de datos, lo que se traduce en 217 meses registrados. Estos datos serán utilizados como entrada en el modelo SARIMA. En la figura 3 se puede ver de manera gráfica el comportamiento de los datos.



Figura 3: Extranjeros que han Arribado a Chile

Algunas conclusiones que se pueden sacar a priori de esta gráfica:

- a) Se ve un gran descenso en la llegada de turistas a Chile en el año 2001, esto es producto al atentado a las torres gemelas en Septiembre de ese año.
- b) En Enero del año 2007 fue posible superar la cantidad máxima registrada antes del atentado, es decir, 6 años después de éste.
- c) La tendencia desde el año 2001 ha sido de aumento en la cantidad de turistas llegados al país.
- d) Se puede observar que la tendencia anteriormente escrita se ve interrumpida en cierto grado el año 2010, este se infiere que es causado por el terremoto registrado el 27 de Febrero de ese año.

Como se puede observar, la llegada de turistas está altamente influenciada por las catástrofes naturales y de otro tipo de catástrofe, como terrorismo, lo que ha provocado una disminución de la totalidad de turistas arribados al país.

La data será estudiada más a fondo en el siguiente capítulo, en donde se expondrá el modelo de predicción SARIMA.

3.3 Detección de estacionalidad

El primero fue el gráfico de función de auto-correlación (FAC), el cual es utilizado para observar la correlación que existe un valor a ciertos periodos de distancia con una variable de la misma serie.

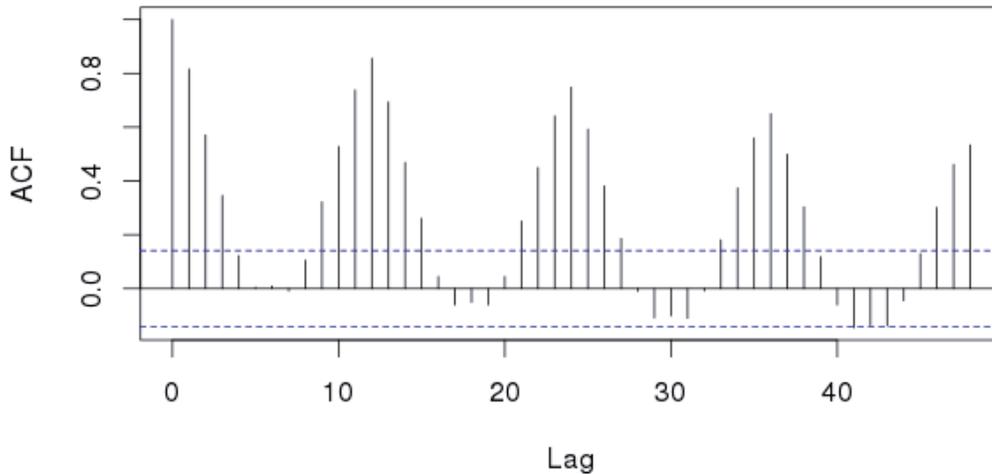


Figura 4: Función de Auto-Correlación a la Serie Original

Como se puede observar de la figura 4, existe una gran correlación entre los datos con una diferencia de 12, 24, 36 y 48. Esto representa un ciclo anual.

Otra manera de distinguir que era una serie estacional fue calculando el espectro de a serie a través de un periodograma suavizado mediante el método denominado suavizado modificado Daniell. Este suavizado lo que realiza es un promedio de móvil, otro tipo de promedio móvil al visto anteriormente. Este tipo de suavizado le otorga a los valores extremos menor relevancia en el resultado que se obtendrá, y a los valores más cercanos al periodo le entrega una mayor relevancia. En el siguiente periodograma se utilizó una suavización de orden (2,2), es decir, a la serie se realiza una doble suavización de orden 2. Una suavización individual de dicho orden genera la siguiente salida:

$$\begin{aligned}\hat{x} &= \frac{x_{t-2} + 2x_{t-1} + 2x_t + 2x_{t+1} + 2x_{t+2}}{8} \\ &= 0.125x_{t-2} + 0.25x_{t-1} + 0.25x_t + 0.25x_{t+1} + 0.125x_{t+2}\end{aligned}$$

El periodograma lo que realiza es transformar la serie desde el dominio del tiempo al dominio de la frecuencia a través de una transformada rápida de Fourier, y entrega como resultado la incidencia de dichas frecuencias en la varianza de la serie, los peaks que se registran en el periodograma corresponde a frecuencia que son relevantes en la serie. Como se cuenta con la frecuencia es sencillo obtener el periodo (3.3.1).

$$t = \frac{1}{f} \quad (3.3.1)$$

donde t es el periodo y f la frecuencia.

En el siguiente periodograma representa en el eje de las abscisas el periodo, en la data un periodo es sinónimo de 12 meses.

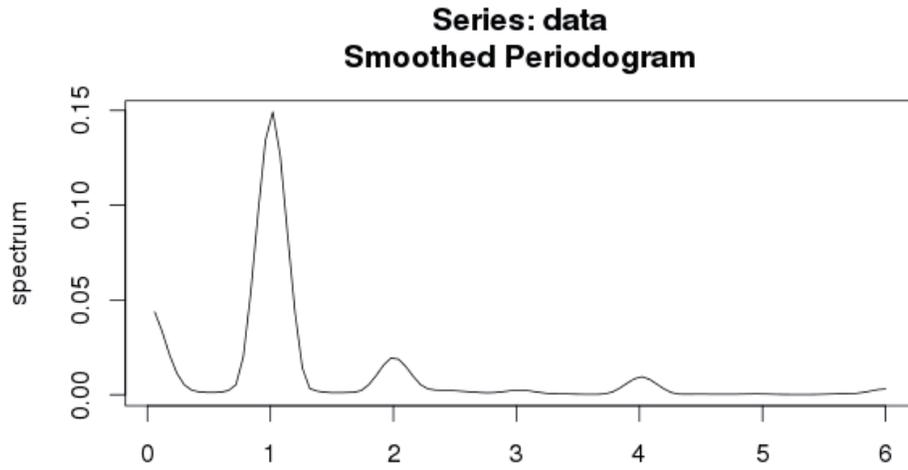


Figura 5: Periodograma Suavizado Aplicado a la Serie

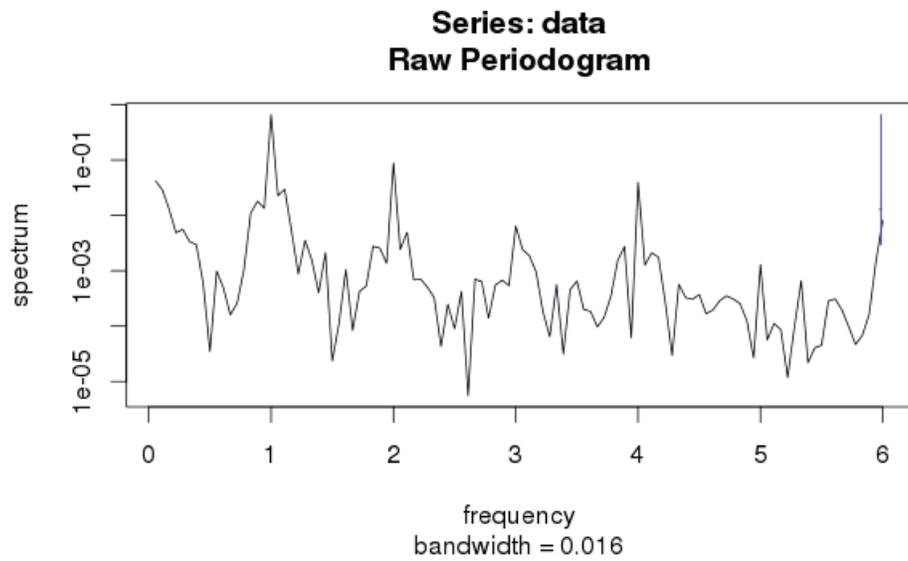


Figura 6: Periodograma sin suavizar

De la figura 5 y figura 6 es posible sacar como conclusión que la serie presenta una gran estacionalidad de grado 12.

3.4 Detección de tendencia

A continuación de haber encontrado que existe una fuerte estacionalidad en la serie, se centralizó en la hipótesis que la serie contaba con una fuerte tendencia (3.4.2). Para aquello se graficó la tendencia de la serie.

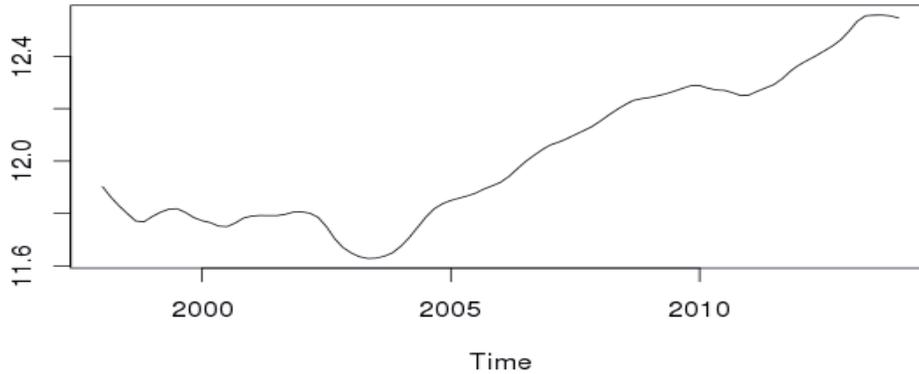


Figura 7: Tendencia de la Serie Original

La tendencia se obtuvo de la siguiente manera:

$$x'_t = \frac{x_t - x_{t-1}}{x_t}, \quad (3.4.2)$$

desde la figura 7 se puede distinguir que la serie tiene una tendencia bastante similar a una lineal.

Y finalmente a la serie de tiempo se le aplicó una conversión logarítmica con el fin de disminuir la varianza de la serie lo que provoca en la general una mejor predicción.

4. Resolución del Caso Mediante un Modelo SARIMA

Los modelos SARIMA son utilizadas principalmente y casi únicamente para predecir valores futuros de series de tiempo que no son estacionarias y que cuentan con una clara estacionalidad en sus datos. Para las series que no son estacionarias y no contienen estacionalidad, es decir, solo tendencia se utiliza un modelo ARIMA, y para las series que no cuentan con tendencia ni estacionalidad, lo que produce que sea una serie estacionaria se utiliza los modelos ARMA.

Los modelos ARMA están constituidos por dos partes:

- a) **Modelo Auto-regresivo:** En los modelos auto-regresivo (AR) se busca obtener una predicción de la variable de interés utilizando una combinación lineal de valores pasados de la variable. El término auto-regresivo indica que es una regresión de la variable entre sí misma.

Un modelo auto-regresivo de orden p puede ser escrito de la siguiente manera:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \quad (4.1)$$

donde c (4.1) es una constante y e_t es ruido blanco. Esto es como una regresión múltiple pero con valores diferenciados de y_t como predictores. Se refiere esto como un modelo AR (p).

- b) **Modelo de promedios móviles:** En vez de utilizar valores pasados para predecir una variable en una regresión, un modelo de promedio móvil utiliza errores pasados de predicciones en un modelo regresivo. Esto puede ser expresado de la siguiente manera:

$$y_t = c + e_t + \phi_1 e_{t-1} + \phi_2 e_{t-2} + \dots + \phi_q e_{t-q} \quad (4.2)$$

donde e_t (4.2) es ruido blanco. Se refiere a esto como un modelo MA (q).

4.1 ARIMA

Si se combina diferenciación con auto-regresión y un modelo de promedios móviles, se obtiene un modelo no estacional llamado ARIMA (4.1.1). Que sea un modelo integrado significa que se ha realizado una diferenciación de sus valores. El modelo completo puede ser descrito de la siguiente manera:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \phi_1 e_{t-1} + \dots + \phi_q e_{t-q} + e_t \quad (4.11)$$

donde y'_t es la serie diferenciada (ésta puede ser diferenciada más de una vez). Los “predictores” al lado derecho de la ecuación incluyen ambos valores de desfase de y_t y errores desfasados. Esto se denomina un modelo ARIMA (p, d, q) donde:

p = orden de la parte auto-regresiva

d = grado de la primera diferenciación involucrada

q = orden de la parte de promedio móvil

4.2 SARIMA

Anteriormente se han descrito utilizados para series sin estacionalidad. Los modelos ARIMA también son capaces de modelar una gran variedad de datos. Un ARIMA estacional, SARIMA (4.2.1), está formado por una parte estacional. Este modelo puede ser descrito de la siguiente manera:

$$ARIMA (p, d, q)(P, D, Q)_m \quad (4.2.1)$$

En donde los valores del primer paréntesis a la izquierda indican los parámetros no estacionales del modelo, mientras que los valores del segundo paréntesis indican los parámetros estacionales del modelo. Donde la m es el número de periodos por estación.

4.3 Aplicación del modelo SARIMA

Se decide utilizar un modelo SARIMA (4.2.1) como primera aproximación a una predicción de la cantidad de turistas extranjeros que arribarán a Chile en los próximos meses. La utilización de un modelo SARIMA y no ARIMA se decidió en base a que se distinguió con claridad que la serie de tiempo cuenta con una estacionalidad marcada, cuya estación tiene una duración de 12 meses. Esta estacionalidad es fácilmente distinguida a primera vista observando el grafico de la serie, pero la decisión final fue en base a dos indicadores que estudiaron los datos originales.

Se utilizó como datos de entrenamiento del modelo la data correspondiente hasta la fecha 14 de Enero de 2014, dejando la data restante para corroborar la precisión del modelo en la fase de prueba. Además se le aplicó una transformación logarítmica a aquellos datos para disminuir la varianza de la serie.

4.4 Detección parámetros del modelo SARIMA con mejor ajuste

Para la detección del modelo que mejor se ajustaba a la serie, se utilizó un algoritmo el cual iteraba hasta encontrar el modelo con lo que se obtenía un menor AIC en la etapa de prueba. Por lo tanto, se ejecutó el siguiente algoritmo:

```
For p in 0:3
  For d in 0:1
    For q in 0:3
      For P in 0:3
        For D in 0:1
          For Q in 0:3
            Forecast <- sarima (Testing, p, d, q, P, D, Q, 12))
            If (Forecast$AIC < AIC) then
              pb <- p
              db <- d
              qb <- q
              Pb <- P
              Db <-D
              Qb <- Q
              AIC <- Forecast$AIC
            End if
          End For
        End For
      End For
    End For
  End For
End For
```

Al término de esta iteración se obtuvo que el modelo que representa un mejor ajuste a la serie, entregando el menor AIC, es el siguiente:

$$(1,0,2)(0,1,1)_{12}$$

Por lo tanto, el modelo con el que se realiza la predicción puede ser representado de la siguiente manera utilizando el operador Backshift:

$$(1 - \phi_1 B)(1 - B^{12})Ln(y_t) = (1 + \omega_1 B + \omega_2 B^2)(1 + \theta_1 B^{12})e_t \quad (4.4.1)$$

Se puede apreciar que solo existe diferenciación en la parte estacional de la serie (4.4.1), la que puede ser observada en la siguiente figura (Figura 8)

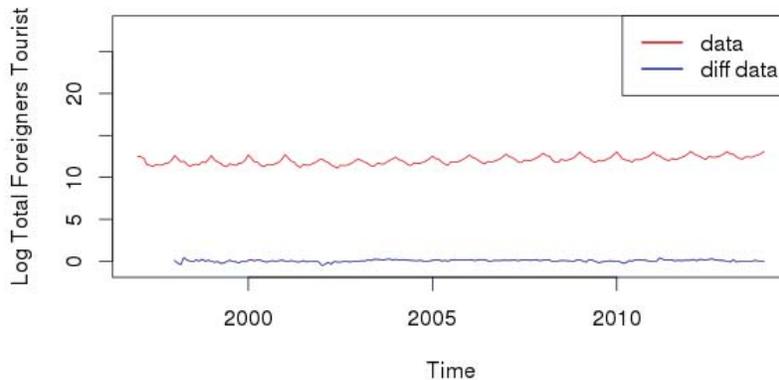


Figura 8: Diferenciación de los datos en primer grado

Por lo tanto, resulta de alta relevancia la información regular de la serie para la obtención de un modelo con un buen ajuste a los datos, no así, la estacionalidad de los datos la cual fue eliminada al realizar la diferenciación estacional.

4.5 Estudios de los residuos del modelo

Una parte primordial en los modelos SARIMA es el estudio de los residuos, un modelo bien ajustado como condición debe entregar residuos que se comporten como ruido blanco, es decir, que estos no guarden correlación estadística, de manera que los residuos no albergue información relevante de la serie. Esto es comprobable verificando la función de autocorrelación y la función parcial de autocorrelación de los residuos.

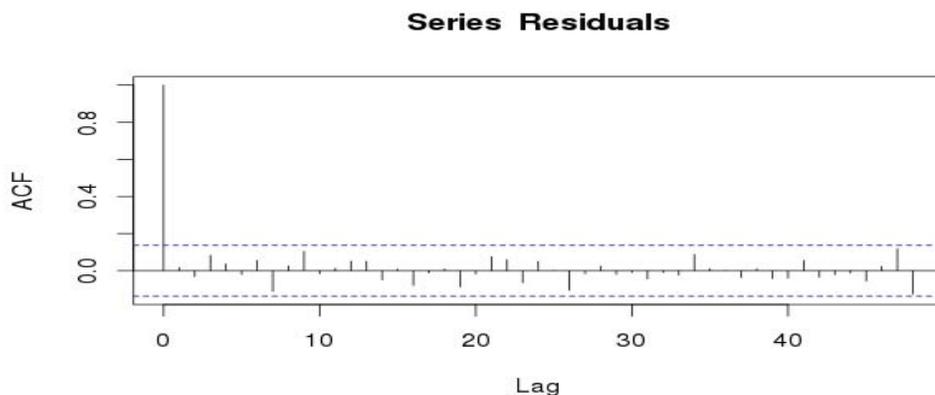


Figura 9: ACF Residuos

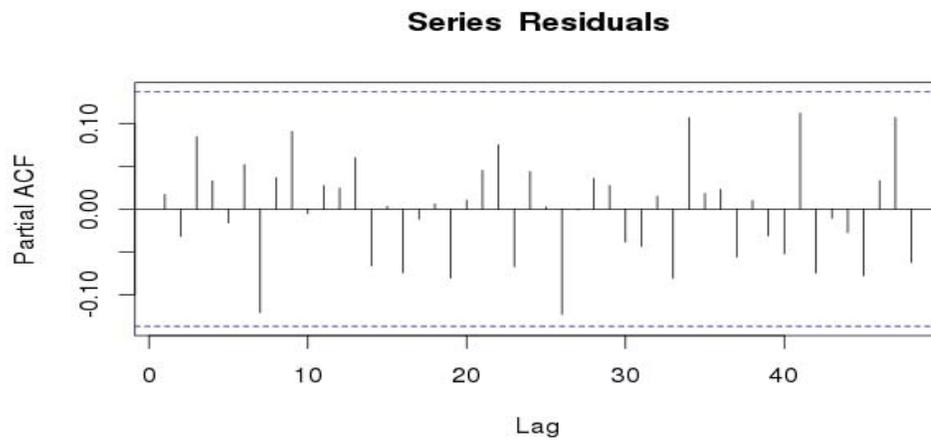


Figura 10: PACF Residuos

Es posible observar que no existe auto correlación alguna entre los diferentes desfases tanto en la acf como en la pacf, a una significancia de 5% y con retraso de hasta 48 periodos.

Otro punto relevante es que los residuos tengan una distribución normal, esto es comprobable mediante la gráfica Cuantil-Cuantil y mediante el histograma.

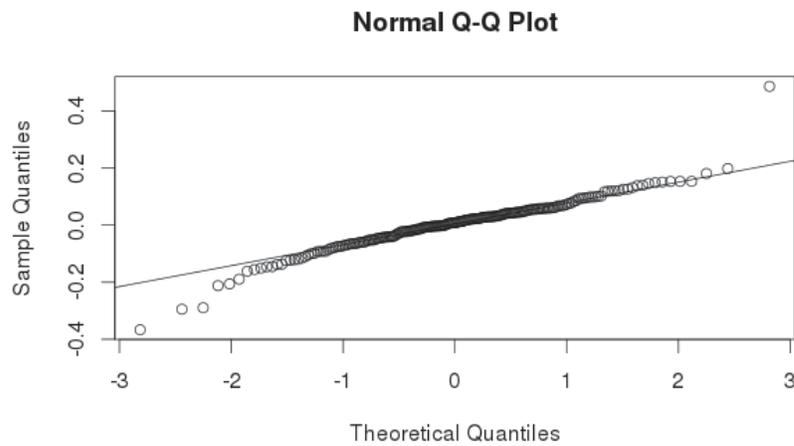


Figura 11: Gráfico Cuantil-Cuantil (QQ) de los residuos

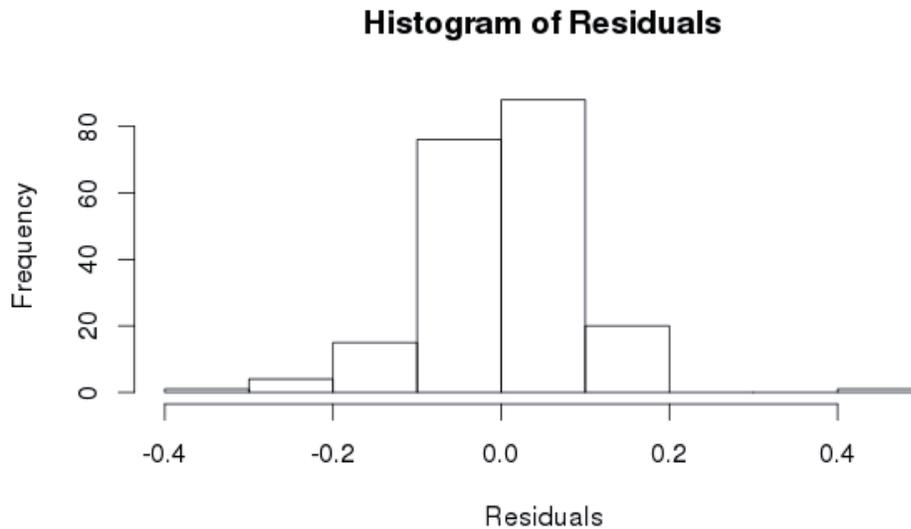


Figura 12: Histograma de los residuos

A través de las Figura 11 y figura 12 es posible detectar la distribución normal de los residuos del modelo, ya que, en el gráfico QQ los datos recaen sobre la línea diagonal y en el histograma se puede apreciar que la mayor frecuencia de los datos ocurren en el centro de la distribución, formando una apariencia cercana a la campana de Gauss.

Debido a lo expuesto anteriormente es posible concluir que los residuos del modelo se comportan como ruido blanco y están normalmente distribuidos.

4.6 Salidas del modelo

El modelo escogido entregó como resultado los siguientes resultados:

$$AIC = -3.61$$

$$AICc = -3.60$$

Como se describe en el Anexo, la diferencia entre el AIC y el AICc no es sustancial debido a que el conjunto de datos es de una extensión mayor a 200 y tan solo con una cantidad de menor a 20 datos la diferencia entre ambos es relevante.

Los parámetros de salida del modelo son los siguientes:

Tabla 1: Factores del modelo

ar1	ma1	ma2	sma1	constante
0.9775	-0.4794	-0.2283	-0.7435	0.0037

4.7 Predicción

Se utilizó para la fase de prueba el último año de datos disponibles, es decir, desde Febrero-2014 a Enero-2015, el resto de los meses hasta Enero-1997 se utilizó en la etapa de entrenamiento. Se transformó la data al dominio inicial de la siguiente manera:

$$\hat{Y}_t = e^{y_t}$$

El modelo entregó los siguientes resultados:

Tabla 2: Resultados de Predicción

Date	Test	Forecast
Feb-14	354125	379205
Mar-14	333977	334985
Abr-14	254063	266000
May-14	228542	212961
Jun-14	193004	191990
Jul-14	270426	256492
Ago-14	251258	241585
Sep-14	244717	242519
Oct-14	302173	284769
Nov-14	312487	318837
Dic-14	437055	403523
Ene-15	560397	549417

Una representación gráfica de la tabla 2:

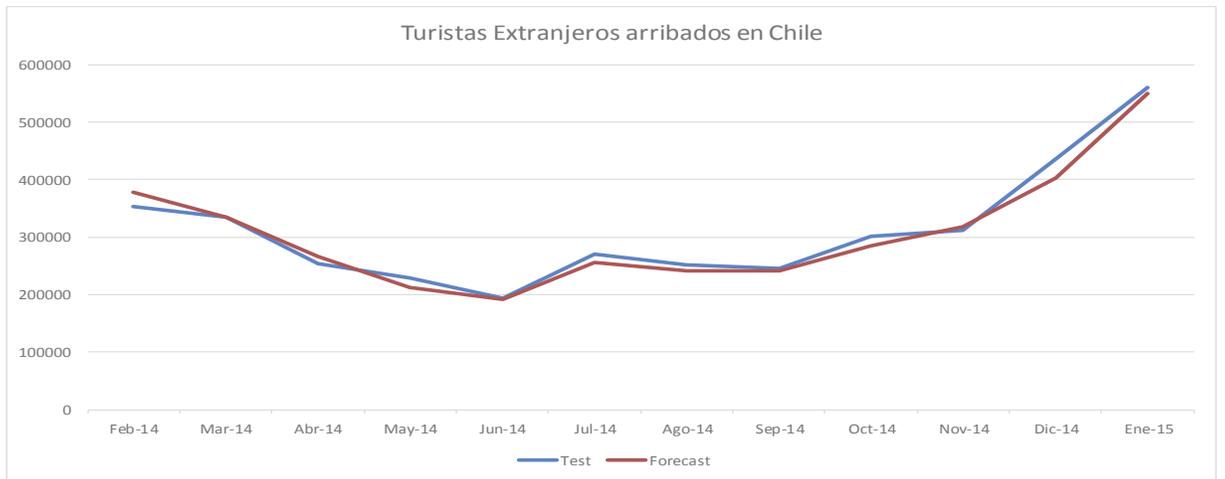


Figura 13: Predicción versus Valores Reales

Es visible una gran precisión en los datos obtenidos desde el modelo, estadísticamente se puede obtener una evaluación del modelo.

$$RMSE = 15537.76$$

$$NSE = 0.97354$$

$$R = 0.9885$$

Estos valores indican que los resultados predichos se ajustan a la realidad.

5. Resolución del Caso Mediante una Red Neuronal Autoregresiva

5.1 Red Neuronal

Las Redes Neuronales Artificiales son sistemas computacionales que intentan homologar el cerebro humano, especialmente en el proceso de aprendizaje. Este aprendizaje es realizado entregándole a la red los datos de entrada o entrenamiento, prueba y esperados, la red empezará a iterar hasta encontrar la configuración que encuentre la combinación con mejor aproximación, es decir, la combinación que obtenga un menor error. A la Red a continuación se le entrega los datos de validación, con los cuales se podrá medir el rendimiento de la red obtenida.

La estructura de una Red Neuronal es la siguiente

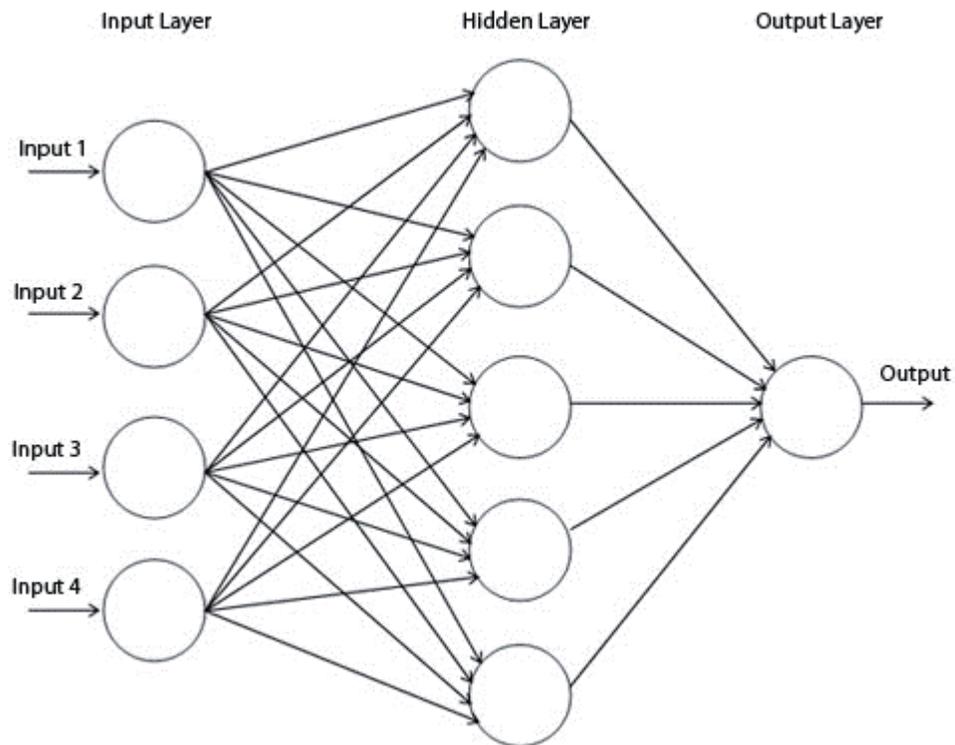


Figura 14: Estructura Red Neuronal

- a) **Capa de Entrada:** En la capa de entrada es por donde ingresan los datos a la red neuronal, es la capa encargada de alimentar al modelo.
- b) **Capa Oculta:** Esta capa recibe la ponderación de los datos de entradas con el peso especificado, a dicha ponderación se le aplica una función de activación, en este caso se utilizó la función de activación sigmoide por su amplia utilización en investigaciones.
- c) **Capa de Salida:** La última capa recibe los valores salidos de la capa oculta ponderado por los pesos, esta función lineal permite obtener finalmente el dato de salida.

Por lo tanto, la Red generará una salida de la siguiente manera:

$$y = w^{sc} + \sum_{j=1}^m w_{1,j}^s \cdot F(w_j^{oc} + \sum_{i=1}^n w_{j,i}^o \cdot x_i) \quad (5.1.1)$$

Donde:

w_j^{oc} : Ponderador de la conexión con entrada de valor unitario y la neurona j de la capa oculta.

w^{sc} : Ponderador de la conexión con entrada de valor unitario y la neurona de capa de salida

$w_{j,i}^o$: Ponderador de la conexión entre la neurona i de la capa de entrada y la neurona j de la capa oculta

$w_{1,j}^s$: Ponderador de la conexión entre la neurona j de la capa oculta y la neurona 1 de la capa de salida.

5.2 Redes Neuronales Propuestas

Para esta tesis se ha decidido utilizar una Red Neuronal Autoregresiva (5.1.1) de mismo orden de estacionalidad utilizado en la modelo SARIMA, es decir, de orden 12, además se iteró desde el orden 13 al doce 24 para así encontrar la configuración con mejor desempeño. La capa de entrada tendrá la misma cantidad de neuronas que el orden de latencia utilizada. Todas las redes propuestas tendrán como cantidad de neuronas en la capa oculta igual a la aproximación al entero más cercano del logaritmo de la cantidad de datos de la muestra; dicha operación nos da un valor de 5 neuronas en la capa oculta.

Además se ha escogido la función sigmoidal como la función de activación en la capa oculta y el algoritmo de Levenberg–Marquardt será utilizado para entrenar a la red con el fin de encontrar los pesos y ponderados. El 85% de la red de los datos son utilizados en la etapa de entrenamiento y el 15% restante en la etapa de prueba.

5.3 Algoritmo de entrenamiento Levenberg–Marquardt

Como fue mencionado anteriormente el algoritmo utilizado para entrenar a la red será el algoritmo llamado Levenberg–Marquardt. A continuación se explicará su funcionamiento.

Este algoritmo tiene como finalidad encontrar la función que mejor se aproxime a los datos, que tenga mejor “ajuste” (5.3.1), de acuerdo con el criterio de los mínimos cuadrados (5.3.2). Dicho de otra manera, lo que intenta es minimizar la suma de las diferencias cuadrada de los residuos. Los residuos son obtenidos de la siguiente manera:

$$r_m = y_m - f(t_m, \theta) \quad (5.3.1)$$

Y la función de mínimos cuadrados es la siguiente:

$$C = \sum_m r_m(\theta)^2 \quad (5.3.2)$$

Por lo tanto, el algoritmo encuentra los valores de $\theta_1, \theta_2, \theta_3, \dots, \theta_m$ tal que C es minimizada. Para aquello existen varios algoritmos, entre ellos está los algoritmos de gradiente descendente, Gauss-Newton y el algoritmo de Levenberg–Marquardt.

5.3.1 Algoritmo de Gradiente Descendente

El algoritmo comienza con un x_0 inicial escogido de manera aleatoria, la ecuación de actualización es la siguiente:

$$x_{n+1} = x_n - \gamma_n \nabla F(x_n), n \geq 0 \quad (5.3.1.1)$$

La ventaja de utilizar este algoritmo es su rápido descenso de la función hacía el mínimo en cada iteración (5.3.1.1). La desventaja de este algoritmo es que al encontrarse cerca del mínimo comienza a zigzaguear y puede tomar bastante tiempo encontrar el mínimo absoluto.

5.3.2 Algoritmo de Gauss-Newton

El algoritmo se inicializa con un x_0 aleatorio. Se ajusta x por Δ en cada iteración, donde Δ está dado resolviendo:

$$(J_r^T J_r) \Delta = J_r^T r \quad (5.3.2.1)$$

La ventaja de utilizar este algoritmo (5.3.2.1) es su rápida convergencia hacía el mínimo absoluto cuando los valores se encuentran cercanos a éste. La desventaja que presenta este algoritmo es que si los valores iniciales escogidos de manera aleatoria están lejanos al mínimo el algoritmo convergerá lentamente o simplemente no convergerá.

5.3.3 Algoritmo de Levenberg–Marquardt

El algoritmo LM combina las ventajas del Gradiente Descendiente (5.3.1.1) y el de Gauss-Newton (5.3.2.1) basado en reglas adaptivas. Empieza con un x_0 inicial escogido de manera aleatoria, x es ajustado por δ solo para iteraciones descendentes.

$$(J^T J + \lambda I) \delta = J^T r \quad (5.3.3.1)$$

J = Matriz de derivadas de los residuos con respecto a los parámetros.

λ = Parámetro de amortiguación

r = vector de residuos

El algoritmo de iteración (5.3.3.1) es el siguiente:

- 1) Inicializar los valores de los parámetros, x , el parámetro Levenberg-Marquardt λ , también los parámetros λ_{down} y λ_{up} utilizados para ajustar el término de amortiguación. Evaluar los residuos r y el Jacobino J en la estimación inicial de los parámetros.
- 2) Calcula la métrica, $g = J^T J + \lambda I$, y el gradiente del costo, $\nabla C = J^T r$, $C = \frac{1}{2} r^2$
- 3) Evaluar los nuevos residuos r_{new} en el punto dado por $x_{new} = x - g^{-1} \nabla C$, y calcular el costo en el nuevo punto, $C_{new} = \frac{1}{2} r_{new}^2$
- 4) Si $C_{new} < C$, acepta la iteración, $x = x_{new}$, y $r = r_{new}$ y $\lambda = \frac{\lambda}{\lambda_{down}}$. De otra manera, rechaza la iteración, mantiene los parámetros de estimación anteriores x , los residuos anteriores r se mantienen y se ajusta $\lambda = \lambda \times \lambda_{up}$
- 5) Chequear convergencia. Si el método ha convergido, retorna x como los parámetros con mejor ajuste. Si el método aún no ha convergido pero la iteración ha sido aceptada, se evalúa el Jacobino J en los nuevos valores de los parámetros. Ir al paso dos.

El algoritmo se detiene cuando alcanza el máximo de iteraciones, o los valores de la función objetivo están por debajo de un valor de corte, o el gradiente del costo es pequeño, o el algoritmo retorna nulos.

5.4 Predicción realizada mediante Red Neuronal utilizando doce Neuronas en la capa de entrada

La primera red a utilizar para la predicción de la cantidad de turistas extranjeros que arribarán a Chile es una red con doce neuronas en la capa de entrada, cada neurona representa un tiempo pasado de la serie temporal, por lo tanto, a la red ingresa un año de datos (doce meses). Esta red gráficamente se puede observar de la siguiente manera:

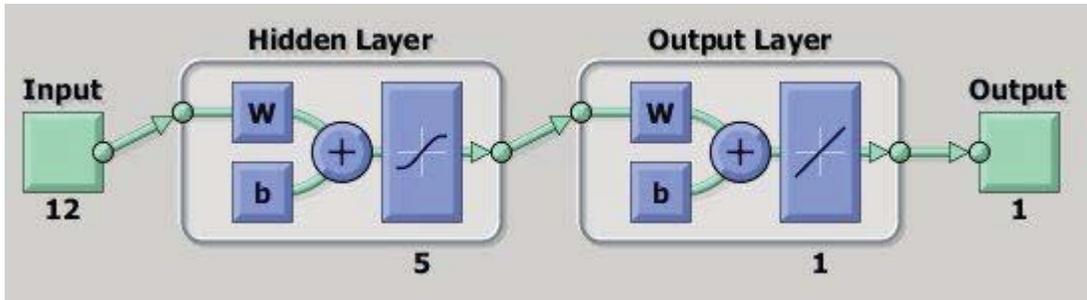


Figura 15: Red Neuronal 12 Neuronas Entrada

5.4.1 Salida Red Neuronal

La red dio como resultado utilizando el 85% de los datos en la etapa de entrenamiento y con 150 iteraciones, escogiendo la configuración con menor MSE, lo siguiente:

Tabla 3: Resultados de Predicción utilizando RNA 12 entradas

Date	Observed	Estimated
Feb-14	354125	480701
Mar-14	333977	321998
Abr-14	254063	281644
May-14	228542	228571
Jun-14	193004	203760
Jul-14	270426	273950
Ago-14	251258	248824
Sep-14	244717	262577
Oct-14	302173	294844
Nov-14	312487	322578
Dic-14	437055	389415
Ene-15	560397	539339

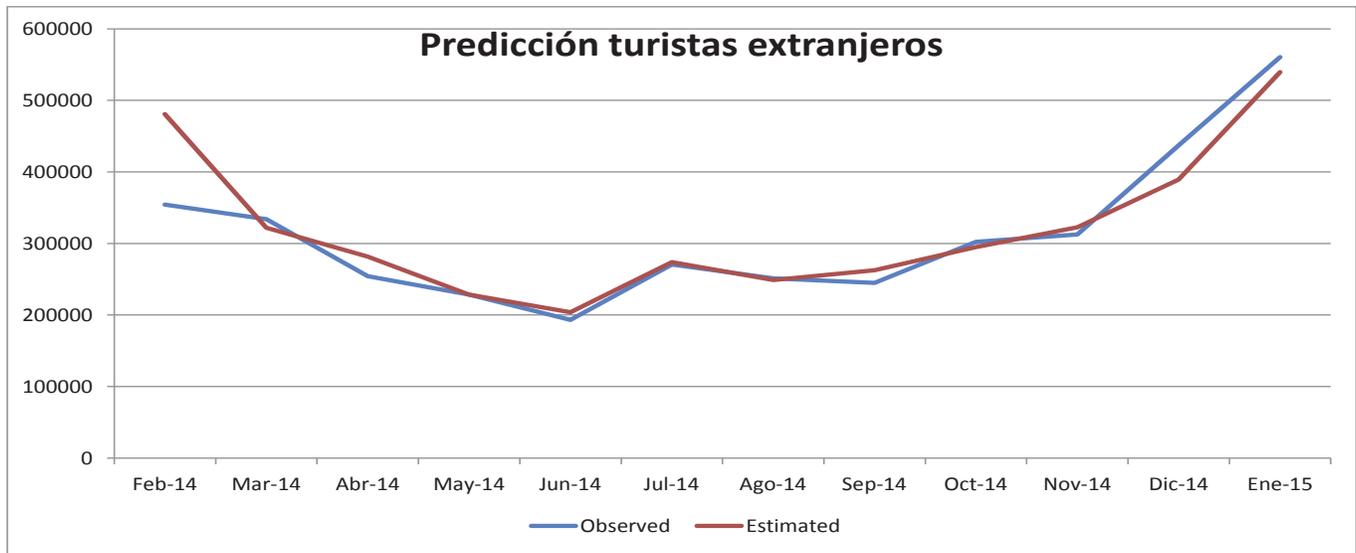


Figura 16: Datos Predichos utilizando RNA

Estadísticamente podemos medir los resultados y los errores obtenidos mediante las siguientes medidas que indican el nivel de poder de predicción de un modelo.

$$RMSE: 41709.84$$

$$NSE: 0.82258$$

$$R: 0.9151$$

5.4.2 Estudios de los Residuos

Los residuos de la red no deben estar correlacionados y deben tener una distribución normal, ya que, en caso contrario se podría concluir que la red debe ser mejorado porque hay información relevante que el modelo no es capaz de incorporar. Por éste motivo es tan importante estudiar los residuos del modelo, para aquello se presentará la autocorrelación a distintos desfases y el histograma de los errores.

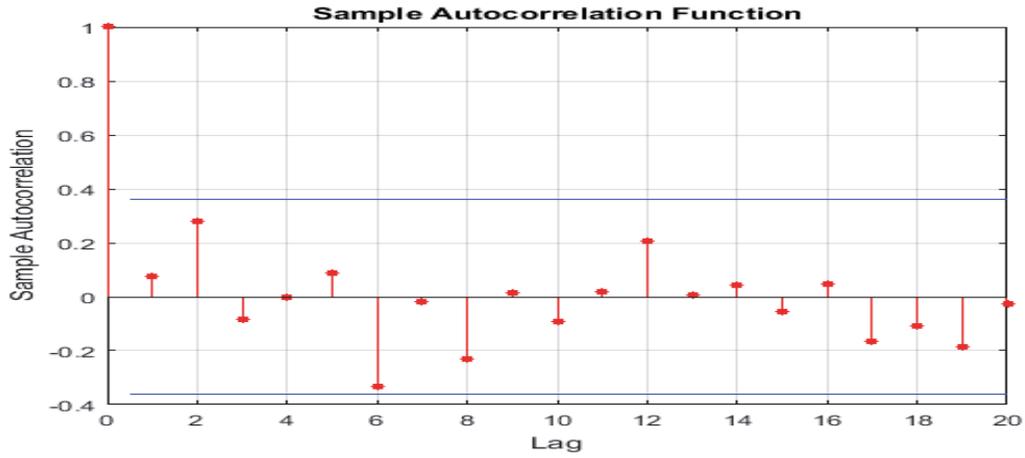


Figura 17: ACF RNA Orden 12

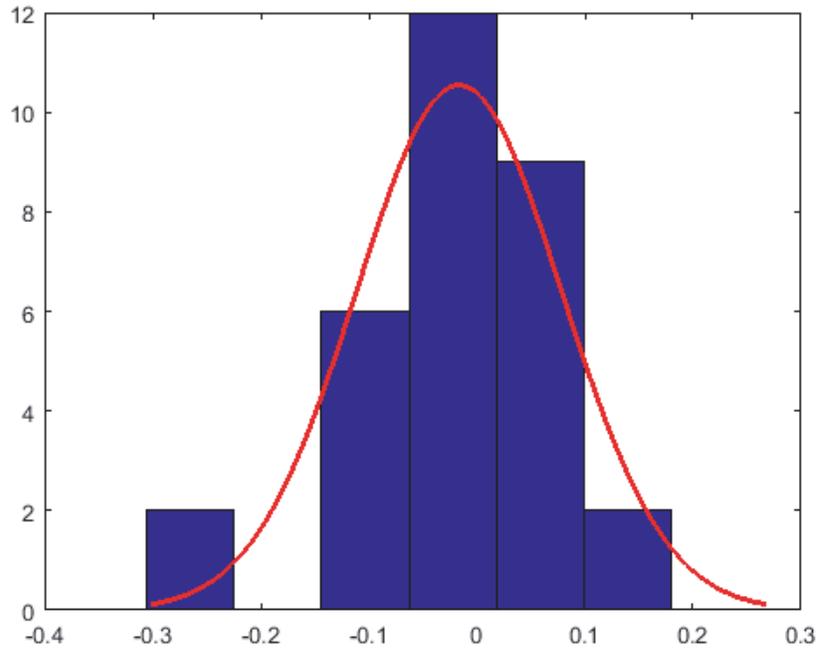


Figura 18: Histograma RNA Orden 12

Es posible observar que los residuos no tienen correlación alguna entre ellos a un 5% de confianza y que los residuos siguen una distribución normal según el histograma. Por lo que se puede concluir que esta red puede ser aceptada al igual que sus resultados, sin necesidad de cambios.

5.5 Predicción realizada mediante Red Neuronal variando las Neuronas en la capa de entrada desde 13 hasta 24 neuronas.

La siguiente etapa del proyecto fue encontrar la configuración que mejor ajuste tenía sobre el conjunto de datos variando la cantidad de neuronas desde 13 hasta 24. Para después contrarrestarla con la red con 12 neuronas de entrada, y así observar cuál de las dos es la que tiene mejor desempeño. La red escogida será aquella que presente el menor MSE (mean squared error) en la iteración.

Todas las redes en prueba utilizaron el 85% de los datos en la etapa de entrenamiento y utilizaron 5 neuronas en la capa oculta con la función de activación sigmoidal y 150 iteraciones fueron ejecutadas para cada red. Los resultados fueron los siguientes:

Figura 19 Resumen MSE

Neurons	13	14	15	16	17	18	19	20	21	22	23	24
MSE training	0.0054	0.0106	0.0112	0.0078	0.0063	0.0087	0.0120	0.0099	0.0154	0.0226	0.0161	0.0103

Por lo tanto, la mejor configuración la presentó la red con 13 neuronas en la capa entrada.

5.5.1 Predicción utilizando una Red Neuronal Autoregresiva de orden 13

Se obtuvo en el capítulo anterior que la red neuronal de orden 13 de autoregresión fue la que presentó el mejor ajuste al conjunto de datos, a menor MSE mejor es el ajuste.

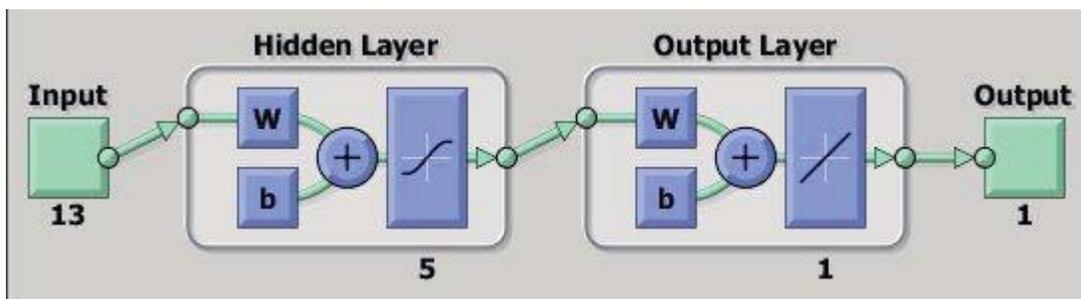


Figura 19: Estructura RNA Orden 13

5.5.2 Salida Red Neuronal

La red entregó como salida los siguientes resultados al término de la etapa de entrenamiento de la red.

Tabla 4: Datos Observados y Estimados por RNA Orden 13

Date	Observed	Estimated
Feb-14	354125	383234
Mar-14	333977	287650
Abr-14	254063	267855
May-14	228542	228914
Jun-14	193004	193165
Jul-14	270426	265774
Ago-14	251258	241277
Sep-14	244717	251249
Oct-14	302173	278786
Nov-14	312487	324292
Dic-14	437055	343589
Ene-15	560397	527603

Gráficamente se puede observar de la siguiente manera:

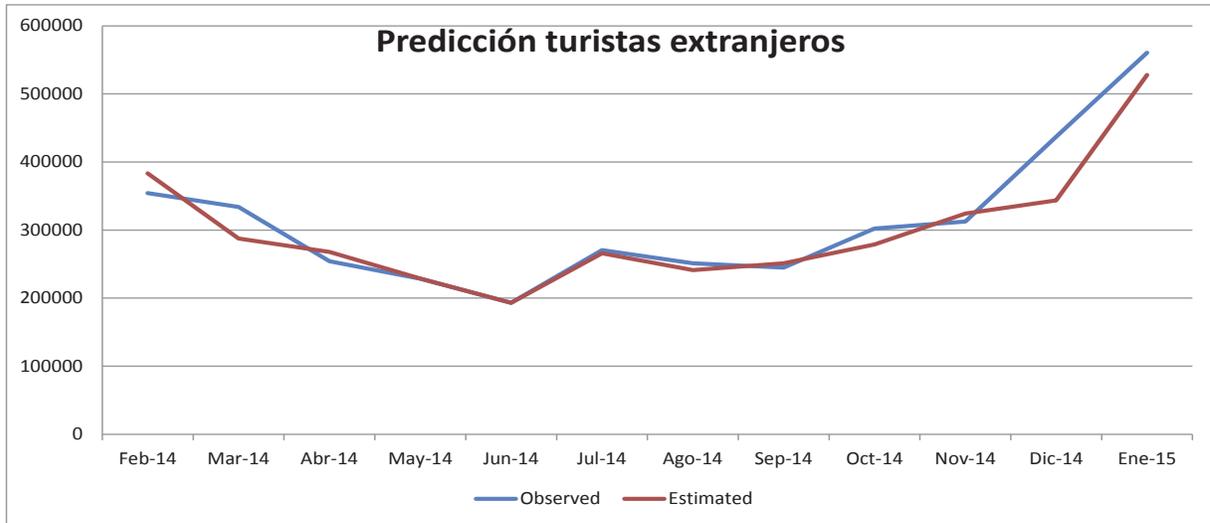


Figura 20: Datos Predichos Utilizando RNA Orden 13

Y empleando las mismas herramientas estadísticas de los modelos anteriores obtenemos:

$RMSE: 33966.31$

$NSE: 0.87867$

$R: 0.9493$

5.5.3 Estudios de los Residuos

Se realizará al igual que en los modelos anteriores un estudio sobre los residuos para descartar la existencia de información relevante en los residuos del modelo implementado. Para aquello nos apoyaremos en los gráficos de autocorrelación e histograma.

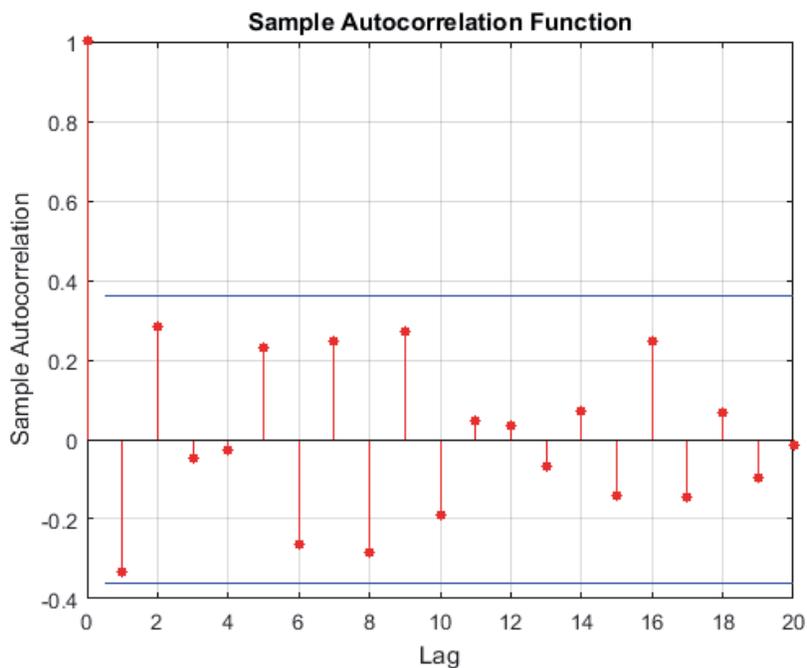


Figura 21: ACF RNA Orden 13

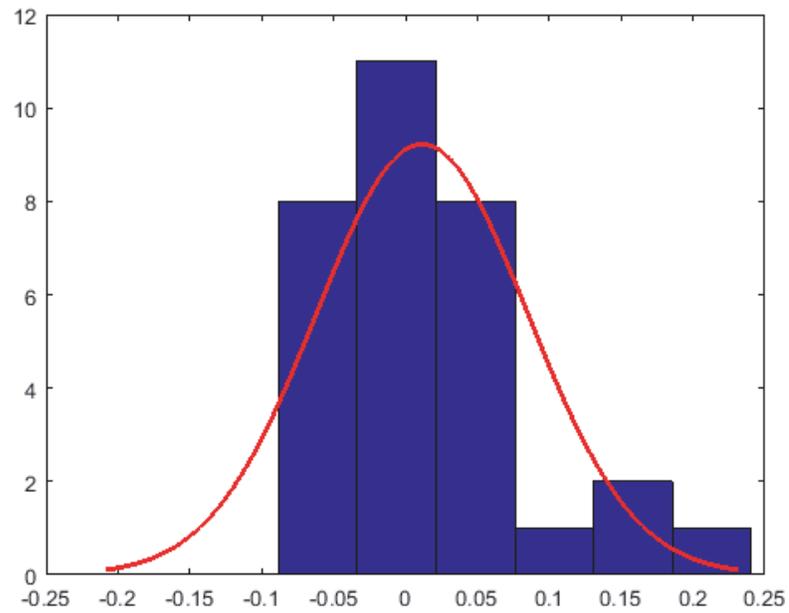


Figura 22: Histograma RNA Orden 13

De acuerdo a las dos gráficas presentadas se puede concluir que el modelo no guarda información relevante en los residuos, porque no hay autocorrelación entre los residuos a distintas latencias, y además éstos se comportan de manera normal según el histograma de los residuos.

6. Comparativa entre los modelos SARIMA y Redes Neuronales Artificiales Autoregresiva

Se puede apreciar que el modelo SARIMA realiza un mejor modelado de la problemática, asimilando de mejor manera la estacionalidad de la serie de tiempo y dando como resultado una salida con mayor precisión. Las redes neuronales obtuvieron un buen rendimiento, obteniendo una salida de los modelos bastante aceptable.

Tabla 5: Resultados Estadísticos

	SARIMA	RNA(12)	RNA(13)
RMSE	15537,76	41709,84	33966,31
NSE	0,97354	0,82258	0,87867
R	0,9885	0,9151	0,9493

Se puede observar mediante el gráfico comparativo entre la salida del modelo SARIMA, Redes Neuronales Artificiales Autoregresiva y los datos observados, que el modelo SARIMA obtiene un mejor desempeño que la Red Neuronal.

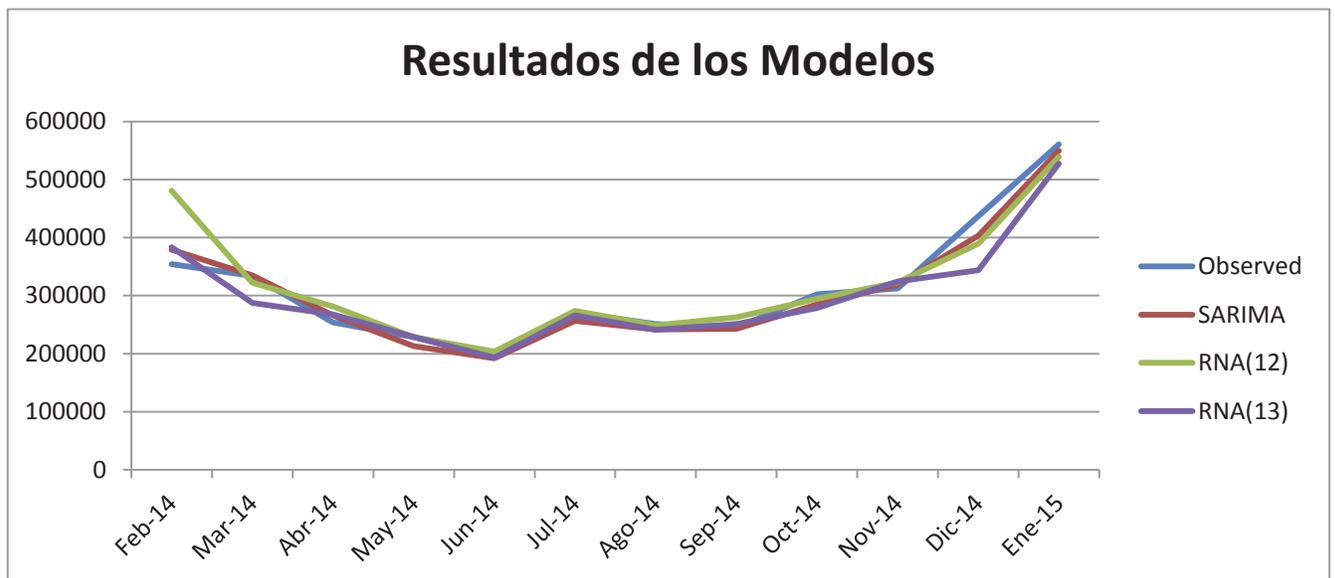


Figura 23: Comparativa datos Predichos

7. Predicción de turistas de los 12 periodos próximos a la serie de tiempo en estudio

Para finalizar este trabajo se realizó la aplicación de los modelos matemáticos obtenidos en las etapas previas del estudio, con el fin de obtener la cantidad mensual de turistas extranjeros que arribarán a Chile dentro de los próximos 12 meses ajenos a la serie de tiempo en estudio, es decir, desde Febrero de 2015 hasta Enero 2016. Datos no observados al momento de comenzar este estudio. El resultado es mostrado en la siguiente figura comparativa.

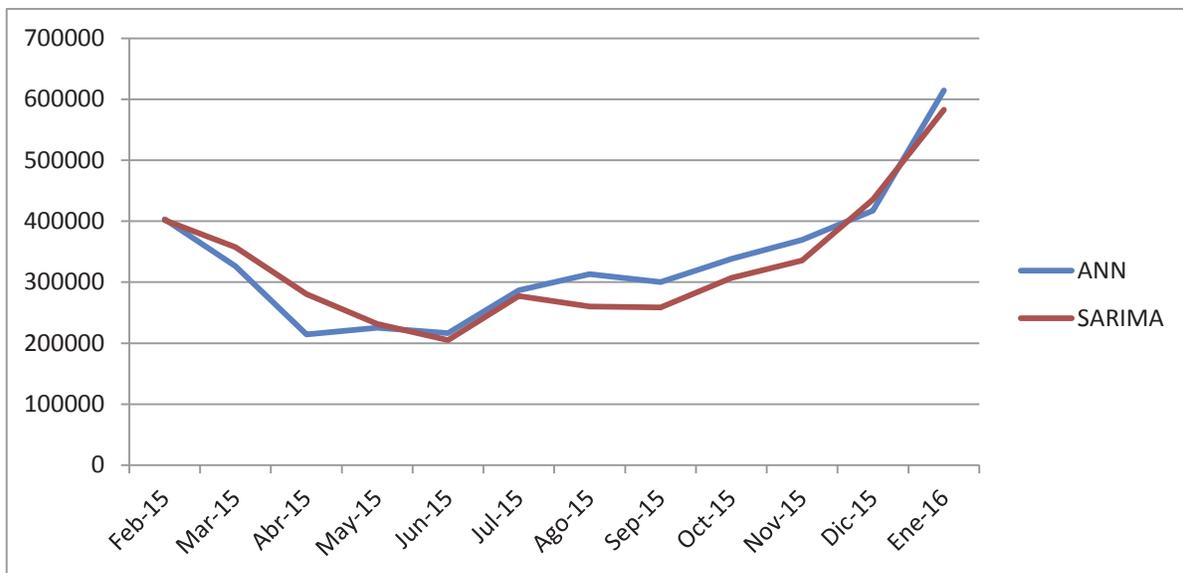


Figura 24: Predicción de los 12 Periodos Siguietes de los Datos Observados

Se puede observar una fuerte correlación entre las dos series, si esta correlación es medida mediante el coeficiente de Pearson se obtiene que existe una correlación de 0,95. Ambas modelos, al igual que en el periodo de prueba, son capaces de asimilar la estacionalidad de la serie, existiendo diferencias de magnitud de 1332 en Febrero de 2015 hasta 66138 en Abril de 2015.

8. Conclusión

La predicción del número del flujo de turistas es relevante para el rubro del turismo y los sub-rubros que este engloba, como el mundo hotelero, aerolíneas, agencias, etc. Permite tener una preparación en lo logístico e inversión de recursos para así afrontar de mejor manera el número de turistas entrantes en las próximas temporadas. Para el Estado también es un tema relevante, porque el turismo es una actividad económica significativa, que aporta un no despreciable porcentaje al PIB chileno y que además es una fuente de miles de empleos, por lo tanto, debe tener en cuenta estos tipos de estudios para realizar una inversión en los espacios públicos, parques nacionales, aeropuertos, terminales de buses, etc. Y así enfrentar con mayor preparación la tendencia al alza del número de turistas que arriban a Chile.

Mediante el uso de modelos autoregresivos para el análisis de flujos de turistas, se concluyó de que la tendencia del número de turistas en el futuro será cada vez mayor, y que el fenómeno del turismo Chileno presenta una fuerte estacionalidad, constante y estable. La tendencia del flujo de turistas que arriban a Chile parecer estar fuertemente influenciada por los desastres naturales que ocurren dentro del territorio nacional, específicamente por los terremotos, además se puede inferir que dicha tendencia fue fuertemente golpeada por la caída de las Torres Gemelas en Estados Unidos, ya que, después de alrededor de 5 años el flujo de turistas se igualó con la cantidad previa a dicho evento.

Los modelos autoregresivos para la predicción, como SARIMA y Redes Neuronales Artificiales, son herramientas ventajosas para la sociedad porque pueden ser aplicadas a diversas problemáticas que afectan a la población. Estos modelos aplicados de manera correcta, es decir, realizar un estudio acabado de los datos y entrenar el modelo con un algoritmo eficiente, pueden ayudar en la preparación a futuros eventos y enfrentarlos con los recursos necesarios, tanto logísticamente, monetariamente, infraestructura, personal, etc.

A través de los resultados obtenidos al aplicar los modelos predictivos en la serie de datos, es posible concluir que ambos modelos fueron capaces de modelar la problemática con bastante precisión según las herramientas estadísticas utilizadas, obteniendo un resultado ajustado a la realidad observada. Además se puede concluir, en base al estudio de los residuos de la serie, que el ruido observado, normalmente distribuido y con baja autocorrelación, es producto de eventos ajenos al rubro del turismo pero que sí afectan fuertemente a éste, como son los desastre naturales e incidentes sociales.

Entrando en la parte técnica del trabajo realizado, se puede concluir que el modelo SARIMA obtuvo una mejor aproximación a los datos observados que las redes neuronales artificiales probadas a una proyección futura de doce meses. La diferencia entre ellas no fue abismal, ya que, todas obtuvieron resultados bastante ajustados a la realidad, pero se puede apreciar a través de las herramientas estadística aplicada a cada modelo, que el modelo

SARIMA fue el que obtuvo el mejor rendimiento. Este modelo se impuso a las dos redes neuronales en todas las mediciones estadísticas (RMSE, NSE y R).

9. Anexo

9.1 NSE

El coeficiente de eficiencia del modelo Nash–Sutcliffe es utilizado para evaluar el poder predictivo de un modelo. La eficiencia Nash–Sutcliffe puede variar desde $-\infty$ hasta 1 (9.1.1). Una eficiencia de uno corresponde a un perfecto ajuste entre los datos entregados por el modelo y los datos reales de prueba. Una eficiencia igual a cero indica que las predicciones del modelo son tan precisas como el promedio de los datos observados, y finalmente si la eficiencia es menor a uno indica que el promedio de los datos observado es un mejor predictor que el modelo, ya que, la varianza de los residuos es mayor que la varianza de los datos. Por lo tanto, entre el coeficiente de eficiencia esté más cerca de uno más preciso es el modelo.

Y está definido de la siguiente manera:

$$E = 1 - \frac{\sum_{t=1}^T (O_t - P_t)^2}{\sum_{t=1}^T (O_t - \bar{O})^2} \quad (9.1.1)$$

donde O es la data observada y P es la salida del modelo y \bar{O} es el promedio de la data observada.

9.2 RMSE

La raíz del error cuadrático medio es frecuentemente utilizada para medir las diferencias entre los valores previstos por un modelo o un estimador y los valores realmente observados. La raíz del error cuadrático medio (9.2.1) representa la desviación estándar de la muestra de las diferencias entre los valores previstos y los valores observados. Por lo tanto, es utilizado para agregar las magnitudes de los errores en las predicciones para diferentes tiempos en una única métrica de poder predictivo. Al medir error se espera tener el más bajo valor posible como resultado.

Se define de la siguiente manera:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}} \quad (9.2.1)$$

donde n es el total de muestras, y los valores observados y \hat{y} los valores previstos por el modelo.

9.3 Operador Backshift

El operador de desplazamiento hacia atrás B (9.3.1) es una notación útil cuando se trabaja con series de tiempo desfasadas:

$$B y_t = y_{t-1} \quad (9.3.1)$$

en otras palabras, B operando en y_t tiene el efecto de desplazar la data hacia atrás en un periodo. Dos aplicaciones de B a y_t desplaza el dato dos periodos hacia atrás (9.3.2):

$$B(B y_t) = B^2 y_t = y_{t-2} \quad (9.3.2)$$

9.4 Función de Autocorrelación

La función de correlación es definida como la correlación cruzada de la señal consigo mismo. Y se calcula de la siguiente manera (9.4.1):

$$R_u = \frac{c_u}{c_0} = \frac{(\sum_{t=1}^{T-|u|} (Z_t - \bar{Z})(Z_{t-|u|} - \bar{Z}))}{(\sum_{t=1}^T (Z_t - \bar{Z})^2)} \quad (9.4.1)$$

Donde Z es el valor observado, \bar{Z} es el promedio de los valores observados y u es la diferencia de periodos con los cuales se quiere calcular la autocorrelación.

9.5 Función de Autocorrelación Parcial

La función de autocorrelación parcial es posible calcularla a partir de la función de autocorrelación (9.4.1) utilizando las ecuaciones de Yule-Walker. Esta función está definida de la siguiente manera para una serie de tiempo Z_t :

$$Z_t = a_{11} Z_{t-1} \quad (9.5.1)$$

donde a_{11} es el valor de la función de autocorrelación parcial de orden uno.

$$Z_t = a_1 Z_{t-1} + a_{22} Z_{t-2} \quad (9.5.2)$$

donde a_{22} es el valor de la función de autocorrelación parcial de orden dos.

$$Z_t = a_1 Z_{t-1} + a_2 Z_{t-2} + \dots + a_{pp} Z_{t-p} \quad (9.5.3)$$

donde a_{pp} es el valor de autocorrelación parcial de orden p .

9.6 Criterio de información de Akaike (AIC)

El criterio de información de Akaike es una de la calidad relativa de un modelo estadístico dado un conjunto de datos. Dado un conjunto de modelos para la data, el criterio de Akaike estima la calidad de cada modelo, relativo a cada uno de los otros modelos. Por lo tanto, AIC proporciona un medio para la selección de un modelo.

Este criterio ofrece una estimación relativa perdida cuando un modelo dado es utilizado para representar el proceso que genera los datos. Al hacerlo, este modelo lidia entre la bondad del ajuste y la complejidad del modelo. AIC no dice nada sobre la calidad del modelo, si todos los modelos candidatos entregan un pobre ajuste, el criterio no dará ninguna advertencia de aquello.

Y se define de la siguiente manera:

$$AIC = 2k - 2Ln(L) \quad (9.6.1)$$

donde L es el máximo valor de la función de verosimilitud del modelo estimado (9.6.2) y k es el números de parámetros estimados en el modelo. Y la función a maximizar para obtener L es:

$$\ln\mathcal{L}(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^n(x_i - \mu)^2 \quad (9.6.2)$$

Dado un conjunto modelos candidatos para los datos, el modelo preferido será aquel que contenga el menor valor AIC (9.6.1). Por lo tanto, AIC premia el buen ajuste (según evaluado por la función de verosimilitud), pero también incluye una penalización que es una función creciente de los números de parámetros estimados. La penalidad desalienta el sobre-ajuste.

9.7 Criterio de información de Akaike corregido (AICc)

AICc es AIC con una corrección para muestras de datos finitas. La fórmula para AICc depende sobre el modelo estadístico. Asumiendo que el modelo es invariante, lineal, y tiene una distribución normal de los residuos, la fórmula para AICc es la siguiente:

$$AICc = AIC + \frac{2k(k-1)}{n-k-1} \quad (9.7.1)$$

donde n denota el tamaño de la muestra y k el número de parámetros.

AICc (9.7.1) esencialmente un AIC (9.6.1) con una mayor penalidad para parámetros extras. Utilizar AIC en vez de su versión corregida incrementa la probabilidad de seleccionar modelos que contienen demasiados parámetros, es decir, sobre-ajustados.

9.8 Diferenciación

El proceso de primera diferenciación de una serie se define como:

$$d(Z_t) = Z_t - Z_{t-1} \quad (9.8.1)$$

Siendo ésta en un modelo SARIMA la diferencia regular de la serie, una diferencia estacional, con estacionalidad anual, se define de la siguiente manera:

$$d^{12}(Z_t) = Z_t - Z_{t-12} \quad (9.8.2)$$

9.9 Función Sigmoide

La función sigmoide (9.9.1) es una función real de variable real diferenciable, de la forma general:

$$y = \frac{1}{1+e^{-x}} \quad (9.9.1)$$

9.10 Estimador Máxima Verosimilitud

Los estimadores utilizados en el modelo SARIMA propuesto fueron encontrados utilizando el modelo máxima verosimilitud. Este modelo devuelve como salida los estimadores con los cuales la muestra puede ser extrapolada a toda la población.

El método lo primero que realiza es la búsqueda de la función de densidad conjunta de todas las observaciones (9.10.1), en este caso al dar el supuesto que la muestra está normalmente distribuida obtenemos que la función de densidad de la muestra con media μ y varianza σ^2 es:

$$f(x_1, \dots, x_n; \mu, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})^n} \prod_{i=1}^n \exp\left(-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2\right) \quad (9.10.1)$$

La función de verosimilitud está dada por:

$$\mathcal{L}(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta) \quad (9.10.2)$$

El método de la máxima verosimilitud estima θ_0 buscando el valor de θ que maximiza $\mathcal{L}(\theta|x)$ (9.10.2).

9.11 Algoritmo Levenberg–Marquardt

El algoritmo de Levenberg–Marquardt (9.11.1) es una combinación del algoritmo del gradiente y del algoritmo Gauss-Newton, el algoritmo de Levenberg–Marquardt intercambia entre los dos algoritmos durante la etapa de entrenamiento. Cuando el coeficientes de combinación es muy pequeño es utilizado el algoritmo de Gauss-Newton, en cambio, cuando el coeficiente de combinación es muy grande se utiliza el algoritmo del gradiente.

El algoritmo puede ser representado de la siguiente manera:

$$w_{k+1} = w_k - (J_k^T J_k + \mu I)^{-1} J_k e_k \quad (9.11.1)$$

w: Vector de los pesos

e: Vector de los errores

J: Matriz Jacobiana

μ : Coeficiente de combinación

I: Matriz Identidad

9.12 Ejemplo numérico predicción utilizando SARIMA

Explicaré con un ejemplo numérico la obtención de la primera predicción del modelo SARIMA utilizado en esta tesis. Cabe recordar que los datos de entrada al modelo están transformados a su base logarítmica, por aquello la inclusión de dicha función en la ecuación que se desarrollará.

Los parámetros del modelo son los siguientes:

Tabla 6: Factores del modelo para ejemplo

ar1	ma1	ma2	sma1
0.9775	-0.4794	-0.2283	-0.7435

Los residuos utilizados para realizar la predicción del primer valor son los siguientes:

Tabla 7: Residuos SARIMA

e_t	e_{t-1}	e_{t-2}	e_{t-3}	e_{t-4}	e_{t-5}	e_{t-6}	e_{t-7}	e_{t-8}	e_{t-9}	e_{t-10}	e_{t-11}	e_{t-12}	e_{t-13}
-0.101	-0.07	-0.058	0.039	-0.030	0.067	0.01	0.053	0.046	-0.096	0.154	-0.077	-0.121	-0.083

Por lo tanto el modelo puede ser representado de la siguiente manera:

$$(1 - \phi B)(1 - B^{12}) \ln(y_t) = (1 + \theta_1 B + \theta_2 B^2)(1 + \mu B^{12}) e_t \quad (9.12.1)$$

ϕ : ar1 θ_1 : ma1 θ_2 : ma2 μ : sma1 e: Residuos B: Operador Backshift (9.3.1)

Si se trabajaba algebraicamente la ecuación anterior y se reemplaza t por $T+1$ y reemplazando e_{T+1} por cero se obtiene:

$$\ln(y_{T+1}) = \ln(y_{T-11}) + \phi(\ln(y_T) - \ln(y_{T-12})) + \mu e_{T-11} + \theta_1 u_1 e_{T-12} + \theta_1 e_T + \theta_2 e_{T-1} + \theta_2 \mu e_{T-13} \quad (9.12.2)$$

Por lo tanto, el primer valor Predicho será igual a:

$$y_{t+1} = e^{\ln(y_{T+1})} \sim 379678 \quad (9.12.3)$$

Existiendo una diferencia de aproximadamente 470 pasajeros, esto puede ser atribuido a la cantidad de decimales que utilizó el Software al momento de computar los valores. Recordar que el valor predicho por el Software fue de 379205.

10 Referencias

- A lvarez-Díaz, M., & Rossello -Nadal, J. (2010). Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. *Tourism Economics*, 16(1)153-168.
- Chan, F., & Lim, C. (2011). Spectral analysis of seasonality in tourism demand. *Mathematics and Computers in Simulation*, 81(7), 1409-1418.
- Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*, 24(3), 323-330.
- Chu, F. L. (2008). A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management*, 29(1), 79-88.
- Chu, F. L. (2011). A piecewise linear approach to modelling and forecasting demand for Macau tourism. *Tourism Management*, 32(6), 1414-1420
- Coshall, J. T. (2009). Combining volatility and smoothing forecasts of UK demand for international tourism. *Tourism Management*, 30(4), 495-511.
- Coshall, J. T., & Charlesworth, R. (2011). A management orientated approach to combination forecasting of tourism demand. *Tourism Management*, 32(4), 759-769.
- Duke University Statistics Department, ARIMA models for time series forecasting. Disponible vía web en: <http://people.duke.edu/~rnau/411arim2.htm> Revisada por última vez el 13 de Abril de 2015
- Fourie, J., & Santana-Gallego, M. (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management*, 32(6), 1364-1370.
- Gil-Alana, L. A. (2005). Modelling international monthly arrivals using seasonal univariate long-memory processes. *Tourism Management*, 26(6), 867-878.
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 23(5), 499-510.
- Gounopoulos, D., Petmezas, D., & Santamaria, D. (2012). Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourist's origin. *Annals of Tourism Research*, 39(2), 641-666.
- Hui, T. K., & Yuen, C. C. (2002). A study in the seasonal variation of Japanese tourist arrivals in Singapore. *Tourism Management*, 23(2), 127-131

Hui, T. K., & Yuen, C. C. (2002). A study in the seasonal variation of Japanese tourist arrivals in Singapore. *Tourism Management*, 23(2), 127-131.

Instituto Nacional de Estadística, Informe anual de turismo. Disponible vía web en: http://www.ine.cl/canales/menu/publicaciones/calendario_de_publicaciones/pdf/turismo_2013.pdf Revisada por última vez el 10 de Abril de 2015.

Jackman, M., & Greenidge, K. (2010). Modelling and forecasting tourist flows to Barbados using structural time series models. *Tourism and Hospitality Research*, 10(1), 1-13.

Lim, C., & McAleer, M. (2001). Forecasting tourist arrivals. *Annals of Tourism Research*, 28(4), 965-977.

Manuel Quesada Pegalajar, Análisis de Series. Modelos Heterocedásticos. Disponible vía web en: [http://masteres.ugr.es/moea/pages/tfm1011/analisisdeseriesmodelosheterocedasticos/!](http://masteres.ugr.es/moea/pages/tfm1011/analisisdeseriesmodelosheterocedasticos/) Revisada por última vez el 04 de Abril de 2015

Page, S., Song, H., & Wu, D. C. (2012). Assessing the impacts of the global economic crisis and swine flu on inbound tourism demand in the United Kingdom. *Journal of Travel Research*, 51(2), 142-153

Santiago de la fuente Fernández, Series temporales: Modelo ARIMA Disponible vía web en: <http://www.fuenterrebollo.com/Economicas/SERIES-TEMPORALES/modelo-arima.pdf> Revisada por última vez el 15 de Abril de 2015

Shareef, R., & McAleer, M. (2007). Modelling the uncertainty in monthly international tourist arrivals to the Maldives. *Tourism Management*, 28(1), 23e45.

Smeral, E. (2010). Impacts of the world recession and economic crisis on tourism: forecasts and potential risks. *Journal of Travel Research*, 49(1), 31-38.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting - a review of recent research. *Tourism Management*, 29(2), 203-220.

Song, H., Dwyer, L., Li, G., & Cao, Z. (2012). Tourism economics research: a review and assessment. *Annals of Tourism Research*, 39(3), 1653-1682.

The Pennsylvania State University, Seasonal Models. Disponible vía web en: <https://onlinecourses.science.psu.edu/stat510/node/67> Revisada por última vez el 15 de Abril de 2015

Wan, S. K., Wang, S. H., & Woo, C. K. (2013). Aggregate vs. disaggregate forecast: case of Hong Kong. *Annals of Tourism Research*, 42, 434-438.