Análisis y predicción de series históricas con datos GPS para evaluar el desplazamiento de la corteza terrestre en Chile

Trabajo de titulación presentado por:
Estephany Romina Holmstrons

para optar al grado de:
Licenciado en Estadística

y al título de:
Ingeniero en Estadística

Profesor guía:
Dra. Orietta Nicolis

Valparaíso, Chile, 2017

Dedicado a
mi familia y sobrinito Nehemías, por ser los que me incentivaron siempre durante mi carrera universitaria, además del apoyo incondicional que me brindaron en todo momento mi madre Carolina Galindo y mi padre Miguel Holmstrons. También quiero dedicar todo esto a mi hermanita Ashley a quien siempre llevo en mi corazón por ser el angelito que me ha cuidado y ayudado a seguir adelante pese a las dificultades.

Agradecimientos

A todas aquellas personas que me motivaron a terminar esta tesis: a mis seres queridos y amistades que me incentivaron y ayudaron en este proceso, como también a mis padres Carolina Galindo y Miguel Holmstrons que siempre dieron todo y trabajaron el doble para que yo pudiera terminar mis estudios, a mis hermanos Staicy y Brian por darme apoyo emocional en ciertas etapas de mi vida académica que tanto necesité, a mi hermanita Ashley que si bien hoy no está con nosotros, es el angelito por el cual yo siempre quise salir adelante.

A mis compañeras y amigas de universidad Nicole Rodríguez, Ruth Flores, Joselin Díaz, Paula Olivares y Francisca olivares, que si bien algunas por distintas circunstancias se cambiaron de carrera en algún momento de mi etapa universitaria fueron fundamentales para mi crecimiento y formación académica.

Finalmente a mis profesores por entregarme formación académica y conocimientos necesarios para enfrentar el día de mañana nuevos desafíos en lo laboral.

Lista de abreviaturas

ADF	Test de Dickey-Fuller Aumentada.
$KPSS$	Test Kwiatkowski Phillips Schmidt Shin.
ACF	Función de autocorrelación simple.
$PACF$	Función de autocorrelación parcial.
AR	Modelo autorregresivo.
MA	Modelo media móvil.
$ARMA$	Modelo autorregresivo de media móvil.
$ARIMA$	Modelo autorregresivo integrado de media móvil.
AIC	Criterio de información Akaike.
BIC	Criterio de información Bayesiana.
$RMSE$	Raíz del error cuadrático medio.
DN	Componente Norte.
DE	Componente Este.
DA	Componente Altura.

Resumen

Chile es considerado uno de los países más sísmicos por encontrarse sobre la placas de Nazca y Sudamericana, las cuales al chocar producen fenómenos sísmicos. El Instituto Geográfico Militar de Santiago (IGM) adquirió observaciones de este movimiento a través de estaciones GPS de monitoreo distribuidas por todo Chile. Con los datos obtenidos se realiza un estudio para estimar y predecir el desplazamiento de la corteza terrestre durante los años 2010, 2011 y 2012.

Se comenzará realizando un análisis preliminar a los datos de cada estación de monitoreo. Posterior a esto se escogen 3 estaciones de cada zona del país (norte, centro y sur) para finalmente aplicar a los datos GPS de cada estación series de tiempo utilizando modelos determinísticos y estocásticos.

Cabe mencionar que no se cuenta con información completa en relación a los datos, sin embargo a los faltantes se le adoptarán técnicas de estimación, para posteriormente aplicar métodos típicos utilizados en series de tiempo.

Índice general

Agradecimientos **II**

Lista de abreviaturas **III**

Resumen **IV**

Lista de figuras **VIII**

Lista de tablas **XIV**

Introducción **1**

1. Fenómenos sísmicos **3**
1.1. Introducción a los fenómenos sísmicos . . . 3
1.1.1. Fenómenos sísmicos en Chile . . . 4
1.2. Geodesia . . . 6
1.2.1. Geodesia espacial: GPS . . . 6
1.2.1.1. Desplazamiento de las placas tectónicas . . . 7
1.2.1.2. Vector desplazamiento . . . 9
1.2.2. Datos GPS . . . 11

2. Series de tiempo **12**
2.1. Introducción a series de tiempo . . . 12
2.2. Análisis de series de tiempo . . . 13
2.2.1. Componentes de una serie de tiempo . . . 13
2.2.2. Análisis de la tendencia en series de tiempo . . . 16
2.2.3. Métodos de suavizamientos . . . 16
2.2.4. Estimación de la tendencia a través de regresión . . . 17
2.2.5. Estimación no paramétrica de la tendencia . . . 20
2.2.5.1. Método de suavizamiento spline . . . 21
2.3. Autocorrelación y correlograma . . . 22
2.3.1. Autocorrelación . . . 22

2.3.1.1. Función de autocorrelación simple (ACF) 22
2.3.1.2. Función de autocorrelación parcial (PACF) 23
2.3.2. Correlograma 23
2.4. Series de tiempo estacionarias y no estacionarias 25
2.5. Procesos estocásticos en series de tiempo 25
2.5.1. Procesos estocásticos estacionarios 27
2.5.2. Proceso ruido blanco 28

3. Modelos de series de tiempo 30
3.1. Modelos univariables 30
3.1.1. Modelo Autorregresivo AR(p) 30
3.1.2. Modelo media móvil MA(q) 36
3.1.3. Proceso autorregresivo AR(p) y media móvil MA(q) 40
3.1.4. Modelo autorregresivo de media móvil (ARMA) 40
3.1.5. Modelo autorregresivo integrado de media móvil (ARIMA) . . 42

4. Identificación, estimación, diagnóstico y pronóstico sobre los distintos modelos en estudio 44
4.1. Prueba de raíz unitaria 44
4.2. Identificación del modelo 46
4.3. Estimación de parámetros 47
4.4. Diagnóstico del modelo 49
4.4.1. Supuestos sobre los residuos 49
4.4.2. Criterios de información Akaike y Bayesiana 50
4.5. Pronóstico 51
4.5.1. Error de pronóstico 52

5. Análisis de series de tiempo respecto al desplazamiento de la corteza terrestre 53
5.1. Información sobre los datos en estudio 53
5.2. Análisis preliminar 54
5.3. Análisis de estacionariedad sobre las series de tiempo 55
5.4. Estimación sobre las series de tiempo 59
5.4.1. Estimación sobre las series de tiempo en Iquique 59
5.4.1.1. Estimación de la tendencia 59
5.4.1.2. Estimación de la componente residual 63
5.4.1.3. Diagnóstico del modelo 67
5.4.1.4. Pronóstico 70
5.4.2. Estimación sobre las series de tiempo en Concepción 76

5.4.2.1. Estimación de la tendencia 76
5.4.2.2. Estimación de la componente residual 82
5.4.2.3. Diagnóstico del modelo 87
5.4.2.4. Pronóstico 92
5.4.3. Estimación sobre las series de tiempo en Llafenco 100
5.4.3.1. Estimación de la tendencia 101
5.4.3.2. Estimación de la componente residual 104
5.4.3.3. Diagnóstico del modelo 107
5.4.3.4. Pronóstico 110

6. Conclusiones **117**

A. Estimación de Modelos: Iquique **119**
A.1. Componente Norte 119
A.2. Componente Este 120
A.3. Componente Altura 120

B. Estimación de Modelos: Concepción **121**
B.1. Componente Norte 121
B.2. Componente Este 121
B.3. Componente Altura 123

C. Estimación de Modelos: Llafenco **124**
C.1. Componente Norte 124
C.2. Componente Este 124
C.3. Componente Altura 125

Bibliografía **126**

Índice de figuras

1.1. Terremoto de Maule (Parra et al., 2010). 5
1.2. Destrucciones estructurales, producto del terremoto del 27 de febrero del 2010 (Parra et al., 2010). 5
1.3. Estaciones GPS de monitoreo en Chile (Parra et al., 2010). 7
1.4. Movimiento GPS (en milímetros) de la componente Norte, Este y Vertical en la estación de Washington, durante los años 2004 hasta el 2006 (issuu, 2012). 8
1.5. Movimiento noreste (en milímetros) en la estación de Washington, durante los años 2004 hasta el 2006 (issuu, 2012). 10
1.6. Movimientos de las placas tectónicas entre Chile y Argentina (Parra et al., 2010). 10
1.7. Gráficos de las componente Norte, Este y Altura, con respecto al desplazamiento de la corteza terrestre en Concepción (Parra et al., 2010). 11

2.1. Serie de tiempo que sigue un comportamiento tendencial. 14
2.2. Serie de tiempo que sigue un comportamiento estacional. 14
2.3. Serie de tiempo que sigue un comportamiento cíclico. 15
2.4. Serie de tiempo que sigue un comportamiento aleatorio. 15
2.5. Linea de ajuste sobre una serie de tiempo con tendencia lineal. 19
2.6. Tendencia exponencial. 20
2.7. Correlograma ACF y PACF, de las temperaturas máximas diarias en Valparaíso, durante el periodo 2012-2013. 24
2.8. Correlogramas ACF con componente tendencial, estacional y cíclico. . 25
2.9. Serie de tiempo y ACF de un proceso de ruido blanco gaussiano. . . . 29

3.1. ACF y PACF de un proceso autorregresivo: $(1 - 0.4B)x_t = \omega_t$. 34

5.1. Serie de tiempo y ACF de la componente Norte de Iquique. 55
5.2. Serie de tiempo y ACF de la componente Este y Altura de Iquique. . 56
5.3. Serie de tiempo y ACF de la componente Norte de Concepción. . . . 56

5.4. Serie de tiempo y ACF de la componente Este y Altura de Concepción. 57
5.5. Serie de tiempo y ACF de la componente Norte de Llafenco. 57
5.6. Serie de tiempo y ACF de la componente Este y Altura de Llafenco. . 58
5.7. Ajuste de la recta sobre las componente Norte y Este de Iquique. . . 60
5.8. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Iquique. 61
5.9. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Iquique. 62
5.10. Gráfico de la primera diferenciación, con sus respectivos correlogramas ACF y PACF sobre los residuos de la componente Norte y serie de tiempo componente Altura de Iquique. 64
5.11. Ajuste del modelo ARIMA(2,1,1) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Norte de Iquique. . 66
5.12. Ajuste del modelo ARMA(2,2) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Este de Iquique. . . . 66
5.13. Ajuste del modelo ARIMA(1,1,1) sobre la serie de tiempo componente Altura de Iquique. 66
5.14. Gráfico de dispersión de los residuos del modelo ARIMA(2,1,1) con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Iquique. 67
5.15. Gráfico de dispersión de los residuos del modelo ARMA(2,2) con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Iquique. 68
5.16. Gráfico de dispersión de los residuos del modelo ARIMA(1,1,1) con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Iquique. 69
5.17. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Norte de Iquique. 71
5.18. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Este de Iquique. 71
5.19. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(2,1,1) sobre los residuos de la componente Norte de Iquique. 72
5.20. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(2,2) sobre los residuos de la componente Este de Iquique. 73

5.21. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(1,1,1) sobre la componente Altura de Iquique. 73
5.22. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Norte de Iquique. 74
5.23. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Este de Iquique. 75
5.24. Ajuste de la recta sobre las componente Norte y Altura de Concepción. 77
5.25. Suavizamiento spline sobre las componente Este. 78
5.26. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Concepción. 79
5.27. Gráfico de los residuos obtenidos en el suavizamiento spline, con sus respectivos correlogramas ACF y PACF, para la componente Este de Concepción. 80
5.28. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Concepción. 81
5.29. Gráfico de la primera diferenciación, con sus respectivos correlogramas ACF y PACF sobre los residuos de la componente Norte y Altura de Concepción. 83
5.30. Ajuste del modelo ARIMA(1,1,1) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Norte de Concepción. 86
5.31. Ajuste del modelo ARMA(2,6) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Este de Concepción. 86
5.32. Ajuste del modelo ARMA(4,7) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Este de Concepción. 86
5.33. Ajuste del modelo ARIMA(2,1,1) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Altura de Concepción. 87
5.34. Gráfico de dispersión de los residuos del modelo ARIMA(1,1,1), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Concepción. 88
5.35. Gráfico de dispersión de los residuos del modelo ARMA(2,6), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Concepción. 89
5.36. Gráfico de dispersión de los residuos del modelo ARMA(4,7), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Concepción. 90

5.37. Gráfico de dispersión de los residuos del modelo ARIMA(2,1,1), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Concepción. 91
5.38. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Norte de Concepción. 93
5.39. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Este de Concepción. 93
5.40. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Altura de Concepción. 94
5.41. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(1,1,1) sobre los residuos de la componente Norte de Concepción. 95
5.42. Gráfico de predicción para el mes de diciembre del 2012, utilizando el modelo ARMA(2,6) sobre los residuos de la componente Este de Concepción. 95
5.43. Gráfico de predicción para el mes de diciembre del 2012, utilizando el modelo ARMA(4,7) sobre los residuos de la componente Este de Concepción. 96
5.44. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(2,1,1) sobre los residuos de la componente Altura de Concepción. 96
5.45. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(2,6) sobre los residuos de la componente Este de Concepción. 97
5.46. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Norte de Concepción. 98
5.47. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Este de Concepción. 99
5.48. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Altura de Concepción. 99
5.49. Suavizamiento spline sobre las componente Norte y Altura de Llafenco. 101

5.50. Gráfico de los residuos obtenidos en el suavizamiento spline, con sus respectivos gráficos ACF y PACF, para la componente Norte de Llafenco. 102
5.51. Gráfico de los residuos obtenidos en el suavizamiento spline, con sus respectivos gráficos ACF y PACF, para la componente Este de Llafenco. 103
5.52. Ajuste del modelo ARMA(3,4) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Norte de Llafenco. 106
5.53. Ajuste del modelo ARMA(4,7) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Este de Llafenco. 106
5.54. Ajuste del modelo ARMA(2,2) sobre los residuos de la serie de tiempo componente Altura de Llafenco. 106
5.55. Gráfico de dispersión de los residuos del modelo ARMA(3,4), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Llafenco 107
5.56. Gráfico de dispersión de los residuos del modelo ARMA(4,7), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Llafenco. 108
5.57. Gráfico de dispersión de los residuos del modelo ARMA(2,2), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Llafenco. 109
5.58. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Norte de Llafenco. 111
5.59. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Este de Llafenco. 111
5.60. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(3,4) sobre los residuos de la componente Norte de Llafenco. 112
5.61. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(4,7) sobre los residuos de la componente Este de Llafenco. 113
5.62. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(2,2) sobre la componente Altura de Llafenco. 113
5.63. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte de Llafenco. 114

5.64. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Este de Llafenco. 115

Índice de tablas

4.1. Características principales del correlograma ACF y PACF, en los modelos AR(p), MA(q) y ARMA(p,q). 47

5.1. Resultados obtenidos con respecto a los datos faltantes, en las 22 estaciones de monitoreo GPS en Chile, durante enero del 2010 hasta diciembre del 2012. 54
5.2. Resultados obtenidos de los test ADF y KPSS sobre las componentes Norte, Este y Altura de Iquique, Concepción y Llafenco. 58
5.3. Resultados obtenidos de la estimación de la tendencia, sobre la componente Norte y Este de Iquique. 60
5.4. Resultados de los test ADF y KPSS, sobre los residuos de la componente Norte y Este de Iquique. 63
5.5. Resultados de los test ADF y KPSS, aplicados sobre la primera diferenciación de los residuos de la componente Norte y serie de tiempo de la componente Altura de Iquique. 63
5.6. Modelos seleccionados para cada componente, con sus respectivos criterios de información AIC y BIC. 65
5.7. Contraste de significación de los parámetros del modelo ARIMA(2,1,1), ARMA(2,2) y ARIMA(1,1,1). 65
5.8. Test Ljung-Box sobre los residuos de cada modelo. 70
5.9. Predicciones obtenidas sobre los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Norte y Este de Iquique. 72
5.10. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto a la componente Altura y los residuos de la componente Norte y Este de Iquique. 74
5.11. Predicciones generales obtenidas para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte y Este de Iquique. 75

5.12. Predicciones finales y módulo del vector desplazamiento obtenidos para los primeros cuatro días de enero del 2013, con respecto a la componente Norte, Este y Altura de Iquique. 76
5.13. Resultados obtenidos de la estimación de la tendencia, sobre la componentes Norte y Altura de Concepción. 77
5.14. Resultados obtenidos de la estimación de la tendencia a través del suavizamiento spline sobre la componente Este de Concepción. 78
5.15. Resultados de los test ADF y KPSS sobre los residuos de la componentes Norte, Este y Altura de Concepción. 82
5.16. Resultados de los test ADF y KPSS, aplicados sobre la primera diferenciación de los residuos de la componente Norte y Altura de Concepción. 82
5.17. Modelos seleccionados para cada componente, con sus respectivos criterios de información AIC y BIC 84
5.18. Contraste de significación de los parámetros del modelo ARIMA(1,1,1), ARMA(2,6), ARMA(4,7) y ARIMA(2,1,1). 85
5.19. Test Ljung-Box sobre los residuos de cada modelo. 91
5.20. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Este de Concepción y la regresión lineal sobre la componente Norte y Altura de Concepción. 94
5.21. Resultados obtenidos del cálculo RMSE, para los modelos ARMA(2,6) y ARMA(4,7) 97
5.22. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto a los residuos de la componente Norte, Este y Altura de Concepción. 98
5.23. Predicciones generales obtenidas para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte, Este y Altura de Concepción. 100
5.24. Predicciones finales y módulo del vector desplazamiento obtenidos para los primeros cuatro días de enero del 2013, con respecto a la componente Norte, Este y Altura de Concepción. 100
5.25. Resultados obtenidos de la estimación de la tendencia sobre la componentes Norte y Este de Llafenco. 101
5.26. Resultados de los test ADF y KPSS sobre los residuos de la componentes Norte y Este de Llafenco. 104
5.27. Modelos seleccionados para cada componente, con sus respectivos criterios de información AIC y BIC 105

5.28. Contraste de significación de los parámetros del modelo ARMA(3,4), ARMA(4,7) y ARMA(2,2). 105
5.29. Test Ljung-Box sobre los residuos de cada modelo. 109
5.30. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Norte y Este de Llafenco. 112
5.31. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto a la componente Altura de Llafenco y los residuos de la componente Norte y Este de Llafenco. 114
5.32. Predicciones generales obtenidas para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte y Este de Llafenco. 115
5.33. Predicciones finales y módulo del vector desplazamiento obtenidos para los primeros cuatro días de enero del 2013, con respecto a la componente Norte, Este y Altura de Llafenco. 116

A.1. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre los residuos obtenidos de la regresión lineal, de la serie de tiempo componente Norte. 119
A.2. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos de la regresión lineal de la serie de tiempo componente Este. 120
A.3. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre la serie de tiempo componente Altura. 120

B.1. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre los residuos obtenidos de la regresión lineal de la serie de tiempo componente Norte. 121
B.2. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos del suavizamiento spline, de la serie de tiempo componente Este. 122
B.3. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre los residuos obtenidos de la regresión lineal, de la serie de tiempo componente Altura. 123

C.1. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos del suavizamiento spline, de la serie de tiempo componente Norte. 124

C.2. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos del suavizamiento spline, de la serie de tiempo componente Este. 125
C.3. Resultados de los distintos modelos ARMA(p,q), aplicados sobre la serie de tiempo componente Altura. 125

Introducción

A lo largo de los años se ha comprobado que Chile es uno de los países más sísmicos a nivel mundial, por este motivo se han realizado diversos estudios para entender de cierta forma estos tipos de fenómenos. Los sismos son producto del movimiento de la corteza terrestre, los cuales generan desplazamientos, hundimientos y alzamientos, dando paso a que ocurran distintos tipos de deformaciones. El movimiento que conlleva al desplazamiento de la corteza terrestre será el tema central de la presente tesis, ya que los datos adquiridos son con respecto a este tema.

Los datos que se utilizan para este estudio son con GPS (Global Positioning System) ya que "la determinación de los desplazamientos de la corteza asociados a los ciclos sísmicos, requiere una metodología de alta precisión para determinar movimientos pequeños sobre superficies grandes" (Ruegg, 1994, pág 202), por lo que la geodesia moderna, específicamente la geodesia espacial como lo es el GPS es ahora un recurso preciso para el estudio del ciclo sísmico.

Es importante mencionar que con las posiciones de GPS existen dispersiones que se caracterizan por errores y tendencias (Oware, 1998, citado por Li et al, 2000, pág 155). Es por este motivo que se pretende aplicar a las series de tiempo con datos GPS, un modelo estadístico capaz de solucionar dichos inconvenientes aproximando y estimando, por ejemplo la tendencia de los datos, como también haciendo que los residuos se modelen a través de procesos estocástico.

Objetivo general:

- Analizar y predecir el desplazamiento de la corteza terrestre durante los años 2010, 2011 y 2012, utilizando datos GPS.

Objetivos específicos:

- Analizar los datos de manera preliminar e imputar datos faltantes.
- Estimar la tendencia de la serie de tiempo.
- Aplicar un modelo estocástico sobre los residuos de la serie de tiempo.

- Comparar los modelos estimados de la zona norte, centro y sur del país.

Hipótesis:

- Suponer que la serie de tiempo GPS sigue un componente determinístico más un componente estocástico.

Capítulo 1

Fenómenos sísmicos

En este capítulo se procede a una breve introducción sobre los fenómenos sísmicos de modo general, para luego dar un enfoque mas específico con respecto a los fenómenos sísmicos en Chile.

1.1. Introducción a los fenómenos sísmicos

Cada cierto tiempo la población se ve sacudida por fenómenos sísmicos, los cuales acarrean consecuencias que resultan evidentes a simple vista. Las causas de los diversos fenómenos sísmicos pueden ser movimientos de las placas tectónicas, hundimientos, deslizamientos, entre otros y las consecuencias que conllevan estas causas suelen ser destrucciones a nivel de infraestructura, derrumbes, maremotos, daños diversos en la corteza terrestre como desplazamientos, etc. En esta ocasión el enfoque principal irá hacia los desplazamientos de la corteza terrestre asociados al ciclo sísmico.

El concepto de ciclo sísmicos según Ruegg (1994), está fundado en la comprobación que los terremotos de cierta importancia van a repetirse a períodos de tiempo regular aproximadamente en el mismo lugar. El ciclo sísmico incluye tanto fenómenos físicos (dos terremotos (magnitud $M > 6$) de gran relevancia consecutivos en una zona sísmica) como procesos de acumulación y de relajación de los esfuerzos, que se repetirán en cada periodo del ciclo de manera similar, pero que no tienen que ser necesariamente idénticos (Ruegg, 1994).

"Una gran parte de sismos producidos en la tierra ocurren a lo largo de fallas en zonas tectónicamente activas"(Gonzáles et al., 2006, pág 488). Al observar las deformaciones del suelo, se logra apreciar que ocurren producto de la actividad tectónica, poniendo en evidencia aquellas fallas que no son visibles en superficie y aquellas que han sido formadas recientemente (Gonzáles et al., 2006). Las deformaciones de la corteza terrestre se definen como cualquier cuerpo de roca que tiene un punto en

el que se fracturará independiente de su dureza. La palabra deformación se utiliza como un término general que se refiere a todos los cambios de tamaño, forma, orientación o posición de una masa rocosa. Las deformaciones de la corteza terrestre suelen tener lugar a lo largo de los márgenes de las placas, por consiguiente los movimientos generados en las placas y las interacciones a lo largo de los límites de placas producen las fuerzas tectónicas que ocasionan la deformación de las unidades de roca (Tarbuck et al., 2005, pág 284).

Comprender el campo de la deformación abre paso a la comprensión del proceso mecánico que ocurre en una falla durante el ciclo sísmico. Las fallas son rupturas en la corteza, que dan lugar a un desplazamiento perceptible. Los desplazamientos co-sísmicos se asocian a la ruptura y por ende entrega información sobre el mecanismo tanto de posición, tamaño, momento, etcétera, permitiendo una modelización de éstas. Para la determinación de los desplazamientos de la corteza terrestre se requiere de una metodología de alta precisión, ésto para lograr determinar movimientos pequeños sobre grandes superficies. Existen varias técnicas para dicha determinación y una de ellas es la geodesia espacial, la cual es muy utilizada, por ser un recurso excelente para el estudio del ciclo sísmico (Ruegg, 1994). La geodesia espacial será explicada más adelante.

1.1.1. Fenómenos sísmicos en Chile

Chile es uno de los países que tiene más registros históricos con respecto a fenómenos sismos, por el hecho de encontrarse ubicada sobre la placa sudamericana, la cual se sitúa justo al límite de la placa tectónica de nazca (Madariaga, 1998). Cuando la placa oceánica de nazca se introduce bajo el continente, se produce una zona de subducción, es decir, el hundimiento de una placa de naturaleza oceánica bajo una placa adyacente. Es por esta razón que "la actividad sísmica dominante es una consecuencia directa de la subducción de la placa de nazca sudamericana" (Madariaga, 1998, pág 221). Los efectos de estos sismos o terremotos (dependiendo de la magnitud) son muy habituales, generando grandes daños, por ejemplo de infrestructura y un gran número de pérdidas de vidas humanas. Se han realizado muchos estudios sobre éstos fenómenos con el propósito de detectar variaciones temporales de la sismicidad o de la deformación que den indicios de un posible próximo terremoto destructor en Chile.

Uno de los últimos sismos ocurridos en Chile que tuvo gran magnitud, denominándose de esta manera terremoto (magnitud $M > 6$), fue el del 27 de febrero del 2010, el cual tuvo su mayor intensidad en el sur del país. A continuación se presenta en la Figura 1.1 la representación de la intensidad y el epicentro del terremoto de Maule, ocurrido el sábado 27 de febrero del 2010 a las 3:34 AM (hora local), con

una magnitud de 8.8°.

Figura 1.1. Terremoto de Maule (Parra et al., 2010).

Las repercusiones producto de este terremoto se dejaron ver de inmediato. En la Figura 1.2 se logran apreciar algunas de las destrucciones estructurales producidas por este terremoto.

(a) **(b)**

(c) **(d)**

Figura 1.2. Destrucciones estructurales, producto del terremoto del 27 de febrero del 2010 (Parra et al., 2010).

1.2. Geodesia

Para adquirir el conocimiento adecuado sobre las deformaciones de la corteza terrestre, se requiere de una metrología de muy alta precisión, capaz de evaluar los movimientos pequeños que se generan sobre grandes superficies. En este caso la técnica más utilizada para este tipo de estudios es la geodesia, la cual tiene como objetivo medir y conocer las dimensiones de la Tierra. A medida que ha pasado el tiempo esta técnica ha adquirido mejoras llegando al desarrollo de la geodesia espacial, que permite la medición de largas distancias, entre dos puntos con visibilidad directa al suelo, o entre puntos del suelo y un objecto celeste, como por ejemplo un satélite artificial (Ruegg, 1994).

1.2.1. Geodesia espacial: GPS

Existen diferentes métodos de geodesia espacial, algunas más utilizadas que otras. En esta ocasión el enfoque será para el sistema GPS (global positioning system).

El sistema GPS es utilizado para el estudio de la Tierra, es decir, identificar el movimiento y deformación de las placas tectónicas. Los monumentos GPS se encuentran unidos al suelo y miden como se mueven las placas. Las unidades del GPS de alta precisión usadas para los estudios científicos miden que tan rápido o lento es el movimiento de las placas de la tierra, llegando a medir algunos milímetros en un año y por muy mínimos que parezcan pueden ser importantes, ya que la roca de movimiento lento puede provocar terremotos de grandes magnitudes.

El GPS utiliza señales electromagnéticas sobre 2 frecuencias (1.5 y 1.2 GHz) con varios códigos que permiten variados tipos de medición. Las señales que emite son a través de satélites que pertenecen a una constelación de más de 24 satélites de la seria NAVSTAR. Las señales que emiten los satélites son recibidas en tierra, por antenas de receptores GPS. Producto que las señales emitidas son fuertes, no existe la necesidad de utilizar instrumentos de gran tamaño, haciendo que éstos sean movibles, con bajo consumo de energía y baratos en cuanto a costo operacional de las mediciones GPS (Ruegg, 1994).

En la Figura 1.3 se presentan las diferentes estaciones GPS de monitoreo ubicadas en Chile, las cuales permiten calcular coordenadas de la actual ubicación espacial de un punto sobre la superficie de la tierra diariamente. Su aplicación conlleva a monitoreos de deformaciones de la superficie, cálculo de velocidades, entre otros.

Figura 1.3. Estaciones GPS de monitoreo en Chile (Parra et al., 2010).

1.2.1.1. Desplazamiento de las placas tectónicas

Con la geodesia espacial, específicamente el sistema GPS, se ha tenido una mejor comprensión de como se mueven las placas tectónicas y la relación de tales movimientos con fenómenos naturales, como son los terremotos y actividades volcánicas. Una de las ventajas que se obtiene con el sistema GPS, es que se logra medir el desplazamiento de las placas tectónicas en las direcciones Norte, Este y Altura. Para comprender lo mencionado, se ilustra en la Figura 1.4 los movimientos registrados a través de GPS en una estación en Washington, que está siendo empujada hacia el noreste por la subducción de la placa de Juan de Fuca (issuu, 2012).

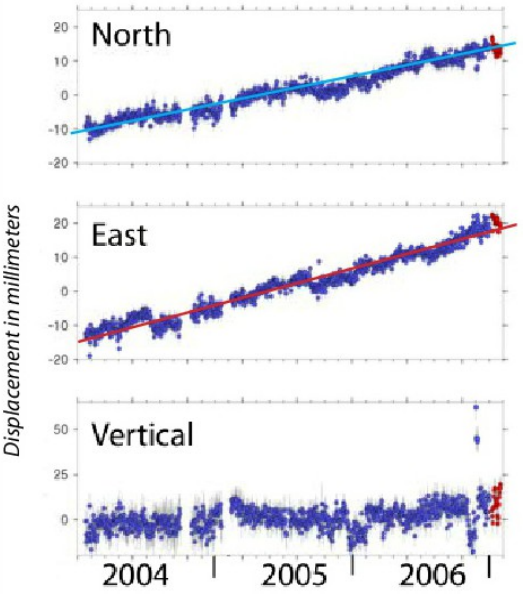


Figura 1.4. Movimiento GPS (en milímetros) de la componente Norte, Este y Vertical en la estación de Washington, durante los años 2004 hasta el 2006 (issuu, 2012).

En esta ocasión para calcular el desplazamiento, se utiliza la fórmula (Tipler y Mosca, 2005):

$$\Delta x = x_f - x_i \ , \tag{1.2.1}$$

donde x_i es la posición en el instante t_i y x_f es la posición en la que se encuentra en un instante posterior t_f.

En el ejemplo de la Figura 1.4, x_f representa el último dato del movimiento y x_i representa el primer dato del movimento, registrado en la dirección Norte y Este. Por consiguiente el desplazamiento entre el 2004 y 2007 en la estación Washington (en la Figura 1.4), se presenta como:

Último dato de la componente Norte: 12 mm
Primer dato de la componente Norte:-11 mm
Último dato de la componente Este: 18 mm
Primer dato de la componente Este: -15 mm

El cálculo del desplazamiento se obtiene de la siguiente manera:

Componente Norte: 12 mm - (-11 mm)= 23 mm
Componente Este: 18 mm - (-15 mm)= 33 mm

Finalmente entre 2004 y 2007 la estación Washington se desplazó 23 mm hacia el Norte y 33 mm hacia el Este.

1.2.1.2. Vector desplazamiento

Cuando existen movimientos de las placas tectónicas, el desplazamiento tiene una dirección en el espacio y un módulo. La magnitud representada como la dirección y distancia entre dos puntos del espacio en linea recta, se determina como vector desplazamiento. Su representación gráfica es a través de una flecha cuya dirección es la misma que la del vector desplazamiento y cuya longitud es proporcional al módulo del vector desplazamiento (Tipler y Mosca, 2005).

El módulo del vector desplazamiento representa las dimensiones de longitud. Debido a que se trata de la hipotenusa de un triángulo rectángulo, su cálculo se realiza mediante el teorema de Pitágoras. Es importante mencionar que el módulo de un vector no puede ser negativo. La fórmula del teorema de Pitágoras se denota como (Tipler y Mosca, 2005):

- En R2:

$$A = \sqrt{A_x^2 + A_y^2} \tag{1.2.2}$$

- En R3 (en tres dimensiones):

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2} \tag{1.2.3}$$

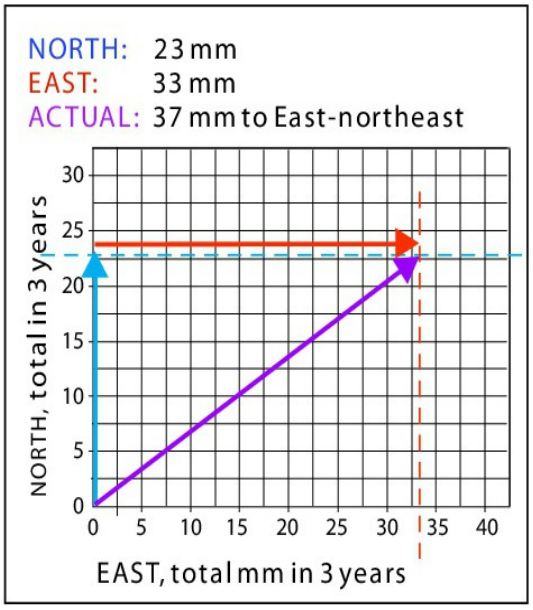


Figura 1.5. Movimiento noreste (en milímetros) en la estación de Washington, durante los años 2004 hasta el 2006 (issuu, 2012).

En la Figura 1.5 se trazan los vectores del desplazamiento Norte y Este, con los cuales se determina el movimiento que fue de 37 mm hacia el noreste.

En la Figura 1.6 se representa el movimiento de las placas tectónicas, a través de GPS entre Chile y Argentina. La longitud de las flechas indica la velocidad del movimiento de esa parte de la placa.

Figura 1.6. Movimientos de las placas tectónicas entre Chile y Argentina (Parra et al., 2010).

1.2.2. Datos GPS

Los datos GPS comúnmente utilizados para realizar diferentes mediciones, se determinan mediante coordenadas. En el caso de medir el desplazamiento de la corteza terrestre, los datos GPS se integran de tres coordenadas, denominadas componente Norte, componente Este y componente Altura. En cada componente se registran series de tiempo que miden el desplazamiento de la corteza terrestre, ya sea diariamente, semanalmente, mensualmente, etcétera.

El sistema GPS en Chile cumple una función muy importante, ya que a través de éste se logran obtener datos que registran mediciones sobre desplazamientos, velocidades, entre otros, de las placas tectónicas producto de los fenómenos sísmicos. En la Figura 1.7 se pueden apreciar tres gráficos que representan las componentes Norte (DN), componente Este (DE) y componente altura (DA), en los cuales se mide el desplazamiento de la corteza terrestre en Concepción, durante los años 2007 hasta 2012. Se puede apreciar en el gráfico de la Componente Norte y Este un gran salto durante febrero del 2010 producto del terremoto que afectó gran parte del territorio chileno. Finalmente queda en evidencia que Concepción sufrió un desplazamiento significativo después del terremoto.

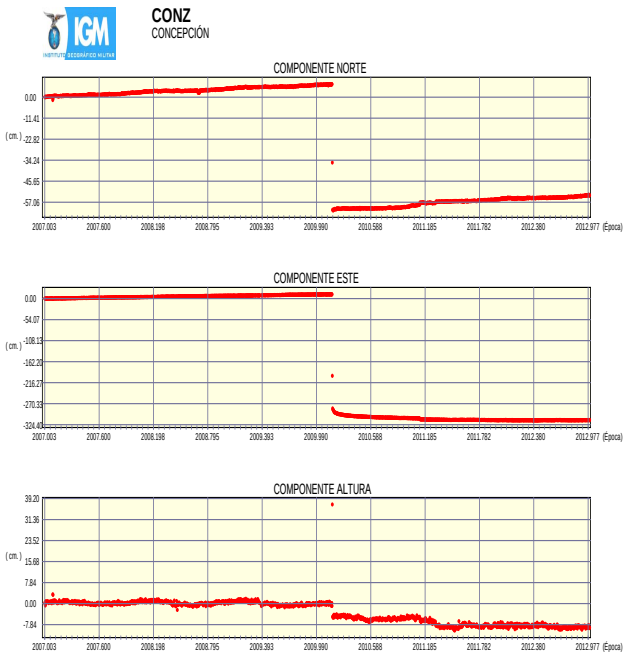


Figura 1.7. Gráficos de las componente Norte, Este y Altura, con respecto al desplazamiento de la corteza terrestre en Concepción (Parra et al., 2010).

Capítulo 2

Series de tiempo

2.1. Introducción a series de tiempo

Al indagar diferentes áreas de estudios se ha podido comprobar que los datos que se utilizan son seleccionados de manera secuencial en el tiempo, es decir, los datos se pueden registrar, observar y analizar, por ejemplo diariamente, semanalmente, mensualmente y entre otros periodos de tiempo. A lo mencionado anteriormente se le llama series de tiempo o también conocidas como series históricas o cronológicas. Los objetivos principales al realizar el análisis son:

- Identificar y describir los componentes (tendencia, estacionalidad, ciclo y aleatoriedad) encontrados en las observaciones de la serie de tiempo.
- Ajustar los componentes de tal manera que se pueda predecir el comportamiento que seguirán en un futuro las observaciones, con el propósito de controlar por ejemplo situaciones fortuitas.

A modo general Chatfield (1989) describe en su libro "The analysis of time series. An introduction", que los objetivos principales del análisis de una serie temporal son describir, explicar, predecir y controlar.

La aplicación de series de tiempo puede ser utilizada en diferentes áreas. Algunos ejemplos son: en economía, cuando se observan las ventas o precios de un artículo en un supermercado, también para las tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etcétera. En meteorología cuando se observan las temperaturas máximas o mínimas diarias, o en demografía al estudiar la tasa de natalidad o mortalidad infantil, entre otros. De esta manera se pueden seguir enumerando distintas áreas que utilizan registros secuenciales en sus estudios.

2.2. Análisis de series de tiempo

Cuando se realizan análisis a series de tiempo, la representación gráfica de las variables es fundamental, porque da un indicio de cómo se comportan y evolucionan los datos en el tiempo. Sin una representación gráfica de los datos resulta difícil el análisis de una series de tiempo, ya que no se lograría tener una visualización de lo que se estudia. Luego de representar gráficamente la secuencia de valores se debe identificar la existencia de tendencia, estacionalidad, variaciones irregulares (si estos son completamente aleatorios) o algún patrón regular a lo largo del tiempo.

Los modelos utilizados generalmente en series de tiempo, suele ser expresados como suma o producto de tres componentes: tendencia, estacionalidad y aleatoriedad. Es importante destacar que dependiendo de cómo se comporten los valores en el tiempo, se aplicaran modelos aditivos multiplicativo o mixtos.

Puede ocurrir que la representación gráfica de la serie temporal por ejemplo muestre tendencia pero no estacionalidad, por ende el modelo solo deberá incluir los componentes tendencia y aleatoriedad. Más adelante se explicará con detalle cada componente.

2.2.1. Componentes de una serie de tiempo

Las técnicas habituales para analizar series de tiempo como ya se ha mencionado anteriormente, consisten en la descomposición de cuatro componentes. Algunos de estos aspectos no pueden ser observables a simple vista, por lo que resulta conveniente disponer de procedimientos que permitan extraer dichos aspectos de las observaciones (Bustamante, 2013).

Para examinar la naturaleza de los componentes de una serie temporal es importante definir cada uno de estos, ya que serán los que expliquen la variación y el comportamiento de los datos. A continuación se definen los cuatro componentes que integran una serie de tiempo, incluyendo un ejemplo para cada uno de éstos.

Componente tendencia T_t: Es el comportamiento creciente o decreciente de la serie observada en torno a la media en un periodo de largo plazo, donde se produce una variación en los datos en comparación de su media o equilibrio. Por lo general se considera un movimiento determinístico y se requiere que la serie posea un número muy elevado de muestras. En la Figura 2.1 se observa una serie de tiempo que presenta la componente de tendencia.

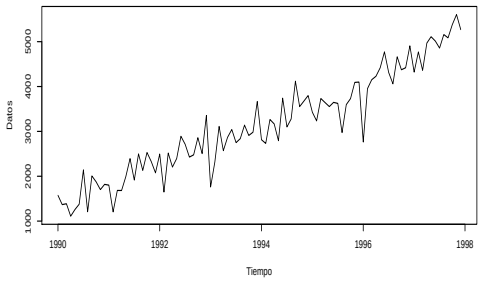


Figura 2.1. Serie de tiempo que sigue un comportamiento tendencial.

Componente estacional E_t: Este comportamiento se presenta cuando la serie de tiempo se encuentra muy marcada para cada estación del año y tiene un movimiento secular, por lo que tiene un comportamiento repetitivo alrededor de la tendencia dentro de un año o menos (trimestrales, mensuales, semanales, etc.) y se relacionan principalmente con las estaciones climáticas. En la Figura 2.2 se observa la serie de tiempo de temperaturas máximas diarias que presenta la región de Valparaíso, durante enero del 2012 y diciembre del 2013. Se puede apreciar que sigue un comportamiento estacional.

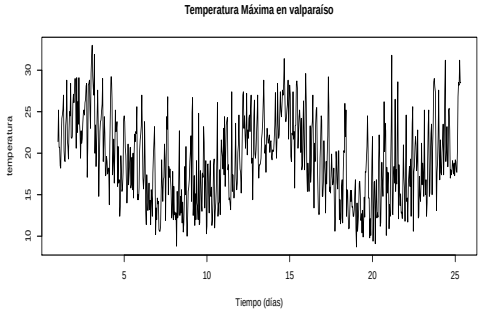


Figura 2.2. Serie de tiempo que sigue un comportamiento estacional.

Componente cíclico C_t: Este comportamiento al igual que la componente estacional se presenta cuando la serie de tiempo se encuentra muy marcada para cada estación del año y el movimiento oscila alrededor de la tendencia. La diferencia que tiene la componente estacional de la cíclica es que en esta última los movimientos oscilatorios son en periodos de tiempo mayores que un año. En la Figura 2.3 se observa una serie de tiempo con componente cíclico.

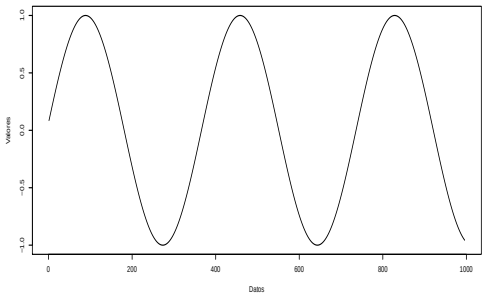


Figura 2.3. Serie de tiempo que sigue un comportamiento cíclico.

Componente aleatorio A_t: Este comportamiento se presenta cuando la serie de tiempo muestra oscilaciones no sistemáticas, es decir, movimientos irregulares, las cuales se debe principalmente a las variaciones aleatorias o repentinas de la serie de tiempo. En la Figura 2.4 se observa la serie de tiempo anual de huelgas en Estados Unidos aproximadamente durante el periodo 1951-1980. Se percibe que los registros fluctúan erráticamente sobre un nivel de variación lenta (Bustamante, 2013).

Figura 2.4. Serie de tiempo que sigue un comportamiento aleatorio.

Tipos de modelos en series de tiempo

Una serie de tiempo puede ser representada en función de los cuatro componentes antes mencionados, conformándose 3 tipos de modelos, de la siguiente manera:

- Aditivo: Suma de cada componente de manera independiente.

$$X_t = T_t + E_t + C_t + A_t \ , \tag{2.2.1}$$

  donde X_t es la serie observada en instante t.

- Multiplicativo: Multiplicación de todos los componentes, relacionándose cada uno de ellos.

$$X_t = T_t * E_t * C_t * A_t \tag{2.2.2}$$

- Mixto: Combinación entre el modelo aditivo y multiplicativo.

$$X_t = (T_t + E_t + C_t) * A_t \tag{2.2.3}$$

2.2.2. Análisis de la tendencia en series de tiempo

Como ya se ha mencionado anteriormente, la representación gráfica entrega indicios del comportamiento que sigue una serie de tiempo, por consiguiente la primera noción sobre la presencia de tendencia T_t en una serie se obtendrá de su representación gráfica.

La tendencia en series de tiempo puede ser detectada y estimada o simplemente eliminada, la elección de alguna de estas dependerá de como se represente en la serie de tiempo.

2.2.3. Métodos de suavizamientos

Un método comúnmente utilizado para suavizar la tendencia de una serie de tiempo y describirla de mejor manera es a través de la aplicación de filtros sobre los datos, es decir, utilizar una función matemática en los valores de la serie, obteniéndose una nueva serie que filtra o suaviza los efectos ajenos a la tendencia (estacionalidad, efectos aleatorios). Los filtros comúnmente utilizados son:

- Promedio móvil: Calcula a través del promedio el valor de una serie en el tiempo deseado, con valores anteriores y posteriores a este. La fórmula se denota como:

$$u_t = \frac{1}{M} \sum_{j=0}^{M-1} z_{i+j} \tag{2.2.4}$$

  La señal de entrada es z_{i+j} y u_i es la señal de salida suavizada (tendencia estimada) y M es el número de puntos en el promedio.

  Si se tiene una serie de tiempo definida como $Z = z_1, z_2, z_3, \ldots, z_t$, el estimador de promedio móvil será el siguiente:

$$\frac{\sum(\text{n términos más próximos})}{M} = \text{valor deseado}$$

- Suavizamiento exponencial simple: El procedimiento de este método es generar una nueva serie de tiempo, pero con promedios calculados de manera exponencial, es decir, que cada valor calculado son en base a las observaciones anteriores pero dando ponderaciones. La fórmula se denota como:

$$\hat{Z}_{t+1} = \alpha Z_t + \alpha(1-\alpha) Z_{t-1} + \alpha(1-\alpha)^2 Z_{t-2} + \ldots + \alpha(1-\alpha)^{n-1} Z_{t-(n-1)} \tag{2.2.5}$$

Escrito de manera más simple:

$$\hat{Z}_{t+1} = \alpha Z_{t+1} + (1-\alpha)\hat{Z}_t \ , \tag{2.2.6}$$

donde Z_t es el valor de la serie en el tiempo t, $\hat{Z}_{t+1}$ es el valor a predecir en el tiempo $t+1$, $\hat{Z}_t$ es el valor estimado en el tiempo t y α es el valor de suavizamiento, el cual se encuentra entre 0 y 1.

- Suavizamiento exponencial doble: Realiza el mismo cálculo que el suavizamiento exponencial simple, pero utilizando los valores obtenidos de éste. La fórmula se denota como:

$$\hat{\hat{Z}}_{t+1} = \alpha \hat{Z}_{t+1} + (1-\alpha)\hat{\hat{Z}}_t \ , \tag{2.2.7}$$

donde $\hat{\hat{Z}}_{t+1}$ es el valor de la predicción en el tiempo $t+1$, $\hat{Z}_t$ es el valor del suavizamiento exponencial simple en el tiempo t, $\hat{\hat{Z}}_t$ es el valor del segundo suavizamiento en el tiempo t y finalmente α es la constate de suavizamiento.

En el caso de que la predicción de la serie de tiempo sea dentro de un periodo más futuro, se debe tener en consideración una interpolación lineal que contemple el componente tendencia. La fórmula se denota como:

$$\bar{Z}_{t+j} = \alpha_t + j * (b_t) \tag{2.2.8}$$

$$a_t = 2 * \hat{Z}_t - \hat{\hat{Z}}_{t+1} \tag{2.2.9}$$

$$b_t = \left(\frac{\alpha}{1-\alpha}\right) * (\hat{Z}_t - \hat{\hat{Z}}_{t+1}) \ , \tag{2.2.10}$$

donde $\bar{Z}_{t+j}$ es el valor pronosticado con tendencia en el periodo $t+j$, α_t es el valor de origen del modelo en el tiempo t, b_t la pendiente de tendencia en el tiempo t y finalmente j es la cantidad de periodos a pronosticar ($j = 1, 2, 3, ...$).

2.2.4. Estimación de la tendencia a través de regresión

Cuando la tendencia T_t se presenta muy marcada en una serie de tiempo resulta importante la aplicación de métodos capaces de estimar dicha tendencia. Existen diferentes métodos para la estimación, uno de ellos es utilizar un modelo de regresión lineal, cuando la serie de tiempo presente una tendencia lineal. Si la serie se presenta con una tendencia no lineal se pueden utilizar otros tipos de regresiones; cuadrática, logística, exponencial, no paramétrica, entre otras.

El ajuste de tendencia debe ser utilizado cuando la serie de tiempo no incluya un componente estacional. Un modelo que solamente adopta la componente tendencia es aquel en donde la observación al tiempo t, está dada por (Chatfield et al, 2003,

citado por Bustamante, 2013, pág 21) :

$$X_t = m_t + \varepsilon_t \ , \tag{2.2.11}$$

donde m_t denota la componente de tendencia (T_t) en el periodo t y ε_t denota un término de error aleatorio (A_t) con media cero.

Las formas matemáticas que más se utilizan para estimar la tendencia, son presentadas a continuación: (Bustamante, 2013, pág 21):

- Lineal
$$X_t = \alpha + \beta t + \varepsilon_t \tag{2.2.12}$$
- Exponencial:
$$X_t = \alpha \, exp(\beta t)\varepsilon_t \tag{2.2.13}$$
- Logística:
$$X_t = \frac{\alpha}{1 + \beta \, exp(\lambda t)}\varepsilon_t \tag{2.2.14}$$
- Gompertz:
$$X_t : \alpha \, exp(\beta \, exp(\lambda t))\varepsilon_t \tag{2.2.15}$$

La tendencia lineal $T_t = \alpha + \beta t$, es la más sencilla de analizar, donde α y β son constantes. El coeficiente β representa el incremento medio en X por unidad de tiempo. Expresado de distinta forma (como otros autores lo prefieren denominar) la tendencia del parámetro β, se representa como la pendiente del modelo lineal. Con respecto a esta última interpretación, la tendencia es el cambio en el nivel medio por unidad de tiempo. La tendencia en la ecuación $T_t = \alpha + \beta t$ se interpreta como una función determinística del tiempo y se conoce como tendencia lineal global. Es importante recordar que (Bustamante, 2013, pág 21):

- Si $\beta > 0$, existe crecimiento lineal
- Si $\beta < 0$, existe una declinación durante un largo periodo
- Si $\beta = 0$, no existe crecimiento o declinación durante un extenso periodo, siendo una serie sin tendencia.

Un procedimiento apropiado para estimar α y β es a través de mínimos cuadrados (Bowerman et al, 2009, citado por Bustamante, 2013, pág 21);

En la Figura 2.5 se presentan datos sobre desviaciones medias anuales de la temperatura en grados centígrados, durante el periodo 1900-1997 (Shumway y Stoffer, 2010). La linea roja que se muestra en la Figura 2.5 representa la linea de ajuste, obtenida de la regresión lineal aplicada sobre los datos.

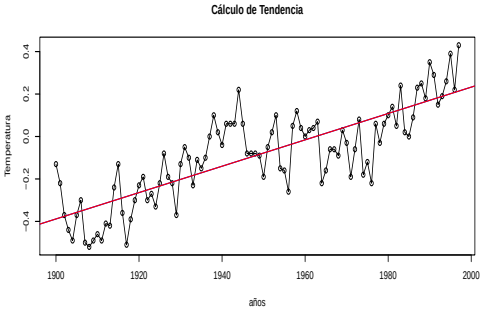


Figura 2.5. Linea de ajuste sobre una serie de tiempo con tendencia lineal.

La ecuación de tendencia exponencial se utiliza cuando la tendencia manifiesta un porcentaje de crecimiento casi constante, y se debería utilizar cuando la gráfica de logaritmo natural de X_t contra t parece casi lineal. Si

$$X_t = \alpha \, exp(\beta t)\varepsilon_t, \tag{2.2.16}$$

entonces

$$log \, X_t = log \, \alpha + \beta t + log \, \varepsilon_t \ , \tag{2.2.17}$$

donde $exp(\beta)$ es la tasa de crecimiento porcentual por unidad de tiempo, mientras que α es el valor de la tendencia en el instante t (Bustamante, 2013). Los coeficientes $log \, \alpha$ y β pueden ser estimados a través de un modelo de regresión lineal de $log \, X_t$ contra t (Hildebrand et al, 1998, citado por Bustamante, 2013, pág 22).

En la Figura 2.6 se generan a través del software R-project datos que siguen una tendencia exponencial.

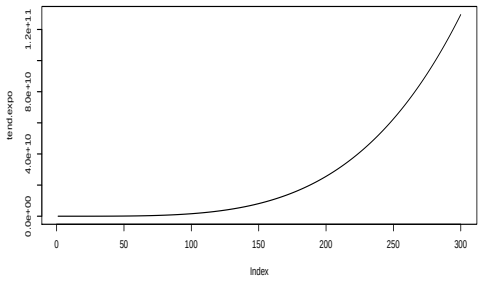


Figura 2.6. Tendencia exponencial.

Las ecuaciones de tendencia logística y de tendencia Gompertz se caracterizan por presentar tendencias similares con forma de S y no se pueden transformar en un modelo de regresión, para su desarrollo existen métodos numéricos más complicados, "por ejemplo el método de Newton-Raphson, o las variantes de este: el método de Newton-Raphson modificado y el método de la secante"(Hildebrand et al, 1998, citado por Bustamante, 2013, pág 22).

2.2.5. Estimación no paramétrica de la tendencia

En la practica, puede ocurrir que la estimación de una tendencia no pueda ser realizada a través de una regresión paramétrica, por lo que se debe proceder a otro tipo de método. La regresión no paramétrica permite estimar el valor promedio de una variable respuesta en función de una o más variables predictoras, al igual que la regresión paramétrica (Härdle, 1992). En algunos casos esta relación no se comporta de manera lineal, por este motivo "la aplicación de un modelo paramétrico pierde firmeza en comparación con las técnicas de regresión no paramétrica"(Reina y Olaya, 2013). El procedimiento que realiza la regresión no paramétrica consiste en modelar el comportamiento de un conjunto de datos sin asumir previamente una función conocida. (Bowman y Azzalini, 1997).

Diversos autores, como por ejemplo Nadaraya (1964), Watson (1964), Priestley y Chao (1972), Cleveland (1979) y Gasser y Müller (1984), han propuesto estimadores denominados como suavizadores, los cuales estiman funciones de regresión no paramétrica. La forma general de representar la ecuación de una regresión no paramétrica es (Meza Santa Cruz, 2013):

$$y_i = f(x_i) + \varepsilon_i \ , \quad i = 1, ..., n \tag{2.2.18}$$

donde y es la variable respuesta aleatoria, f es la función de regresión, x la variable

independiente y finalmente ε los errores del modelo con media cero y varianza común σ^2. Por lo general la función f es no lineal en x, siendo frecuente tratarla como una relación no lineal, aproximando f a un polinomio de orden m (Meza Santa Cruz, 2013):

$$f(x) = \beta_0 + \beta_1 x + ... + \beta_{m-1} x^{m-1} \tag{2.2.19}$$

La técnica de regresión no paramétrica suele ser aplicada en diversas partes del análisis de regresión, por ejemplo en la construcción de modelos. De hecho, la suavización spline es una de las herramientas más versátil para la construcción de modelos (Meza Santa Cruz, 2013).

2.2.5.1. Método de suavizamiento spline

Autores como Wegman y Wright (1983) y Green y Silverman (2000) describen un spline como una "delgada tira de madera", la cual "se puede hacer pasar a través de un conjunto de puntos gracias a que posee una determinada energía de deflexión"(Barrientos et al., 2007). La construcción de un spline tiene como objetivo hallar una aproximación al comportamiento de una función, que resulta desconocida y de la cual sólo se conoce un conjunto de puntos (Julio et al., 2002).

Según Isern y Cuesta (2013) un modelo de regresión spline consiste en un ajuste por partes, donde cada una de ellas es una región del campo de variación de la variable explicativa, en la que se ajusta un modelo de regresión polinómica (en general de bajo orden) y que están unidos en los extremos ("nodos") para dar continuidad a la curva. A continuación se define un modelo spline, considerando una sola variable regresora, y un polinomio de grado 1, de la siguiente manera (Isern y Cuesta, 2013):

$$y_i = f(x_i) + \varepsilon_i = \beta_0 + \beta_1 x_i + \sum_{k=1}^{K} \beta_{1k} (x_i - N_k)_+ + \varepsilon_i \quad , \tag{2.2.20}$$

donde las expresiones $(x_i - N_k)_+$ son funciones de "bases truncadas"; las mismas toman el valor $(x_i - N_k)_+$ si $x_i > N_k$ y cero en otro caso.

A medida que incrementa la cantidad de nodos, el ajuste se vuelve más "rugoso", lo que implica que pueden estar "sobre-ajustandos" los datos, debido a que las pequeñas fluctuaciones observadas podrían deberse, por ejemplo al azar y no responder a una causa justificada, por el contrario un modelo con pocos nodos puede "sub-ajustar" los datos, existiendo la posibilidad de no captar saltos o movimientos propios de la verdadera estructura subyacente. Finalmente es importante mencionar que un buen ajuste dependerá en gran medida de la ubicación de los nodos más que de su cantidad (Isern y Cuesta, 2013).

2.3. Autocorrelación y correlograma

En series de tiempo una herramienta comúnmente utilizada para el análisis de independencia entre las observaciones es el coeficiente de autocorrelación. Este permite determinar si los valores pasados influyen en los futuros, es decir, identifica si los valores que toma una variables en el tiempo son o no independientes entre sí. Cuando sean independientes un valor determinado no dependerá de los valores anteriores, sin embargo cuando sean dependientes será todo lo contrario. Es importante también destacar que dicha herramienta entrega información con respecto a los componentes y comportamiento de una serie de tiempo (tendencia, estacionalidad, ciclicidad y la condición de estacionariedad). Existen dos formas de medir la independencia de las variables, las cuales se presentan a continuación.

2.3.1. Autocorrelación

2.3.1.1. Función de autocorrelación simple (ACF)

La función de autocorrelación simple (ACF) mide la correlación entre dos observaciones de una serie de tiempo, las cuales se encuentran separadas por h periodos (Z_t, Z_{t+h}). Por ejemplo, se tiene una serie de tiempo de la siguiente manera:

$$Z_1, Z_2, \ldots\ldots, Z_{t-1} Z_t, Z_{t+1}, Z_{t+2}, \ldots\ldots,$$

En este caso la función de autocorrelación simple determina la correlación entre Z_1 y Z_2, Z_2 y Z_3 y así sucesivamente. La fórmula de esta función se define como:

$$\rho_h = corr(Z_t, Z_{t+h}) = \frac{cov(Z_t, Z_{t+h})}{\sigma Z_t \, \sigma Z_{t+h}} \tag{2.3.1}$$

Cuando:

- $\rho_h = 0$, no existe correlación lineal.
- $\rho_h = 1$, existe correlación lineal positiva.
- $\rho_h = -1$, existe correlación lineal negativa.

La función de autocorrelación simple (ACF) tiene las siguientes propiedades:

- $\rho_0 = 1$
- $-1 \leq \rho_h \leq 1$
- simetría $\rho_h = \rho - h$

El ACF también puede ser expresado de la siguiente forma:

$$\tau_h = \frac{\sum_{t=1}^{n-h}(Z_t - \bar{Z})(Z_{t+h} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2} \ , \tag{2.3.2}$$

donde τ_h es el coeficiente de autocorrelación simple para un retraso de h periodos, $\bar{Z}$ es la media de los valores en la serie de tiempo, Z_t es la observación en el periodo t, Z_{t+h} es la observación en el periodo $t+h$ y n es el número total de periodos.

2.3.1.2. Función de autocorrelación parcial (PACF)

La función de autocorrelación parcial (PACF), al igual que la función de autocorrelación simple (ACF) se utiliza para determinar la correlación entre dos observaciones de una serie de tiempo. La diferencia está en que la autocorrelación parcial no considera la dependencia que se generan por los retardos intermedios existentes entre las variables. La fórmula de esta función se denota como:

$$\pi_h = corr(Z_t, Z_{t+h}|Z_{t+1}, Z_{t+2}, ..., Z_{t+h-1}) \tag{2.3.3}$$

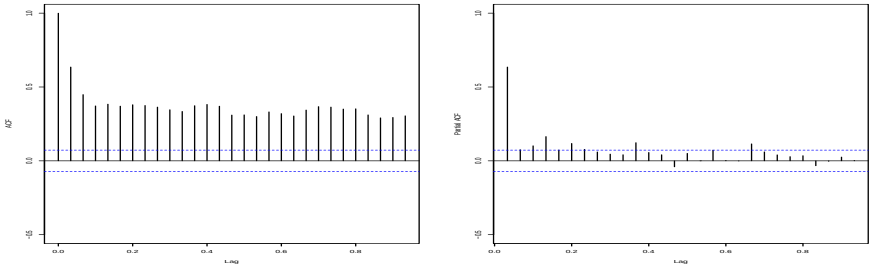
Calculada mediante el coeficiente de regresión:

$$Z_t = \phi_{1h}Z_{t-1} + \phi_{2h}Z_{t-2} + \phi_{3h}Z_{t-3} + ... + \phi_{hh}Z_{t-h} + \omega_t \ , \tag{2.3.4}$$

donde ϕ_{jh} representa el parámetro j de la regresión y ω_t el término de error con media cero y no correlacionado con $Z_{t-1},\ Z_{t-2}, ..., Z_{t-h}$.

2.3.2. Correlograma

Los coeficientes de autocorrelación simple y parcial obtenidos para los diferentes periodos, pueden ser representados gráficamente, denominándose correlograma o función de autocorrelación. En ella se grafican las autocorrelaciones para varios retardos h de tiempo de una serie. A continuación, utilizando el ejemplo de la Figura 2.2, se obtiene el correlograma de los coeficientes de autocorrelación simple (ACF) y parcial (PACF).



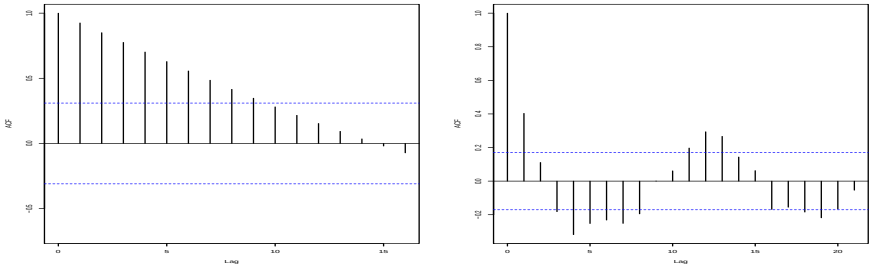
(a) Correlograma ACF **(b)** Correlograma PACF

Figura 2.7. Correlograma ACF y PACF, de las temperaturas máximas diarias en Valparaíso, durante el periodo 2012-2013.

El correlograma ACF y PACF de la Figura 2.7 muestra en la escala horizontal de la parte baja del gráfico cada uno de los retardos de tiempo que resultan de interés (h=0.1, 0.2,..., etc.), mientras que la escala vertical del lado izquierdo muestra el posible rango de un coeficiente de autocorrelación (-1 a 1). Cada barra representa la autocorrelación obtenida mediante las fórmulas presentadas anteriormente. Bajo el supuesto de independencia τ_h, los valores siguen una distribución $N\left(\frac{-1}{n}, \frac{1}{n}\right)$, por consiguiente las bandas de confianza del correlograma quedan dadas por $\frac{-1}{n} \pm \frac{2}{\sqrt{n}}$.

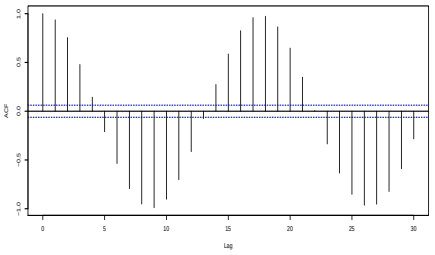
Como ya fue mencionado anteriormente, a través del correlograma se pueden identificar los distintos componentes de una serie de tiempo. En una serie con tendencia los coeficientes de autocorrelación comienzan a caer lentamente hacia cero (nunca anulándose). Sin embargo en una serie estacionaria al estar presente la variación estacional muestra correlaciones significativas en los periodos $s, 2s, 4s, ..,$. Por ejemplo si la serie de tiempo tiene valores con estacionalidad mensual, tendrá coeficiente significativo cada 12 retardos. En una serie de tiempo cíclica el comportamiento que seguirán los coeficientes de autocorrelación se comportarán de manera similar a la estacionalidad, con la diferencia de que los valores del ACF decaen en forma de seno o coseno, dependiendo de cómo se comporten los valores de la serie.

En la Figura 2.8 se presentan tres correlogramas de series de tiempo con componente tendencia, estacionalidad y ciclicidad. La imagen (a) presenta el correlograma con componente tendencial lineal: $y = 2t + 4$, la imagen (b) presenta el correlograma con componente estacional de una serie de tiempo de pasajeros (incluido en el paquete del software R-project) y finalmente la imagen (c) muestra el correlograma con componente cíclico: $y = sen\left(\frac{2 * pi * t}{12}\right)$.



(a) ACF: tendencia

(b) ACF: estacionalidad



(c) ACF: ciclicidad

Figura 2.8. Correlogramas ACF con componente tendencial, estacional y cíclico.

2.4. Series de tiempo estacionarias y no estacionarias

Una serie de tiempo puede ser estacionaria o no estacionaria, por consiguiente la identificación de alguna de estas resulta importante para determinar si la serie cambia o no en el tiempo.

Una serie de tiempo es estacionaria cuando no hay cambios sistemáticos en ella, es decir, es estable en el tiempo, por ende la media y varianza son constantes y no hay tendencia en la serie. Gráficamente los valores de la serie tienden a oscilar alrededor de la media y la variabilidad con respecto a esa media también permanece constante en el tiempo. Sin embargo, una serie de tiempo es no estacionaria cuando sigue una tendencia y/o variabilidad que cambian en el tiempo. Los cambios en la media determinan una tendencia creciente o decreciente a largo plazo, por lo que la serie no va a oscila alrededor de un valor constante.

2.5. Procesos estocásticos en series de tiempo

Para comprender el concepto de un proceso estocástico, se procede a continuación a mostrar su definición.

Definición 1.1 Un proceso estocástico es una colección de variables aleatorias $\{X_t : t \in T\}$ parametrizada por un conjunto T, llamado espacio parametral, en donde las variables toman valores en un conjunto S llamado espacio de estados (Rincón, 2012).

Explicado de manera más sencilla, un proceso estocástico se describe como una sucesión de datos que evolucionan a medida que transcurre el tiempo. Una manera fácil de describir un procesos estocástico es definiendo sus momentos, específicamente el primer y segundo momento, correspondientes a la media y autocovarianza, respectivamente.

Función de medias: En cada una de las distribuciones marginales se obtiene la media a través del tiempo, denotándose como:

$$\mu_t = E(Z_t), \forall\, t \tag{2.5.1}$$

Función de varianzas: En cada una de las distribuciones marginales se obtiene la varianza a través del tiempo, denotándose como:

$$\sigma_t^2 = Var(Z_t) = E[(Z_t - \mu)^2], \forall\, t \tag{2.5.2}$$

Función de autocovarianza: Es la covarianza descrita en dos tiempos distintos, denotándose como:

$$\gamma_h = Cov(Z_t, Z_{t+h}) = E[(Z_t - \mu)(Z_{t+h} - \mu)], \forall\, t \tag{2.5.3}$$

Función de autocorrelación: Es la correlación presente en distintos tiempos, denotándose como:

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\gamma_h}{\sigma^2} \tag{2.5.4}$$

Los procesos estocásticos también son expresados como:

Función de media muestral:

$$\bar{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t \tag{2.5.5}$$

Función de varianza muestral:

$$Var(Z) = c_0 = \frac{1}{n-1}\sum_{t=1}^{n}(Z_t - \bar{Z})^2 \tag{2.5.6}$$

Función de autocovarianza muestral:

$$c_j = \frac{1}{n}\sum_{t=1}^{n-j}(Z_t - \bar{Z})(Z_{t+j} - \bar{Z}) \tag{2.5.7}$$

Función de autocorrelación muestral:

$$\rho_j = \frac{c_j}{c_0} \tag{2.5.8}$$

2.5.1. Procesos estocásticos estacionarios

Los procesos estocásticos pueden ser clasificados como estacionarios en sentido débil o estricto. Un proceso aleatorio será estrictamente estacionario cuando la distribución conjunta de n observaciones no varía con respecto al tiempo cuando se desplazan las variables en h periodos. Por ejemplo, si se tienen n observaciones; $z_1, z_2, z_3, \ldots\ldots, z_n$, en los tiempos $t = 1, 2, 3, \ldots, n$, se puede decir que es la misma que la que siguen los n datos en los tiempos $t = 1+h, 2+h, 3+h, \ldots., n+h$, donde h es cualquier salto en el tiempo,

$$f(z_1, z_2, z_3, ..., z_n) = f(z_{1+h}, z_{2+h}, z_{3+h}, ..., z_{n+h}) \tag{2.5.9}$$

Un proceso aleatorio será estacionario en sentido débil, cuando la media, la varianza y la autocovarianza del proceso estocástico sea constante y finito en el tiempo, denotándose como:

Función de medias:

$$E(Z_t) = \mu \ , \ \forall \ t \ \epsilon \ \mathbb{Z} \tag{2.5.10}$$

Función de varianzas:

$$\sigma_t^2 = Var(Z_t) = \sigma^2 \ , \ \forall \ t \ \epsilon \ \mathbb{Z} \tag{2.5.11}$$

Función de autocovarianza:

$$\gamma(t, t+h) = Cov(Z_t, Z_{t+h}) = \gamma(h) \ , \ \forall \ t, h \ \epsilon \ \mathbb{Z} \tag{2.5.12}$$

Puede ocurrir que el proceso estocástico sea invertible cuando exista una función lineal $H(.)$ y un proceso $\omega \sim$ Ruido Blanco en cada t, denotándose de la siguiente manera:

$$Z_t = H(Z_{t-1}, Z_{t-2}, Z_{t-3}, ...) + \omega_t \tag{2.5.13}$$

2.5.2. Proceso ruido blanco

Una serie de tiempo se dirá un proceso de ruido blanco, cuando sus variables sean independientes e idénticamente distribuidas (iid), con:

Media:

$$E(\omega_t) = 0 \ , \ \forall \ \ t \tag{2.5.14}$$

Varianza

$$Var(\omega_t) = \sigma_\omega^2 \ , \ \forall \ t \tag{2.5.15}$$

Covarianza:

$$Cov(\omega_t, \omega_{t+h}) = 0 \ , \ \forall \ k \neq 0 \tag{2.5.16}$$

Las variables observadas cuando sigan una distribución normal, se denotarán como:

$$\omega_t \sim N(0, \sigma^2)$$

Finalmente se puede decir que algunos procesos de series de tiempo siguen un modelo ruido blanco, cuando tenga los parámetros antes vistos, denotándose como:

$$\omega_t \sim RB(0, \sigma_\omega^2)$$

Por consiguiente, un proceso de ruido blanco es estacionario si:

La autocovarianza (γ_h) es:

$$\gamma_h = \begin{cases} \sigma_\omega^2 & \text{si } h = 0 \\ 0 & \text{si } h > 0 \end{cases}$$

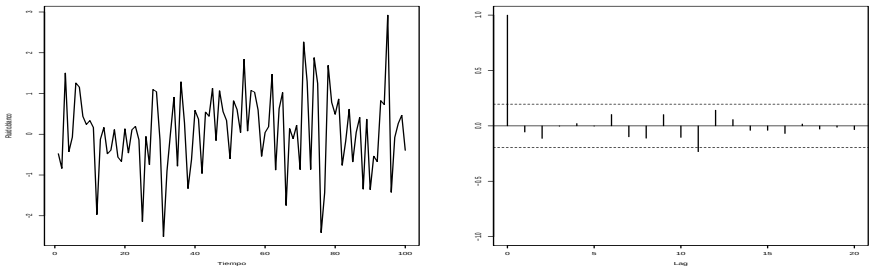
La función de autocorrelación (ρ_h) es:

$$\rho_h = \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } h > 0 \end{cases}$$

Un proceso de ruido blanco puede ser observado mediante el proceso gaussiano (normal), donde sus variables aleatorias son normales estándar (media cero y varianza constante) e independientes, denotándose como:

$$\omega_t \sim iid \, N(0, \sigma_\omega^2) \ ,$$

y gráficamente quedando de la siguiente manera:



(a) Ruido blanco **(b)** ACF: ruido blanco

Figura 2.9. Serie de tiempo y ACF de un proceso de ruido blanco gaussiano.

Se tiene que las funciones para los valores que muestra la Figura 2.9, están dados por:

$$\mu_\omega = 0 \quad \gamma_h = \begin{cases} \sigma_\omega^2 & \text{si} h = 0 \\ 0 & \text{si } h \neq 0 \end{cases} \quad \rho_h = \begin{cases} 1 & \text{si } h = 0 \\ 0 & \text{si } h \neq 0 \end{cases}$$

Capítulo 3

Modelos de series de tiempo

Los modelos de series temporales se desarrollan sobre la afirmación de que las series de tiempo "tienen una historia estadística recurrente particular que puede ser modelada y explotada para fines de pronóstico" (del Alcázar y Quiróz, 1996). En este capítulo se examinan los modelos univariables, los cuales se caracterizan por emplear únicamente la serie de tiempo de la variable a predecir. Las variaciones que presenta la serie de tiempo se expresan como una función de términos autorregresivos (valores pasados de la variable) y términos de promedios móviles (errores contemporáneos y pasados) (del Alcázar y Quiróz, 1996). Estos tipos de modelos se denotan como:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \omega_t - \theta_1 \omega_{t-1} - ... - \theta_q \omega_{t-q} \tag{3.0.1}$$

3.1. Modelos univariables

Los modelos univariables de predicciones lineales procuran modelar el comportamiento de una serie de tiempo, a través de una función lineal y en función de la propia variable. Los modelos AR, MA, ARMA, ARIMA y SARIMA son los que aplican este método y se caracterizan por ser procesos estocásticos que tienen una metodología común (Pérez, 2014, pág 14).

3.1.1. Modelo Autorregresivo AR(p)

Un proceso autorregresivo se representa como un modelo en donde la variable endógena $\{x_t\}$ es determinada por sus valores pasados $\{x_{t-1}, x_{t-2}, \ldots, x_{t-p}\}$ y por un ruido blanco $\{\omega_t\}$. El propósito de este modelo es utilizar los valores pasados de la serie como variables independientes, para estimar el valor siguiente. La representación del proceso autorregresivo de orden p $\{x_t \rightarrow AR(p)\}$, se denota a través de

la siguiente ecuación (Shumway y Stoffer, 2010):

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} +, ..., + \phi_p x_{t-p} + \omega_t \ , \tag{3.1.1}$$

donde x_t es estacionario, $\phi_1, \phi_3, \phi_3, ..., \phi_p$ son parámetros del modelo y por ende constantes ($\phi_p \neq 0$) y $\omega_t \sim RB(0, \sigma_\omega^2)$. La media de x_t en la ecuación 3.1.1 es cero. Si la media μ de x_t no es cero, se reemplaza x_t pór $x_t - \mu$ en la ecuación 3.1.1, es decir,

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + ... + \phi_p(x_{t-p} - \mu) + \omega_t \ , \tag{3.1.2}$$

o escribir

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} +, ..., \phi_p x_{t-p} + \omega_t \ , \tag{3.1.3}$$

donde $\alpha = \mu(1 - \phi_1 - ... - \phi_p)$.

Existen algunas dificultades técnicas al utilizar el modelo de la ecuación 3.1.3, ya que los regresores $x_{t-1}, \ldots, x_{t-p}$ son componentes aleatorios, mientras que z_t se supone que es fijo. Una manera de solucionar el problema es utilizando el operados Backward shift. El modelo autorregresivo de la ecuación 3.1.1 utiliza el operador de retardo B, que retrasa el índice de tiempo, es decir,

$$Bx_t = x_{t-1} \tag{3.1.4}$$

El polinomio autorregresivo se define de forma generalizada, de la siguiente manera:

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p \tag{3.1.5}$$

Por ende, se puede reescribir el modelo autorregresivo AR(p) como:

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p) x_t = \omega_t \ , \tag{3.1.6}$$

lo que implica que

$$\phi(B) x_t = \omega_t \tag{3.1.7}$$

En un proceso autorregresivo el orden p se determina por la cantidad de observaciones retrasadas, por ende, si un modelo $AR(p)$ tiene un valor $p = 1$, significa que es un proceso autorregresivo de orden uno, si $p = 2$ entonces será un proceso autorregresivo de orden dos y así sucesivamente.

Otra forma de escribir el modelo autorregresivo AR(p) de manera abreviada es:

$$x_t = \sum_{j=1}^{p} \phi_j \, x_{t-j} + \omega_t \tag{3.1.8}$$

Considerando el operador de retardo B, la ecuación queda como:

$$x_t = \sum_{j=1}^{p} \phi_j \, B^j \, x_t + \omega_t \tag{3.1.9}$$

Una característica importante de los procesos autorregresivos de orden p es que siempre son invertibles, debido a que $\sum_{j=1}^{p} |\phi_j| < \infty$. Por otra parte, la condición de estacionariedad se cumplirá sí y sólo sí las raíces del polinomio autorregresivo se encuentran fuera del círculo unitario $(\phi_p B = 0) > 1$.

De los procesos autorregresivos se pueden obtener las siguientes propiedades:

- Esperanza: $E(x_t) = 0$
- Varianza: $Var(x_t) = \gamma_0 = \phi_1\gamma_1 + ... + \phi_p\gamma_p$
- Autocovarianza: $Cov(x_t, x_{t-h}) = \gamma_h = \phi_h\gamma_{h-1} + ... + \phi_p\gamma_{h-p}\,, \;\; h > 0$
- Función de autocorrelación simple (ACF):

$$\rho_h = \phi_1\rho_{t-1} + ... + \phi_p\rho_{t-p}\,, \;\; h > 0$$

- Función de autocorrelación parcial(PACF):

$$\phi_{hh} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \rho_h \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \rho_h \end{vmatrix}}\,, \qquad h > 0$$

Modelo autorregresivo de orden uno AR(1)

La variable x_t en un modelo AR(1) se explica por su valor pasado, además del ruido blanco, denotandose de la siguiente manera (Shumway y Stoffer, 2010):

$$x_t = \phi_1 x_{t-1} + \omega_t \ , \quad \omega_t \sim RB(0, \sigma_\omega^2) \tag{3.1.10}$$

Considerando el operador de retardo B, expresado en la fórmula 3.1.4, el modelo autorregresivo AR(1) se puede reescribir como:

$$(1 - \phi_1 B)x_t = \omega_t \tag{3.1.11}$$

Para verificar si se cumple la condición de estacionariedad para x_t, se debe comprobar que las raíces del polinomio de la ecuación 3.1.11 están dentro del círculo unitario. Para comprobar lo dicho anteriormente, se debe realizar lo siguiente:

$$\begin{aligned} 1 - \phi_1 B &= 0 \\ \phi_1 B &= 1 \\ B &= \frac{1}{\phi_1} \ , \end{aligned}$$

donde se debe cumplir que $|B| = \left|\frac{1}{\phi_1}\right| > 1$. De esta manera el modelo autorregresivo AR(1) cumple con la condición de estacionariedad.

Las propiedades para el modelo autorregresivo AR(1) quedan de la siguiente manera:

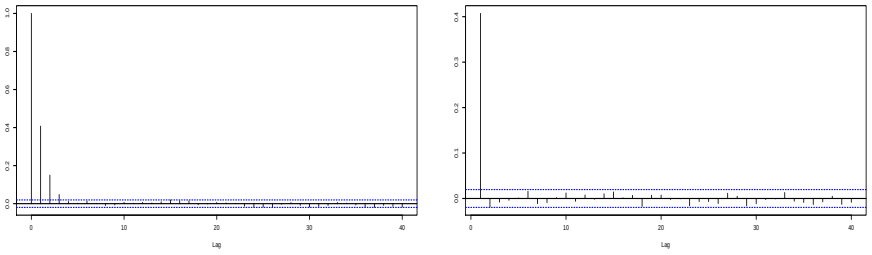
- Esperanza: $E(x_t) = 0$
- Varianza: $Var(x_t) = \gamma_0 = \frac{\sigma_w^2}{1 - \phi_1^2} \ , \quad t = 1, 2, ...,$
- Autocovarianza: $Cov(x_t, x_{t-h}) = \gamma_h = \phi_1 \frac{\sigma_1^2}{1 - \phi_1^2} = \phi_1 \gamma_0 \ , \quad t = 1, 2, ...,$
- Función de autocorrelación simple (ACF):

$$\rho_h = \frac{\gamma_h}{\gamma_0} = \phi_1 \rho_{h-1} \ , \quad h = 1, 2, ...,$$

- Función de autocorrelación parcial (PACF):

$$\phi_{hh} = \begin{cases} \rho_1 = \phi_1 & h = 1 \\ 0 & h > 0 \end{cases}$$

En la Figura 3.1 se presenta la simulación de un proceso autorregresivo de orden 1 AR(1), con un $\phi = 0.4$. La Figura 3.1a presenta el correlograma ACF y la imagen (b) presenta el correlograma PACF.



(a) Correlograma ACF **(b)** Correlograma PACF

Figura 3.1. ACF y PACF de un proceso autorregresivo: $(1 - 0.4B)x_t = \omega_t$.

Como se puede apreciar en la Figura 3.1a el ACF muestra que las correlaciones decaen de forma exponencial y disminuyen rápidamente hacia cero y en la Figura 3.1b el PACF se observa que sólo la primera autocorrelación es distinta de cero, identificándose de esta manera el orden del proceso autorregresivo AR(1).

Es de gran importancia tener en consideración que sólo la primera autocorrelación en un AR(1) es distinta de cero y puede ser positiva o negativa dependiendo del valor de ϕ_1.

Modelo autorregresivo de orden dos AR(2)

La variable x_t en un modelo AR(2) se explica por sus dos valores pasados, además del ruido blanco, denotándose de la siguiente manera:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \omega_t \ , \ \ \omega_t \sim RB(0, \sigma_\omega^2) \tag{3.1.12}$$

Considerando el operador de retardo B, expresado en la fórmula 3.1.4, el modelo autorregresivo AR(2) se puede reescribir como:

$$(1 - \phi_1 B - \phi_2 B^2)x_t = \omega_t \tag{3.1.13}$$

Para verificar la condición de estacionariedad para x_t, se realiza lo siguiente:

$$1 - \phi_1 B - \phi_2 B^2 = 0$$

$$B_1, B_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \ ,$$

donde se debe cumplir que $|B_1, B_2| = \left| \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} \right| > 1$. De esta manera el modelo autorregresivo AR(2) cumple con la condición de estacionariedad.

Las propiedades para el modelo autorregresivo AR(2), quedan expresadas de la siguiente manera:

- Esperanza: $E(x_t) = 0$
- Varianza: $Var(x_t) = \gamma_0 = \frac{\sigma_\omega^2 + 2\phi_1\phi_2\gamma_1}{1 - \phi_1^2 - \phi_2^2}$
- Autocovarianza: $cov(x_t, x_{t-h}) =$

$$\gamma_h = \begin{cases} \frac{\sigma_\omega^2 + 2\phi_1\phi_2\gamma_1}{1 - \phi_1^2 - \phi_2^2} & h = 0 \\ \frac{\phi_1\gamma_0}{1 - \phi_2} & h = 1 \\ \phi_1\gamma_{h-1} + \phi_1\gamma_{h-2} & h > 1 \end{cases}$$

- Función de autocorrelación simple (ACF):

$$\rho_h = \begin{cases} 1 & h = 0 \\ \frac{\phi_1}{1 - \phi_2} & h = 1 \\ \frac{\phi_1^2 - \phi_2^2 + \phi_2}{1 - \phi_2} & h = 2 \\ \phi_1\rho_{h-1} - \phi_2\rho_{h-2} & h > 2 \end{cases}$$

- Función de autocorrelación parcial (PACF):

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{bmatrix}}{\begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = \phi_2$$

Es de gran importancia tener en consideración que sólo las dos primeras autocorrelaciones en un modelo AR(2), son distintas de cero y pueden ser positivas o negativas dependiendo del valor de ϕ_1 y ϕ_2.

3.1.2. Modelo media móvil MA(q)

Un proceso de media móvil se representa como un modelo expresado por la suma de los errores pasados y actuales de manera ponderada. La representación del proceso media móvil de orden q $\{x_t \rightarrow MA(q)\}$, se denota a través de la siguiente ecuación Shumway y Stoffer (2010):

$$x_t = \omega_t - \theta_1\omega_{t-1} - \theta_2\omega_{t-2} - ... - \theta_q\omega_{t-q} \tag{3.1.14}$$

Donde se pueden observar q retardos en el modelo, $\theta_1, \theta_2, ..., \theta_q$ $(\theta_q \neq 0)$ son parámetros del modelo y $\omega_t \sim RB(0, \sigma_\omega^2)$.

Considerando el operador de retardo B, de la misma manera que en AR(p), se retrasa la serie un periodo hacia atrás, es decir,

$$B\omega_t = \omega_{t-1} \tag{3.1.15}$$

El polinomio de media móvil se define de forma generalizada, de la siguiente manera:

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q \tag{3.1.16}$$

Por ende, se puede reescribir el modelo media móvil MA(q) como:

$$(1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)x_t = \omega_t \,, \tag{3.1.17}$$

lo que implica que

$$x_t = \theta(B)\omega_t \tag{3.1.18}$$

En un proceso media móvil el orden q depende de la cantidad de errores anteriores, por ende, si un modelo MA(q) tiene un valor $q = 1$, significa que es un proceso de media móvil de orden uno, si $q = 2$ entonces será un proceso media móvil de orden dos y así sucesivamente.

Otra forma de escribir el modelo media móvil MA(q), de manera abreviada es:

$$x_t = \sum_{j=1}^{q} \theta_j \omega_{t-j} - \omega_t$$

Considerando el operador de retardo B, la ecuación se denota como:

$$x_t = \sum_{j=1}^{q} \theta_j B^j \omega_t - \omega_t$$

Una característica importante de los procesos de media móvil de orden q es que siempre son estacionarios, debido a que $1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2 < \infty$. Por otra parte la condición de invertibilidad se cumplirá sí y sólo sí las raíces del polinomio de media móvil, se encuentran fuera del círculo unitario, $\theta_q B = 0 > 1$.

De los procesos de media móvil se pueden obtener las siguientes propiedades:

- Esperanza: $E(x_t) = 0$
- Varianza: $Var(x_t) = \gamma_0 = (1 + \theta_1^2 + ... + \theta_q^2)\sigma_\omega^2$
- Autocovarianza: $cov(x_t, x_{t-h}) =$

$$\gamma_h = \begin{cases} (-\theta_h + \theta_1\theta_{h+1} + ... + \theta_q\theta_{q-h})\sigma_\omega^2 & h = 1, 2, ..., q \\ 0 & h > q \end{cases}$$

- Función de autocorrelación simple (ACF):

$$\rho_h = \begin{cases} \dfrac{(-\theta_h + \theta_1\theta_{h+1} + ... + \theta_q\theta_{q-h})}{1 + \theta_1^2 + ... + \theta_q^2} & h = 1, 2, ..., q \\ 0 & h > q \end{cases}$$

- Función de autocorrelación parcial(PACF):

$$\phi_{hh} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \rho_h \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & \rho_h \end{vmatrix}}, \quad h > 0$$

Modelo media móvil de orden uno MA(1):

La variable x_t en un modelo MA(1) es explicada por el error actual y el que está inmediatamente antes que éste, además del ruido blanco, denotándose de la siguiente manera (Shumway y Stoffer, 2010) :

$$x_t = \omega_t - \theta_1 \omega_{t-1} \ , \quad \omega_t \sim RB(0, \sigma_\omega^2) \tag{3.1.19}$$

Considerando el operador de retardo B, expresado en la fórmula 3.1.15, el modelo de media móvil MA(1), se puede reescribir como:

$$(1 - \theta_1 B) x_t = \omega_t \tag{3.1.20}$$

Para verificar si se cumple la condición de invertivilidad para x_t, se debe comprobar que las raíces del polinomio de la ecuación 3.1.20, están fuera del círculo unitario. Para comprobar lo dicho anteriormente, se debe realizar lo siguiente:

$$\begin{aligned} 1 - \theta_1 B &= 0 \\ \theta_1 B &= 1 \\ B &= \frac{1}{\theta_1} \quad , \end{aligned}$$

donde se debe cumplir que $B = \dfrac{1}{\theta_1} > 1$. De esta manera el modelo de media móvil MA(1) cumple con la condición de invertibilidad.

Las propiedades para el modelo de media móvil MA(1) quedan de la siguiente manera:

- Esperanza: $E(x_t) = 0$
- Varianza: $Var(x_t) = \gamma_0 = (1 + \theta_1^2)\sigma_\omega^2$
- Autocovarianza: $cov(x_t, x_{t-h}) =$

$$\gamma_h = \begin{cases} (1 + \theta_1^2)\sigma_\omega^2 & h = 0 \\ -\theta_1 \sigma_\omega^2 & h = 1 \\ 0 & h > 1 \end{cases}$$

- Función de autocorrelación simple (ACF):

$$\rho_h = \begin{cases} \dfrac{-\theta_1}{1 + \theta_1^2} & h = 1 \\ 0 & h > 1 \end{cases}$$

- Función de autocorrelación parcial (PACF):

$$\phi_{hh} = \frac{-\theta_1^h(1-\theta_1^2)}{1-\theta_1^{2(h+1)}}\ , \quad h > 0$$

Modelo media móvil de orden dos MA(2):

La variable x_t en un modelo MA(2) se explica por sus dos errores, que están inmediatamente antes del error actual, además del ruido blanco, denotándose de la siguiente manera:

$$x_t = \omega_t - \theta_1\omega_{t-1} - \theta_2\omega_{t-2}\ , \quad \omega_t \sim RB(0, \sigma_\omega^2) \tag{3.1.21}$$

Considerando el operador de retardo B, expresado en la fórmula 3.1.14, el modelo de media móvil MA(2), se puede reescribir como:

$$(1 - \theta_1 B - \theta_2 B^2)x_t = \omega_t \tag{3.1.22}$$

Para verificar la condición de invertibilidad para x_t se realiza lo siguiente:

$$1 - \theta_1 B - \theta_2 B^2 = 0\ , \tag{3.1.23}$$

donde se debe cumplir que $|B_1, B_2| = \left|\frac{\theta_1 \pm \sqrt{\theta_1^2 + 4\theta_2}}{-2\theta_2}\right| > 1$. De esta manera el modelo de media móvil MA(2), cumple con la condición de invertibilidad.

Las propiedades para el modelo de media móvil MA(2), quedan de la siguiente manera:

- Esperanza: $E(x_t) = 0$
- Varianza: $Var(x_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2)\sigma_\omega^2$
- Autocovarianza: $cov(x_t, x_{t-h}) =$

$$\gamma_h = \begin{cases} (1+\theta_1^2+\theta_2^2)\sigma_\omega^2 & h = 0 \\ -\theta_1(1-\theta_2)\sigma_\omega^2 & h = 1 \\ -\theta_2\sigma_\omega^2 & h = 2 \\ 0 & h > 2 \end{cases}$$

- Función de autocorrelación simple (ACF):

$$\rho_h = \begin{cases} \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2} & h=1 \\ \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2} & h=2 \\ 0 & h>2 \end{cases}$$

- Función de autocorrelación parcial(PACF):

$$\begin{aligned} \phi_{11} &= \rho_1 \\ \phi_{22} &= \frac{\rho_2-\rho_1^2}{1-\rho_1^2} \\ \phi_{33} &= \frac{\rho_1^3-\rho_1\rho_2(2-\rho_2)}{1-\rho_2^2-2\rho_1^2(1-\rho_2)} \\ &\vdots \end{aligned}$$

3.1.3. Proceso autorregresivo AR(p) y media móvil MA(q)

Los procesos autorregresivos AR(p) y de media móvil MA(q) se encuentran relacionados entre sí, ya que un proceso autorregresivo de orden finito puede equivaler a un proceso de media móvil de orden infinito. Para que se entienda de mejor manera lo dicho anteriormente se procede a mostrar un ejemplo.

Al escribir el modelo autorregresivo de orden uno AR(1), en términos del operador de retardo B, se tiene que:

$$\begin{aligned} (1-\phi B)x_t &= \omega_t \ , \\ x_t &= \frac{\omega_t}{(1-\phi B)} \end{aligned}$$

Por convergencia de una serie geométrica se tiene que:

$$\begin{aligned} x_t &= \frac{\omega_t}{(1-\phi B)} = (1+\phi B+\phi^2 B^2+...)\omega_t \\ x_t &= \omega_t+\phi\omega_{t-1}+\phi^2\omega_{t-2}+... \end{aligned}$$

3.1.4. Modelo autorregresivo de media móvil (ARMA)

Un proceso autorregresivo de media móvil, es un modelo compuesto por los dos procesos básicos antes mencionados (AR(q) y MA(q)). El propósito de este proceso

es modelar series parcialmente autorregresivas y de medias móviles. Estos modelos al ser combinados adquieren la abreviación mediante las siglas ARMA(p,q). Al ser una composición de los dos modelos básicos, el valor de la serie x_t es una combinación lineal de las q perturbaciones más recientes y de las p observaciones más recientes de sí misma. La representación del modelo ARMA(p,q), se denota a través de la siguiente ecuación (Shumway y Stoffer, 2010):

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \omega_t - \theta_1 \omega_{t-1} - ... - \theta_q \omega_{t-q} \ , \tag{3.1.24}$$

donde p y q son los respectivos parámetros autorregresivos y de medias móviles, además $\phi_p \neq 0$, $\theta_q \neq 0$, $\sigma_\omega^2 > 0$ y $\omega_t \sim RB(0, \sigma_\omega^2)$. Si x_t tiene una media μ no nula, se debe poner $\alpha = \mu(1 - \phi_1 - ... - \phi_p)$ y se vuelve a escribir el modelo como:

$$x_t = \alpha + \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + \omega_t - \theta_1 \omega_{t-1} - ... - \theta_q \omega_{t-q} \tag{3.1.25}$$

En un modelo ARMA(p,q) los pronósticos dependerán de los valores pasados de x_t como de los valores actuales y pasados de los errores. En estos procesos cuando $q = 0$, se presentará un modelo AR(p), mientras que cuando $p = 0$, se obtendrá un modelo MA(q).

Considerando el operador de retardo B en los modelos autorregresivo y de media móvil, expresado en las ecuaciones 3.1.4 y 3.1.15 y los polinomios de éstos en las ecuaciones 3.1.5 y 3.1.16, se reescribe el modelo autorregresivo de media móvil ARMA(p,q) de la siguiente manera:

$$(1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p) x_t = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q) \omega_t \ , \tag{3.1.26}$$

lo que implica que

$$\phi(B) x_t = \theta(B) \omega_t \tag{3.1.27}$$

Si el proceso es estacionario su representación MA(∞) es:

$$x_t = \frac{\phi_q(B)}{\theta_p(B)} \omega_t \ , \tag{3.1.28}$$

entonces

$$x_t = \omega_t + \varphi_1 \omega_{t-1} + \varphi_2 \omega_{t-2} + \varphi_3 \omega_{t-3} + ... \tag{3.1.29}$$

Si el proceso es invertible su representación AR(∞) es:

$$\frac{\phi_p(B)}{\theta_q(B)} x_t = \omega_t \ , \tag{3.1.30}$$

entonces

$$x_t = \omega_t + \pi_1 x_{t-1} + \pi_2 x_{t-2} + \pi_3 x_{t-3} + ... \tag{3.1.31}$$

La condición de estacionariedad en un modelo ARMA(p,q) está dada por la parte AR(p) y la condición de invertibilidad está dada por la parte MA(q). Es importante recordar que un modelo MA(q) es siempre estacionaria y un modelo AR(p) es siempre invertible.

Finalmente se debe tener en consideración, que los modelos ARMA(p,q) siempre van a compartir las características de los modelos AR(p) y MA(q), ya que se encuentra compuesta por ambas estructuras a la vez.

3.1.5. Modelo autorregresivo integrado de media móvil (ARIMA)

Un proceso autorregresivo integrado de media móvil ARIMA(p,d,q), es considerado un proceso lineal no estacionario, por ende, se utiliza en series donde la media y la varianza no son constantes en el tiempo.

Estos tipos de procesos son denominados procesos integrados, por consiguiente para hacer que la serie de tiempo sea estacionaria, se debe diferenciar d veces, para finalmente aplicar a la serie diferenciada un modelo ARMA(p,q) (Villavicencio, 2010).

De todo lo mencionado anteriormente se deduce que una serie de tiempo x_t sigue un proceso ARIMA(p,d,q), si al diferenciarla sigue un proceso ARMA(p,q). La representación del modelo ARIMA(p,d,q) se denota a través de la siguiente ecuación:

$$\nabla^d x_t = \phi_1 \nabla^d x_{t-1} + ... + \phi_p \nabla^d x_{t-p} + \omega_t - \theta_1 \omega_{t-1} - ... - \theta_q \omega_{t-q} \,, \tag{3.1.32}$$

donde $\omega_t \sim RB(0, \sigma_\omega^2)$, p es el número de términos autorregresivos, d el número de veces que la serie es diferenciada para que se vuelva estacionaria y q el número de términos de la media móvil invertible.

La diferenciación en un modelo ARIMA(p,d,q) se representa como:

- Primera diferenciación: $\nabla x_t = x_t - x_{t-1}$
- Segunda diferenciación $\nabla^2 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$
- Tercera diferenciación: $\nabla^3 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) - (x_{t-2} - x_{t-3}) \cdots$

Así sucesivamente. La serie de tiempo se diferenciará cuantas veces sea necesario para que ésta se vuelva estacionaria, por consiguiente si una serie se diferencia dos veces $d = 2$. Cuando $d = 0$ se considera un proceso estacionaria, en cambio cuando $d > 0$ se considera un proceso no estacionario diferenciado d veces.

Considerando el operador de retardo B, el modelo ARIMA(p,d,q) se reescribe como:

$$\phi_p(B)(1 - B)^d x_t = \theta_q(B)\omega \,, \tag{3.1.33}$$

donde $\phi_p(B)$ y $\theta_q(B)$ son los polinomios autorregresivos y de medias móviles. Las raíces de las ecuaciones $\phi_p(B) = 0$ y $\theta_q(B) = 0$, se encuentran fuera del círculo unitario.

Finalmente se debe considerar que en los modelos ARIMA(p,d,q), se pueden identificar ausencias tanto de términos de medias móviles como de términos autorregresivos, siendo expresados mediante las siglas ARI(p,d) y IMA(d,q).

Capítulo 4

Identificación, estimación, diagnóstico y pronóstico sobre los distintos modelos en estudio

4.1. Prueba de raíz unitaria

Antes de aplicar cualquier modelo (AR, MA, ARMA, ARIMA) sobre una serie de tiempo, se debe identificar si es estacionarias, es decir, comprobar que tanto la media como la varianza sean constantes en el tiempo. Si bien, mediante los gráficos se puede identificar la no estacionariedad en una serie de tiempo, para comprobar formalmente esta condición se deben aplicar test de estacionariedad.

Existen diversos test para identificar estacionariedad, entre los mas utilizados se encuentan el test Dickey Fuller (DF), Dickey Fuller Aumentada (ADF) y Kwiatkowski Phillips Schmidt Shin (KPSS), los cuales verifican si existe o no raíz unitaria en las series de tiempo, por consiguiente un proceso es no estacionario cuando exista raíz unitaria, en cambio cuando no tenga raíz unitaria se hablará de un proceso estacionario.

- Test Dickey Fuller (DF): Dickey y Fuller(1979, 1981) diseñaron un procedimiento capaz de comprobar la presencia o no de raices unitarias en una serie de tiempo. Las ecuaciones presentadas por estos autores , se definen como (Gujarati, 1997, pág 703):

$$\triangle Y_t = \delta Y_{t-1} + u_t \tag{4.1.1}$$

$$\triangle Y_t = \alpha + \delta Y_{t-1} + u_t \tag{4.1.2}$$

$$\triangle Y_t = \alpha + \beta T + \delta Y_{t-1} + u_t \tag{4.1.3}$$

La diferencia entre las tres regresiones presentadas, involucran la presencia de componentes deterministicos; intercepto (drift) y tendencia (T). La ecuación 4.1.1 representa un modelo puramente aleatorio, la ecuación 4.1.2 añade un intercepto o término a la deriva drift y la ecuación 4.1.3 incluye intercepto y un término de tendencia.

El parámetro de interés en las 3 regresiones es δ, por lo que si $\delta = 1$, entonces la serie Y_t no es estacionaria, por el contrario si $\delta < 1$ se dice que la serie Y_t es estacionaria.

El problema que presenta el test DF es que "asume que el proceso estocástico subyacente a los datos sigue un AR(1)", es por este motivo que cuando el proceso siga otro esquema se obtendrán complicaciones con respecto a la estimación de la regresión auxiliar que comprende el test, arrojando incumplimiento con respecto a la condición que presenta el ruido blanco para los residuos. Por este motivo se dispone del test ampliado de Dickey-Fuller (ADF).

- Test Dickey Fuller Aumentada(ADF): Este test es considerado como la versión aumentada del test Dickey Fuller y es utilizado porque permite considerar otros esquemas de autocorrelación, añadiendo términos diferenciados de la variable dependiente x. Para este test el modelo que se propone es el siguiente:

$$\nabla Y_t = \alpha + \beta T + \delta Y_{t-1} + \gamma \sum_{i=1}^{\rho} Y_{t-i} + e_t \tag{4.1.4}$$

Es recomendable utilizar el estadístico ADF cuando la prueba DF no pueda corregir la correlación serial en los residuos. Los retardos $\gamma \sum_{i=1}^{\rho} \triangle Y_{t-i}$ del test ADF aseguran que los residuos sean ruido blanco.

Las hipótesis a contrastar para el test DF y ADF, estan dadas por:

$$\begin{aligned} H_0 &: \delta = 0 \quad \text{Tiene raíz unitaria (No estacionaria)} \\ H_1 &: \delta \neq 0 \quad \text{No tiene raíz unitaria (Estacionaria)} \end{aligned}$$

- Test Kwiatkowski Phillips Schmidt Shin (KPSS): El test KPSS consiste en contrastar la hipótesis nula, la cual indica que la serie de tiempo es estacionaria en torno a una tendencia lineal o en torno a nivel, en cambio la hipótesis alternativa indica todo lo contrario, es decir, se aprecia raíz unitaria en la serie de tiempo y por consiguiente la no estacionariedad. La formulación del test

KPSS se expresa como (Casquero, 2001):

$$x_t = \gamma\, D_t + \mu_t$$
$$\mu_t = \mu_{t-1} + \omega_t \quad \omega_t \sim RB(0, \sigma_\omega^2)$$

Donde D_t está definida por la parte determinística (por la tendencia, constante o ambas) y μ.

Las hipótesis a contrastar para el test KPSS son:

$$H_0 : \sigma_\omega^2 = 0 \qquad \text{Es Estacionaria}$$

$$H_1 : \sigma_\omega^2 > 0 \quad \text{No estacionaria}$$

Cuando se cumpla la hipótesis nula, μ_t será una constante, por ende la serie de tiempo resulta estacionaria.

En la práctica es recomendable utilizar el test ADF y KPSS, por el motivo de que la prueba ADF suele mostrar muchos falsos positivos, es decir, se encuentra un poco sesgada por manifestar a través de la hipótesis nula que la serie de tiempo presenta raíz unitaria (no es estacionaria) cuando realmente no la hay, en cambio el test KPSS tiene la ventaja de tener la hipótesis nula contraria a la del test ADF, es decir, en la hipótesis nula el test KPSS prueba que la serie de tiempo no presenta raíz unitaria(es estacionaria). Finalmente se puede concluir que cuando los dos test presenten diferencias, es recomendable optar por el test KPSS.

4.2. Identificación del modelo

Los distintos modelos estadísticos estudiados, por lo general requieren de una secuencia de pasos para la óptima obtención de resultados. Puede ocurrir la necesidad de redefinir alguna de estas secuencias, por el hecho de obtener resultados insatisfactorios, por consiguiente los diferentes pasos a seguir si bien son sucesivos, de acuerdo a los resultados de cada uno de éstos puede surgir un retroceso. La identificación del modelo es el primer paso a seguir y consta de estudiar la función de autocorrelación simple y parcial a través del correlograma de cada uno éstos (ACF y PAC), los cuales ayudarán a determinar qué modelo podría ser el más adecuado.

Las características principales del comportamiento del correlograma ACF y PACF para cada uno de los procesos que fueron estudiados anteriormente, son explicados en la siguiente tabla:

Tabla 4.1. Características principales del correlograma ACF y PACF, en los modelos AR(p), MA(q) y ARMA(p,q).

Proceso	Función de autocorrelación simple (ACF)	Función de autocorrelación parcial (PACF)
Modelo autorregresivo AR(p)	Las correlaciones decaen exponencialmente hacia cero	Las correlaciones se cortan cuando los retardos sean superiores a p
Modelo media móvil MA(q)	Las correlaciones se cortan cuando los retardos sean superiores a q	Las correlaciones decaen exponencialmente hacia cero
Modelo autorregresivo de media móvil ARMA(p,q)	Las correlaciones decaen exponencialmente hacia cero cuando se corta después del retardo $q-p$	Las correlaciones decaen exponencialmente hacia cero cuando se corta después del retardo $p-q$

4.3. Estimación de parámetros

Identificado el modelo a utilizar se debe ir al siguiente paso, el cual considera la estimación de los parámetros. Para esto se presentan dos posibles métodos; el de los momentos o el de máxima verosimilitud. Si se pone como ejemplo un modelo ARMA(p,q) (expresada en la ecuación 3.1.24), los parámetros a estimar serían ϕ, μ, θy σ_ω^2

Métodos para la estimación de los parámetros:

- Método de momentos: Este método consiste en realizar un reemplazo a cada momento, es decir, a la media, varianza y función de autocorrelación simple, por el estimador de momentos. Por ejemplo, en un modelo AR(p) el estimador de momento más simple para la media resulta ser $\widehat{\mu} = \overline{x}_t$ y para ϕ se utiliza $\rho_h = \phi_1\rho_{h-1} + ... + \phi_p\rho_{h-p}$, obteniendose las ecuaciones de YULE-WALKER expresada como:

$$\rho_1 = \phi_1 + \phi_2\rho_1 + ... + \phi_p\rho_{p-1}, ..., \rho_p = \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + ... + \phi_p \tag{4.3.1}$$

  Si es reemplazado ρ_h por:

$$\widehat{\rho}_h = \frac{\sum_{t=1}^{n-h}(x_t - \overline{x})(x_{t+h} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2} \ , \tag{4.3.2}$$

  y es resuelto, se pueden obtener $\widehat{\phi}_1, ..., \widehat{\phi}_p$. A continuación de obtener los estimadores de ϕ, se procede a calcular γ_0, de la siguiente manera:

$$\begin{aligned} \gamma_0 &= E(x_t x_t) && (4.3.3)\\ &= E[x_t(\phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + \omega_t)] && (4.3.4)\\ &= \phi_1\gamma_1 + \phi_2\gamma_2 + ... + \phi_p\gamma_p + \sigma_\omega^2 && (4.3.5) \end{aligned}$$

Finalmente se puede estima σ_ω^2 como:

$$\sigma_\omega^2 = \gamma_0(1 - \widehat{\phi}_1\widehat{\rho}_1 - ... - \widehat{\phi}_p\widehat{\rho}_p) \tag{4.3.6}$$

Para obtener los estimadores de θ, se presenta un ejemplo considerando un proceso MA(1), de la siguiente manera:

$$x_t = \omega_t - \theta_1\omega_{t-1}$$

Para calcular el estimador de θ_1, se debe realizar lo siguiente:

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{-\theta_1}{1+\theta_1^2}$$

Resolviendo la ecuación y obteniendo el estimador de θ_1 , se procede a reemplazar ρ_1 por $\widehat{\rho}_1$. Posterior a su resolución se obtiene $\widehat{\theta}_1$, de la siguiente manera:

$$\widehat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4\widehat{\rho}_1^2}}{2\widehat{\rho}_1} \tag{4.3.7}$$

Finalmente se puede estima σ_ω^2 :

$$\sigma_\omega^2 = \frac{\widehat{\gamma}_0}{1 + \widehat{\sigma}_1^2} \tag{4.3.8}$$

- Estimador de máxima verosimilitud: Es uno de los método más utilizados para la estimación de parámetros en los modelos de series de tiempo. Para comprender el procedimiento de este método, se toma como ejemplo un modelo ARMA(p,q), expresada en la ecuación 3.1.24, donde se debe recordar que $\omega_t \sim RB(0, \sigma_\omega^2)$. La densidad de probabilidad condicional de $\omega = (\omega_1, \omega_2, ..., \omega_n)$, se expresa como:

$$P(\omega|\phi, \theta, \mu, \sigma_\omega^2) = (2\pi\sigma_\omega^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_\omega^2}\sum_{i=1}^{n}\omega_t^2\right] \ , \tag{4.3.9}$$

donde $\phi = (\phi_1 + \phi_2 + ... + \phi_n$ y $\theta = (\theta_1 + \theta_2 + ... + \theta_n)$. Si se escribe la ecuación 3.1.24 en función de ω_t :

$$\omega_t = x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - ... - \phi_p x_{t-p} + \theta_1\omega_{t-1} + \theta_2\omega_{t-2} + ... + \theta_q\omega_{t-q} \tag{4.3.10}$$

Sean entonces $x = (x_1, x_2, ..., x_n)$, $x^* = (x_{1-p}, ..., x_{-1}, x_0)$ y $\omega^* = (\omega_{1-q}, ..., \omega_{-1}, \omega_0)$ las condiciones iniciales conocidas. La función log-verosimilitud condicional se

puede escribir como:

$$lnL(\phi, \theta, \mu, \sigma_\omega^2) = -\frac{n}{2}\, ln2\pi\sigma_\omega^2 - \frac{S(\phi, \theta, \mu)}{2\sigma_\omega^2} \ , \tag{4.3.11}$$

donde:

$$S(\phi, \theta, \mu) = \sum_{t=1}^{n} \omega_t^2(\phi, \theta, \mu | x^*, \omega^*, x) \tag{4.3.12}$$

Las cantidades que maximizan la función condicional de máxima verosimilitud en ϕ, θ y μ, serán los estimadores de máxima verosimilitud condicional ($\widehat{\phi}, \widehat{\theta}$ y $\widehat{\mu}$). Cabe mencionar que para encontrar estos estimadores se debe minimizar la ecuación 4.4.12.

Con la obtención de los estimadores se logra conseguir $\widehat{\sigma}_\omega^2$, de la siguiente manera:

$$\widehat{\sigma}_\omega^2 = \frac{S(\widehat{\theta}, \widehat{\mu}, \widehat{\theta})}{d.f} \ , \tag{4.3.13}$$

donde d.f corresponde a la cantidad de términos utilizados en $S(\phi, \theta, \mu)$ menos el numero de parámetros estimados, es decir, d.f serán los grados de libertad.

4.4. Diagnóstico del modelo

Obtenida las estimación de los parámetros del modelo, se procede al diagnóstico, es decir, a la comprobación de ciertos supuestos sobre los residuos (ω_t). El cumplimiento de los supuestos ayudará a indicar si el modelo es adecuado para realizar pronósticos en un futuro.

4.4.1. Supuestos sobre los residuos

Los supuestos en los residuos del modelo deben indicar si éstos siguen un ruido blanco, es decir, deben ser independientes, tener media cero y la varianza con respecto a la media debe ser constante.

Para identificar independencia se debe calcular el ACF y PACF. Si no se observan autocorrelaciones significativas, los residuos cumplen con la independencia. El test que comprueba estadísticamente independencia es el test Ljung-Box, el cual consiste en probar si las autocorrelaciones de los residuos son distintos de cero. Esta prueba se definida como:

$$LB = n(n+2)\sum_{k=1}^{m}(\frac{\widehat{\rho}_k^2}{n-k}) \sim \chi_{(m)}^2 \ , \tag{4.4.1}$$

donde n es el tamaño de la muestra y m la longitud del rezago.

Las hipótesis a contrastar para este test son:

H_0 : Las autocorrelaciones son iguales a cero (independientes)

H_1 : Las autocorrelaciones son distintas de cero (no son independientes)

En la aplicación, si el Q calculado es mayor que el valor Q crítico de la tabla Chi-cuadrado ($Q_{\text{calculado}} > Q_{\text{crítico}}$), no se acepta la hipótesis nula, indicando que por lo menos algunos de los coeficientes de autocorrelación deben ser distintos de cero.

Para identificar media cero y varianza constante, se debe realizar lo siguiente:

- Media cero: Prueba t-student
- Varianza constante: Se debe apreciar el ACF de los residuos al cuadrado e identificar si se observan o no autocorrelaciones significativas. Si no se observan autocorrelaciones significativas la varianza resulta constante (homocedástica). Sin embargo, si las autocorrelaciones resultan significativas, la varianza resulta no constante (heterocedástica). Para este último caso se debe ajustar un modelo heterocedástico que explique la volatilidad de los datos.

Otro supuesto a comprobar es el de normalidad, es decir, identificar si los residuos sigan una distribución normal. Un test de normalidad comúnmente utilizado es el llamado Shapiro-Wilk, el cual consiste en contrastar la normalidad de un conjunto de de datos. Por ejemplo:

$$x_t = x_1, x_2, ..., x_n \ ,$$

donde n debe ser menor o igual a 50 ($n \leq 50$). Las hipótesis a contrastar para este test son:

$$H_0 : x_t \sim \text{Normal}$$
$$H_1 : x_t \nsim \text{Normal}$$

Es importante destacar que el supuesto de normalidad no siempre se cumple, debido a que los datos no siempre se logran distribuir normal. Esto no quiere decir que el modelo no sea el adecuando, ya que el no cumplimiento de normalidad sobre los residuos no resulta relevante.

4.4.2. Criterios de información Akaike y Bayesiana

En la elección del mejor modelo, existe la posibilidad de que más de uno cumpla con los supuestos antes mencionados, por lo que se deben realizar otros procedi-

mientos capaces de seleccionar el mejor modelo. Los criterios de información Akaike y Bayesiana ayudan con esta elección. El procedimiento consiste en identificar el modelo que contenga el menor valor de los dos criterios, ya que de esta manera se estaría indicando que es el modelo más adecuando. A continuación se procede con la explicación de los criterios de información.

- Criterio de información Akaike (AIC): Para este tipo de criterio se utiliza la función de verosimilitud marginal y resulta maximizando una aproximación de la verosimilitud, el cual es integrado sobre los efectos aleatorios (del Carmen García et al., 2014). La expresión de este criterio se denota como:

$$\text{AIC} = -2\, log(\hat{L}) + 2s \ , \tag{4.4.2}$$

  donde L corresponde a la verosimilitud, $log(L)$ son funciones del logaritmo de la verosimilitud y s es la cantidad de parámetros del modelo. Finalmente en este criterio se selecciona el modelo que minimice el valor AIC.

- Criterio de información Bayesiana (BIC): Este tipo de criterio al igual que el Akaike asume un modelo con s parámetros, el cual se ajusta a los datos. La diferencia está en su formulación, la cual se denota como:

$$\text{BIC} = -2\, log(\hat{L}) + s\, log(n) \ , \tag{4.4.3}$$

  donde se aprecia la diferencia del AIC en el segundo término n. Finalmente en este criterio se selecciona el modelo que minimice el valor BIC. Desde un enfoque bayesiana, BIC se utiliza para encontrar, dado los datos el modelo más probable.

  Es importante destacar que estos dos criterios tienen un factor de castigo, ya que incluyen parámetros al modelo, pero el criterio BIC tiene un castigo mayor por el hecho de tener número de parámetros adicionales con respecto al criterio AIC. Por consiguiente, cuando se elige el BIC que presenta el menor valor, se estará escogiendo un modelo con menor número de parámetros en comparación a la selección de un modelo bajo el criterio AIC.

  Finalmente, si bien estos dos criterios ayudan a seleccionar un buen modelo entre varios, no asegura el cumplimiento de los supuesto.

4.5. Pronóstico

La identificación y confirmación de los supuestos conllevan finalmente al pronóstico del modelo, es decir, al conocimiento anticipado de lo que sucederá en un futuro.

En otras palabras a la predicción de una serie con respecto a su evolución. Cuando se realiza el análisis de una serie de tiempo, lo que se pretende es descifrar el comportamiento de los datos, con el objetivo de obtener predicciones a futuro, es decir, con las observaciones pasadas se procura obtener predicciones en un futuro próximo. Por ejemplo, si se tiene una serie de tiempo de la forma:

$$x_t = x_1, x_2, ..., x_n \ ,$$

donde n es el último periodo del que se tiene registro. El valor predicho se representa como $x_{n+\tau}$ donde τ es la cantidad de periodos en el futuro a considerar.

4.5.1. Error de pronóstico

Escoger el mejor modelo en ocasiones puede provocar confusiones, ya que en la práctica existe la probabilidad de tener más de un posible modelo para predecir. Existen diversas herramientas para escogen el modelo más adecuado (además de los criterios de información AIC y BIC) de todos los posibles que existan. Una de estas herramientas es el error de pronóstico.

En toda predicción existe un error de pronóstico, el cual se denota como:

$$e_\tau = x_{n+\tau} - \hat{x}_{n+\tau} \ , \tag{4.5.1}$$

donde $x_{n+\tau}$ serán los valores originales y $\hat{x}_{n+\tau}$ los valores pronosticados.

La raíz del error cuadrático medio (que en inglés se escribe como root mean squared error (RMSE)), es una de las formas de representar el error de pronóstico, el cual mide el promedio de los errores al cuadrado.

A continuación se presenta la ecuación de la raíz del error cuadrático medio (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{\tau=1}^{N} e_\tau^2} \ , \tag{4.5.2}$$

donde N es la cantidad de datos a pronosticar.

Capítulo 5

Análisis de series de tiempo respecto al desplazamiento de la corteza terrestre

En este capítulo se aplicará todo lo presentado anteriormente, utilizando datos reales. Las series de tiempo en estudio serán modeladas con el propósito de obtener predicciones sobre el desplazamiento de la corteza terrestre en Chile.

5.1. Información sobre los datos en estudio

Los datos en estudio fueron adquiridos por el Instituto Geográfico Militar de Santiago (IGM). La información con respecto a estos datos es la siguiente:

- A través del sistema GPS se obtienen datos diarios en centímetros (cm), con respecto al desplazamiento de la corteza terrestre en Chile, durante enero del 2010 hasta diciembre del 2012.
- Existen 22 estaciones GPS de monitoreo en estudio, de las cuales no se cuenta con información completa con respecto a los datos.
- En cada estación GPS de monitoreo se cuenta con tres componentes; componente Norte (DN), componente Este (DE) y componente Altura (DA), que representan las coordenadas del sistema GPS.

Es importante mencionar que todo análisis y desarrollo estadístico es realizado a través del software R-project. Programa gratuito que se puede descargar desde la página: `www.r-project.org`.

5.2. Análisis preliminar

Para tener un registro de la cantidad de datos faltantes, se procede a realizar un análisis preliminar a cada estación de monitoreo, con el propósito de obtener una estación al norte centro y sur del país que contenga la menor cantidad de datos faltantes. En la Tabla 5.1 se presentan los resultados obtenidos.

Tabla 5.1. Resultados obtenidos con respecto a los datos faltantes, en las 22 estaciones de monitoreo GPS en Chile, durante enero del 2010 hasta diciembre del 2012.

Estaciones de monitoreo GPS	Región	Zona	Total datos faltantes
Iquique	Región de Tarapacá (I)	Norte	*183
Atacama	Región de Atacama (III)	Norte	236
Copiapó	Región de Atacama (III)	Norte	200
Valparaíso	Región de Valparaíso (V)	Centro	267
Los Andes	Región de Valparaíso (V)	Centro	209
Rancagua	Región de O'Higgins (VI)	Centro	205
Pichilemu	Región de O'Higgins (VI)	Centro	399
Cauquenes	Región del Maule (VII)	Centro	331
Hualañe	Región del Bío-Bío (VIII)	Centro	568
Concepción	Región del Bío-Bío (VIII)	Centro	*182
Ninhue	Región del Bío-Bío (VIII)	Centro	334
Cobquecura	Región del Bío-Bío (VIII)	Centro	719
Chillan	Región del Bío-Bío (VIII)	Centro	447
Arauco	Región del Bío-Bío (VIII)	Centro	523
Cabrero	Región del Bío-Bío (VIII)	Centro	327
Antuco	Región del Bío-Bío (VIII)	Centro	187
Los Ángeles	Región del Bío-Bío (VIII)	Centro	215
Llafenco	Región de Araucanía (IX)	Sur	*198
Coyhaique	Región de Aysén (XI)	Sur	427
Punta Arenas	Región de Magallanes (XII)	Sur	222
Santiago	Región Metropolitana (XIII)	Centro	253
Niebla	Región de los Ríos (XIV)	Sur	352

Se puede apreciar en la Tabla 5.1 con asterisco (*), las estaciones de monitoreo que presentaron la menor cantidad de datos faltantes en cada zona del país. En esta ocasión fueron Iquique en la zona norte, Concepción en la zona centro y Llafenco en la zona sur. Cabe mencionar que se torna complicado un análisis con la existencia de datos faltantes, por lo que se aplicó la estimación de la media sobre los faltantes, es decir, se obtuvo el promedio del valor anterior y posterior del dato faltante.

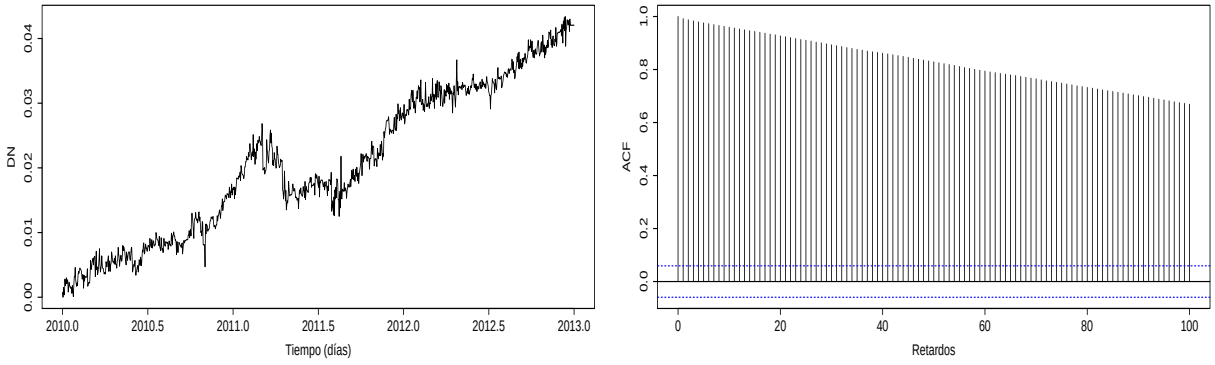
Con la imputación de datos, se obtiene para la estación de monitoreo de cada componente (Norte, Este y Altura) 1.096 datos durante los años 2010 hasta 2012.

5.3. Análisis de estacionariedad sobre las series de tiempo

Se procede con el análisis de estacionariedad sobre cada serie de tiempo en estudio. Es importante recordar que las tres estaciones de monitoreo seleccionadas son: Iquique, Concepción y Llafenco, las cuales se componen de tres series de tiempo cada una, definidas como: componente Norte, Este y Altura.

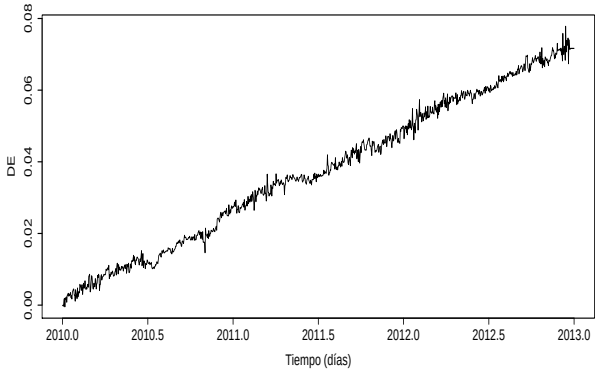
En la Figura 5.1-5.6 se presentan las observaciones diarias del desplazamiento de la corteza terrestre para las componentes Norte, Este y Altura de Iquique, Concepción y Llafenco. En la Figura 5.1a, 5.2a, 5.3a, 5.4a, 5.4b, 5.5a y 5.6a se presentan series de tiempo que muestran una tendencia muy marcada, por lo que a simple vista se podría pensar que son series no estacionarias, ya que la media y varianza no serían constantes en el tiempo. Para comprobar la existencia de estacionariedad en cada serie de tiempo se procede con la aplicación de los test ADF y KPSS.

Estación de monitoreo: Iquique

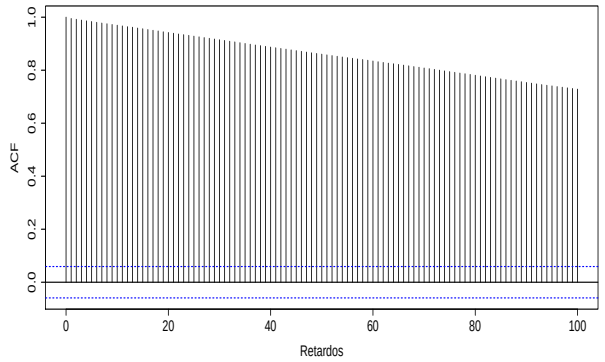


(a) Serie de tiempo componente Norte **(b)** ACF: componente Norte

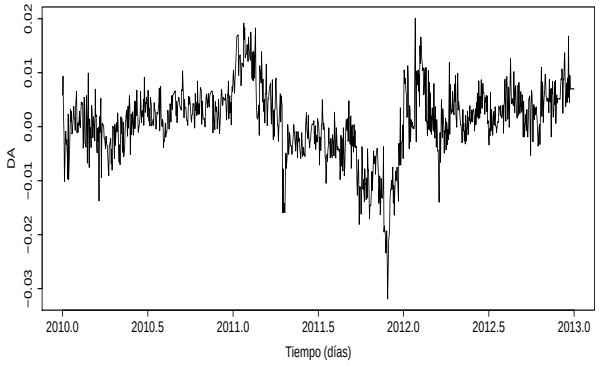
Figura 5.1. Serie de tiempo y ACF de la componente Norte de Iquique.



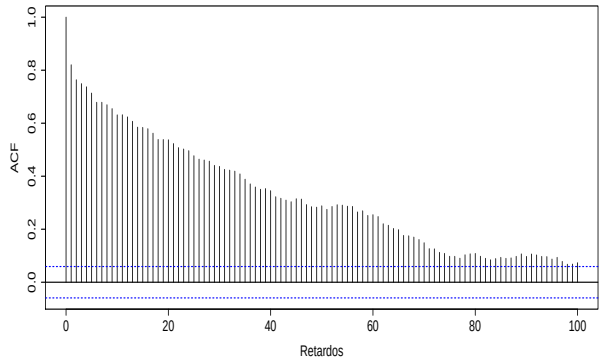
(a) Serie de tiempo componente Este



(b) ACF: componente Este



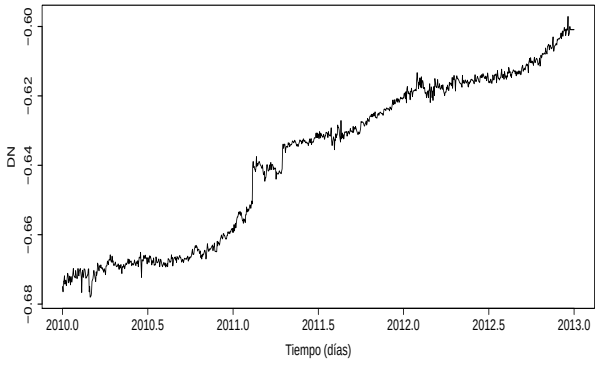
(c) Serie de tiempo componente Altura



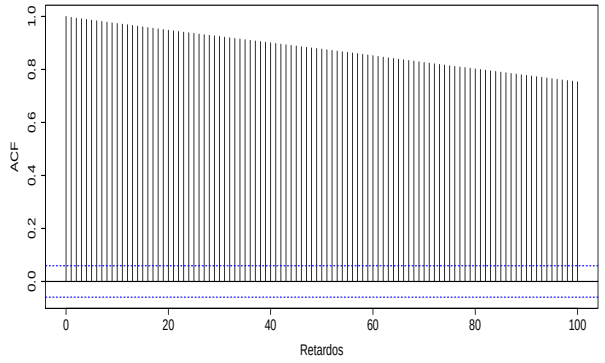
(d) ACF: componente Altura

Figura 5.2. Serie de tiempo y ACF de la componente Este y Altura de Iquique.

Estación de monitoreo: Concepción

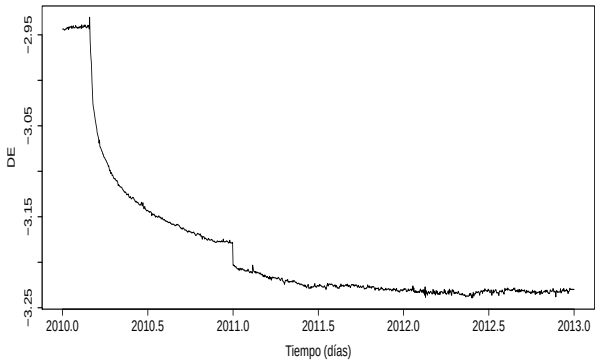


(a) Serie de tiempo componente Norte

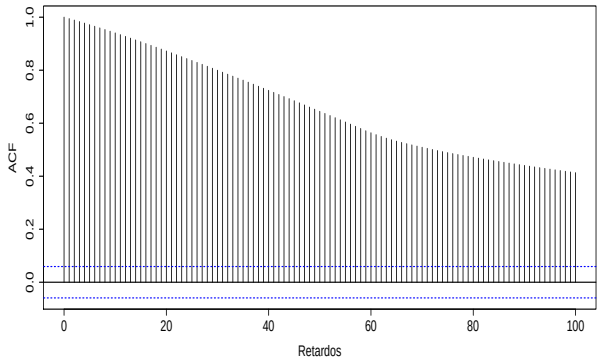


(b) ACF: componente Norte

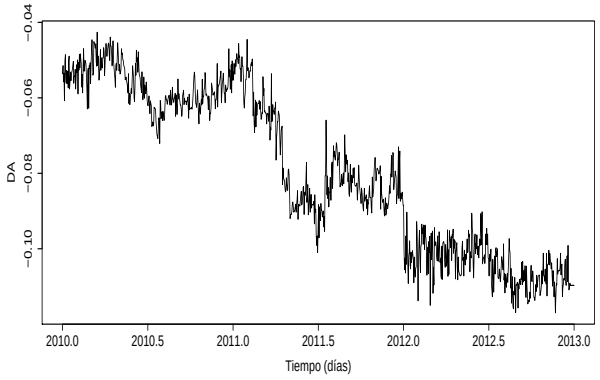
Figura 5.3. Serie de tiempo y ACF de la componente Norte de Concepción.



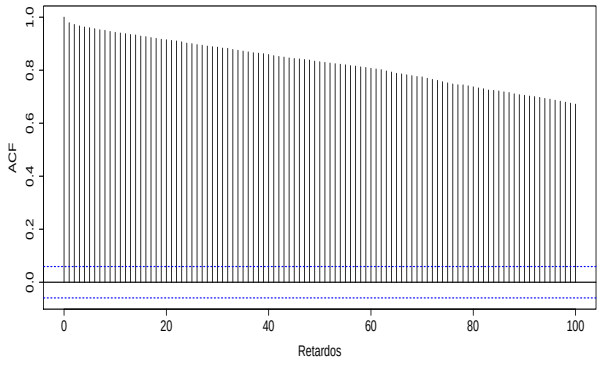
(a) Serie de tiempo componente Este



(b) ACF: componente Este



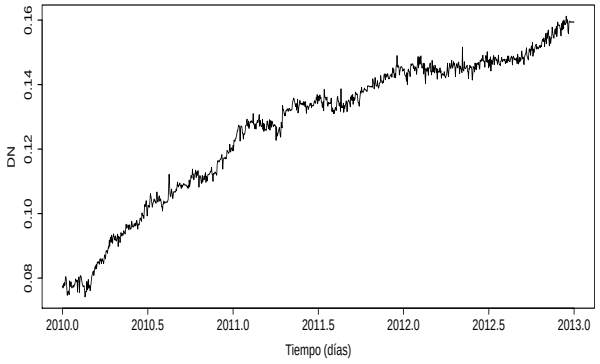
(c) Serie de tiempo componente Altura



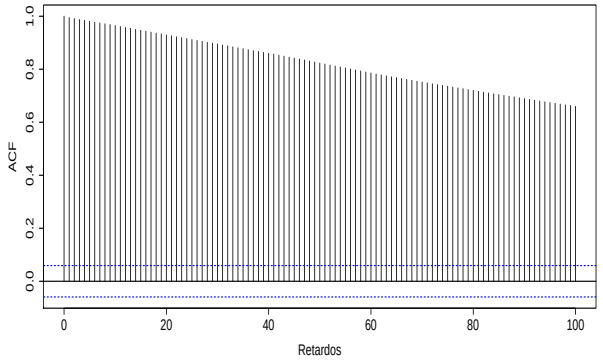
(d) ACF: componente Altura

Figura 5.4. Serie de tiempo y ACF de la componente Este y Altura de Concepción.

Estación de monitoreo: Llafenco

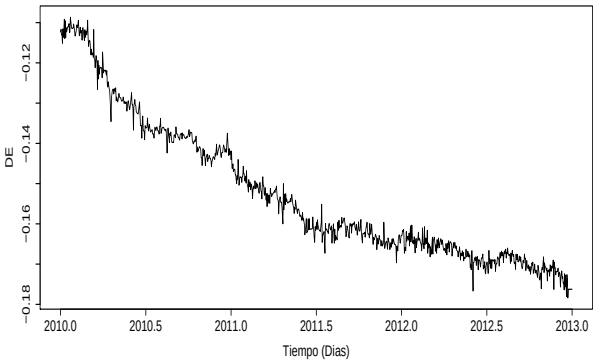


(a) Serie de tiempo componente Norte

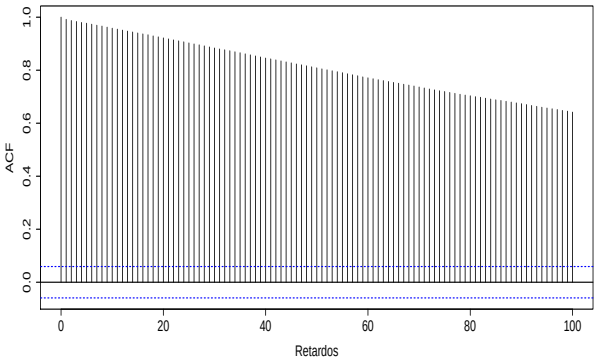


(b) ACF: componente Norte

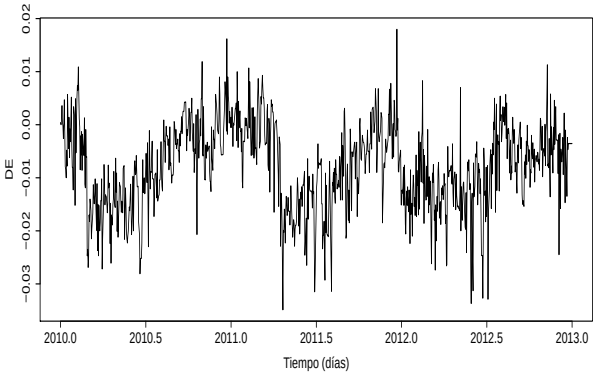
Figura 5.5. Serie de tiempo y ACF de la componente Norte de Llafenco.



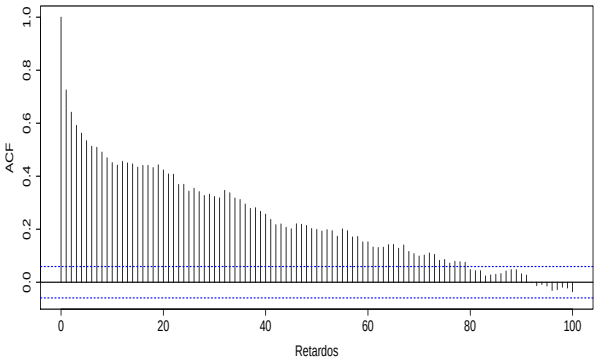
(a) Serie de tiempo componente Este



(b) ACF: componente Este



(c) Serie de tiempo componente Altura



(d) ACF: componente Altura

Figura 5.6. Serie de tiempo y ACF de la componente Este y Altura de Llafenco.

A continuación se presentan los resultados obtenidos del test ADF y KPSS sobre las series de tiempo de la Figura 5.1-5.6. Los resultados se presentan en la Tabla 5.2.

Tabla 5.2. Resultados obtenidos de los test ADF y KPSS sobre las componentes Norte, Este y Altura de Iquique, Concepción y Llafenco.

Componente	Test ADF p-valor	Resultado	Test KPSS p-valor	Resultado
Iquique				
Norte	0.38	Serie de tiempo no estacionaria	0.01	Serie de tiempo no estacionaria
Este	0.01	Serie de tiempo estacionaria	0.01	Serie de tiempo no estacionaria
Altura	0.08	Serie de tiempo no estacionaria	0.01	Serie de tiempo no estacionaria
Concepción				
Norte	0.61	Serie de tiempo no estacionaria	0.01	Serie de tiempo no estacionaria
Este	0.05	Serie de tiempo estacionaria	0.01	Serie de tiempo no estacionaria
Altura	0.05	Serie de tiempo estacionaria	0.01	Serie de tiempo no estacionaria
Llafenco				
Norte	0.61	Serie de tiempo no estacionaria	0.01	Serie de tiempo no estacionaria
Este	0.47	Serie de tiempo no estacionaria	0.01	Serie de tiempo no estacionaria
Altura	0.01	Serie de tiempo estacionaria	0.08	Serie de tiempo estacionaria

De la Tabla 5.2 se puede deducir que las series de tiempo de la componente Norte y Altura de Iquique no resultan ser estacionarias en ambos test. Sin embargo, para la componente Este se presenta una contradicción, ya que el test KPSS entrega como resultado que la serie de tiempo no es estacionaria, en cambio el test ADF resulta todo lo contrario. En este caso se toma en consideración el test KPSS, ya que al observar la Figura 5.2b se logra apreciar que la función de autocorrelación simple (ACF) decae lentamente hacia cero, indicando presencia tendencial y no de estacionariedad. En Concepción la serie de tiempo de la componente Norte no resulta ser estacionaria. Sin embargo, en las series de tiempo de la componente Este y Altura los test presentan una contradicción, pero se tomará en consideración el test KPSS, ya que al observar la Figura 5.4b y 5.4d se logran apreciar que las funciones de autocorrelación simple (ACF) decaen lentamente hacia cero, indicando presencia tendencial y no de estacionariedad. Finalmente en Llafenco se observa que las series de tiempo de la componente Norte y Este no resultan ser estacionarias y la componente Altura resulta estacionaria.

5.4. Estimación sobre las series de tiempo

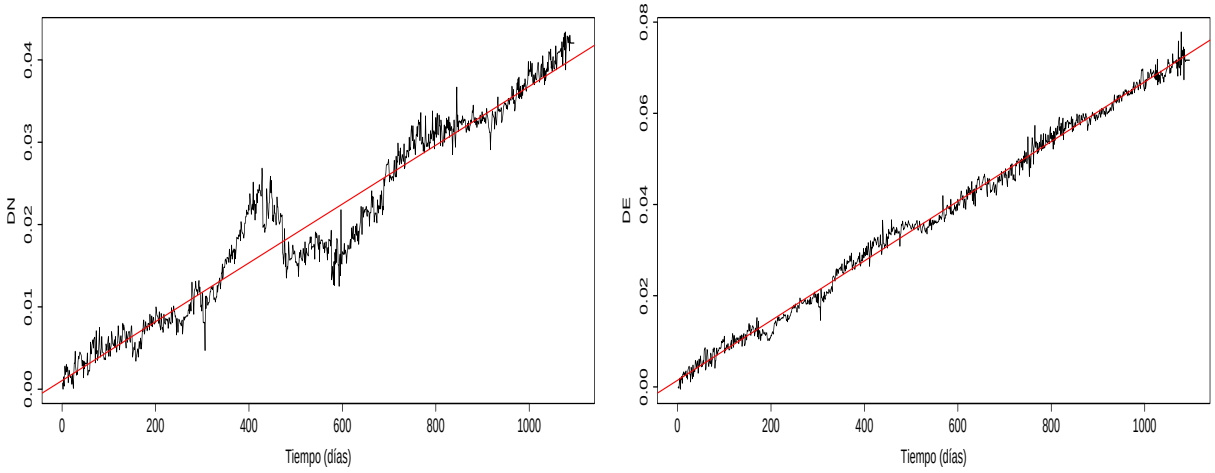
En esta sección se procede con la estimación de las series de tiempo de la componente Norte, Este y Altura de Iquique, Concepción y Llafenco.

5.4.1. Estimación sobre las series de tiempo en Iquique

En la Figura 5.1a y 5.2a se observa que cada serie de tiempo presenta una tendencia lineal muy marcada, por consiguiente se procede con la aplicación de la estimación de la tendencia a través de regresión lineal.

5.4.1.1. Estimación de la tendencia

En la Figura 5.7a y 5.7b se aprecia a simple vista un buen ajuste de la recta sobre las series de tiempo componente Norte y Este de Iquique.



(a) Componente Norte **(b)** Componente Este

Figura 5.7. Ajuste de la recta sobre las componente Norte y Este de Iquique.

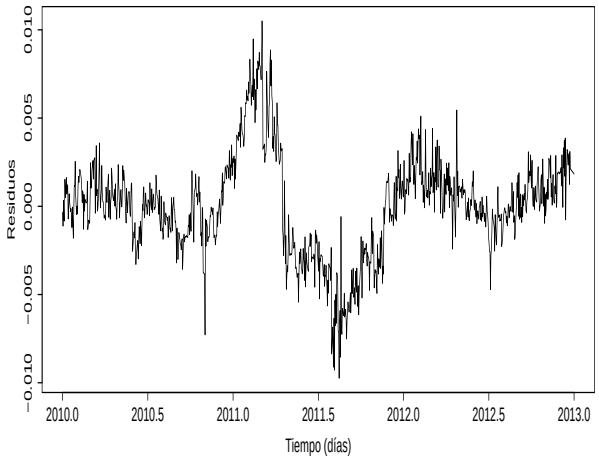
De la Tabla 5.3 se deduce que existe un muy buen ajuste de la recta para la componente Norte y Este de Iquique, ya que el coeficiente de determinación R^2 presenta un porcentaje sobre el 90 %, lo que indica que la recta se ajusta de muy buena manera sobre los datos. También se presenta en cada componente la desviación estándar, el intercepto (β_0) y la pendiente (β_1) que resulta positiva en ambos casos.

Tabla 5.3. Resultados obtenidos de la estimación de la tendencia, sobre la componente Norte y Este de Iquique.

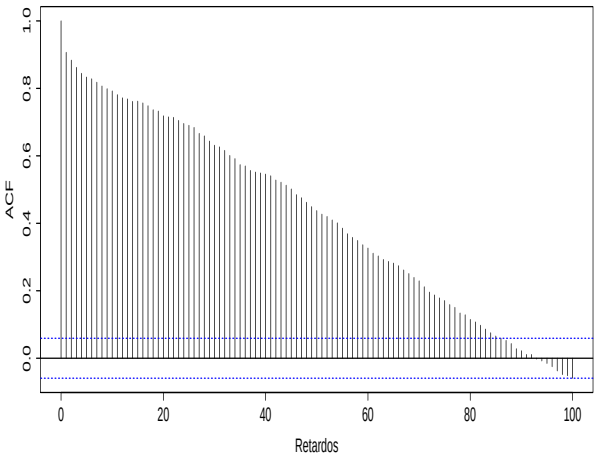
Componente	R^2	β_0	β_1	desv. estándar
Norte	94.08 %	0.0011	0.00004	0.0028
Este	99.40 %	0.0015	0.00007	0.0016

Residuos obtenidos de la regresión lineal

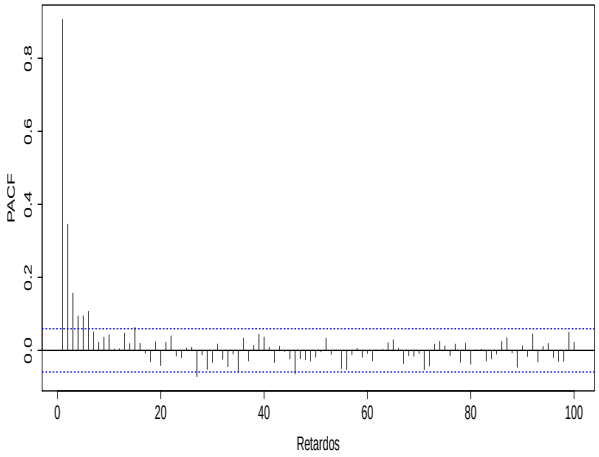
Para la estimación de modelos sobre la componente Norte y Este de Iquique, se utilizarán los residuos obtenidos en cada regresión lineal (analizadas anteriormente).



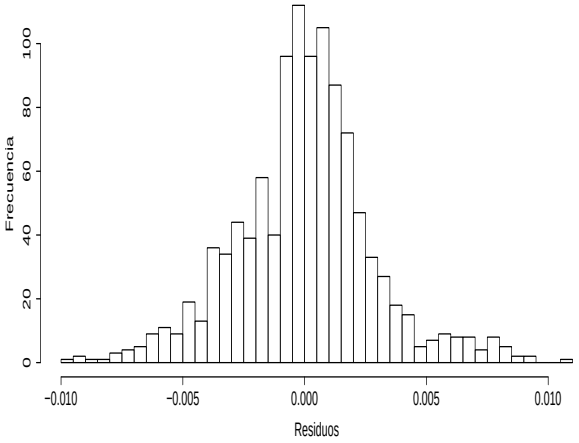
(a) Residuos componente Norte



(b) ACF

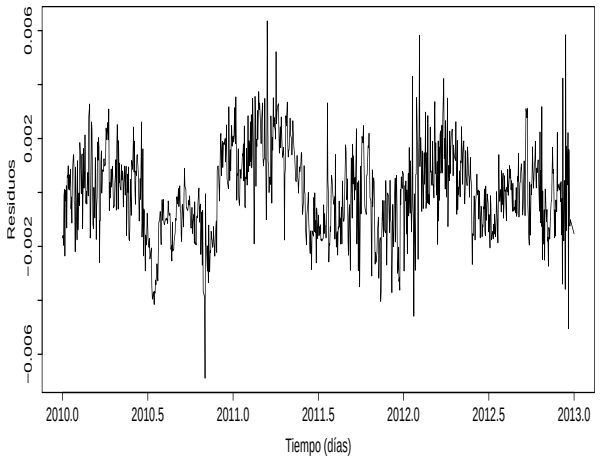


(c) PACF

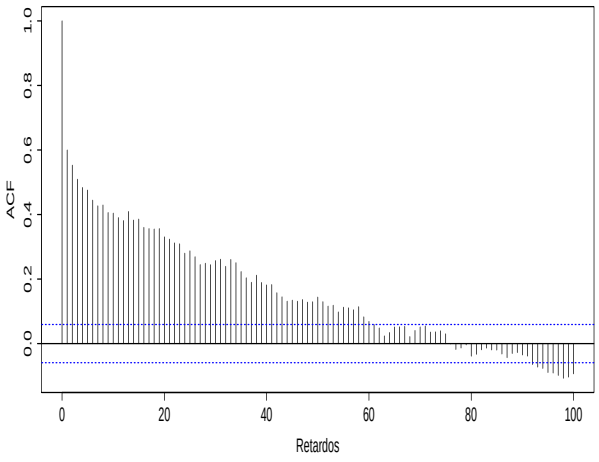


(d) Distribución de frecuencias

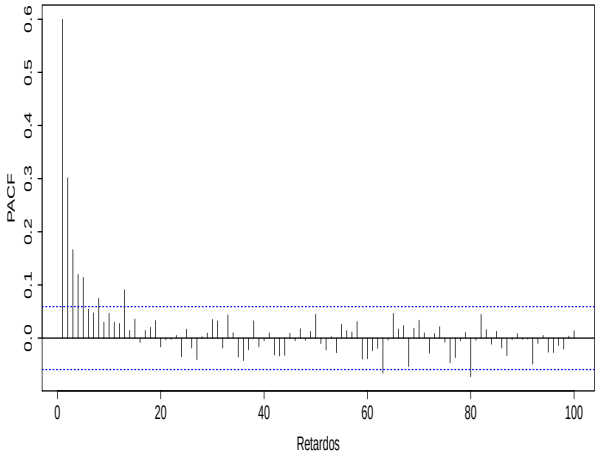
Figura 5.8. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Iquique.



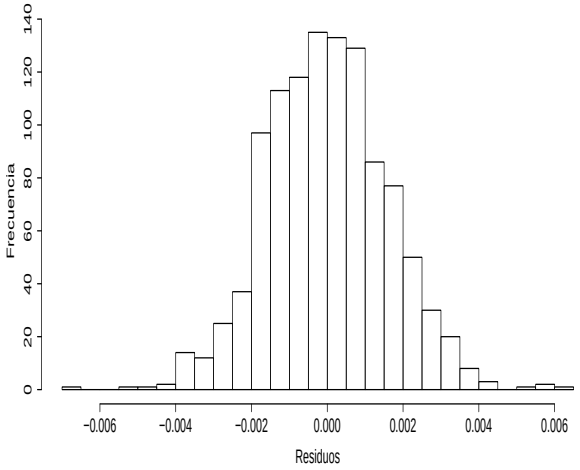
(a) Componente Este



(b) ACF



(c) PACF



(d) Distribución de frecuencias

Figura 5.9. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Iquique.

En la Figura 5.8b y 5.9b se observa que el ACF de los residuos no presentarían independencia, ya que las autocorrelaciones se encuentran fuera de los intervalos de confianza. En la Figura 5.8d y 5.9d se observa que los residuos no seguirían un comportamiento normal. Finalmente sobre los residuos de la Figura 5.8a y 5.9a se debe determinar la existencia de estacionariedad, por lo que se procede con la aplicación del test ADF y KPSS. Los resultados se presentan en la Tabla 5.4.

Tabla 5.4. Resultados de los test ADF y KPSS, sobre los residuos de la componente Norte y Este de Iquique.

Residuos	Test ADF p-valor	Resultado	Test KPSS p-valor	Resultado
Componente Norte	0.38	Serie de tiempo no estacionaria	0.02	Serie de tiempo no estacionaria
Componente Este	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria

De los resultados obtenidos en la Tabla 5.4 se puede deducir que los residuos de la componente Norte de Iquique no son estacionarios y los residuos de la componente Este de Iquique son estacionarios.

5.4.1.2. Estimación de la componente residual

A continuación se identifican los modelos a utilizar sobre los residuos obtenidos de la regresión lineal de la componente Norte, Este y la serie de tiempo de la componente Altura de Iquique.

Identificación de modelos

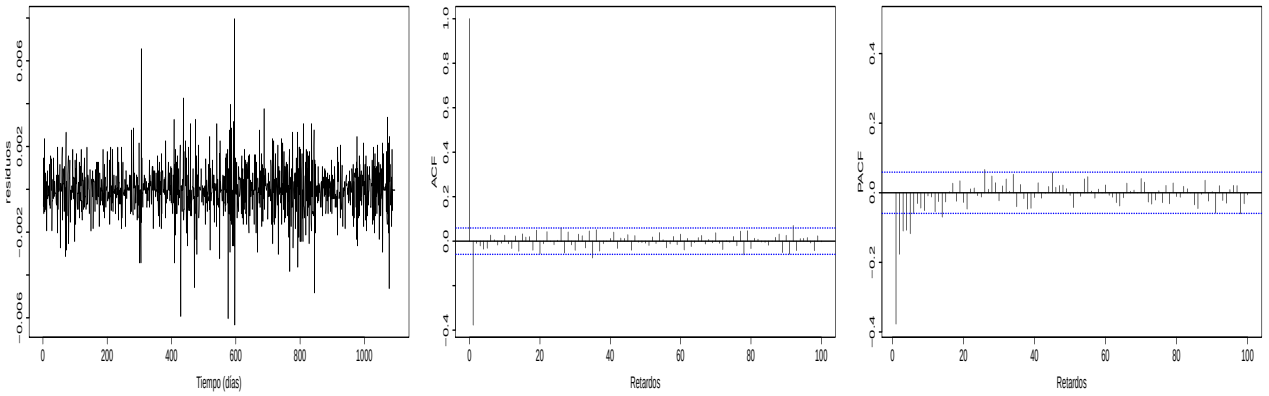
Como se explicó previamente, los residuos de la componente Norte y la serie de tiempo de la componente Altura de Iquique no son estacionarios, por lo que se aplica diferenciación para cumplir con la condición de estacionariedad. Sin embargo, los residuos de la componente Este de Iquique resultan estacionarios, por lo que no necesitan diferenciación. A continuación se presentan en la Tabla 5.5 los resultados obtenidos del test ADF y KPSS, respecto a la aplicación de la primera diferenciación.

Tabla 5.5. Resultados de los test ADF y KPSS, aplicados sobre la primera diferenciación de los residuos de la componente Norte y serie de tiempo de la componente Altura de Iquique.

Serie de tiempo	Test ADF p-valor	Resultado	Test KPSS p-valor	Resultado
Residuos componente Norte	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria
Componente Altura	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria

En la Tabla 5.5 ambos test afirman que la primera diferenciación logra que los residuos de la componente Norte y la serie de tiempo de la componente Altura de Iquique se vuelvan estacionarios, por consiguiente en ambos casos es aceptable para que se ajusten modelos ARIMA(p,d,q), con d=1.

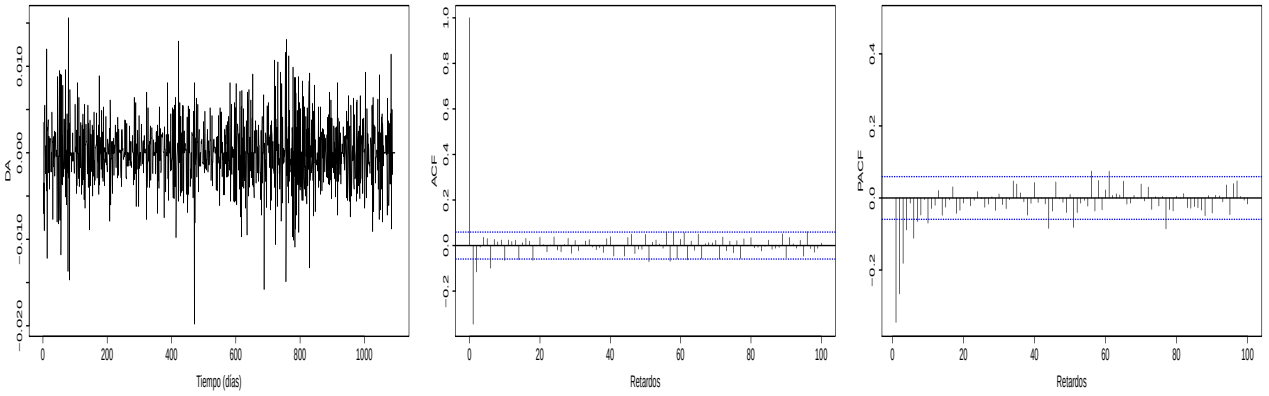
A continuación se presenta gráficamente la primera diferenciación sobre los residuos de la componente Norte y Altura de Iquique con sus respectivos correlogramas.



(a) Primera diferenciación: residuos componente Norte

(b) ACF

(c) PACF



(d) Primera diferenciación: componente Altura

(e) ACF

(f) PACF

Figura 5.10. Gráfico de la primera diferenciación, con sus respectivos correlogramas ACF y PACF sobre los residuos de la componente Norte y serie de tiempo componente Altura de Iquique.

En la Figura 5.10a y 5.10d se observa que los gráficos de la primera diferenciación presentan una media constante, ya que oscilan alrededor del cero y una variabilidad con respecto a esa media también constante. En los correlogramas de la Figura 5.10b, 5.10c, 5.10e y 5.10f no se logra apreciar una clara identificación del orden del modelo, por lo que se ajustan distintos modelos ARIMA(p,d,q).

Con respecto a los residuos de la componente Este de Iquique presentados en la Figura 5.9a, se logra aprecia una media constante, ya que los residuos oscilan alrededor del cero y la variabilidad con respecto a esa media también resulta constante. En los correlogramas de la Figura 5.9b y 5.9c no se logra apreciar una clara identificación del orden del modelo, por lo que se ajustan distintos modelos ARMA(p,q).

En la Tabla 5.6 se presentan los modelos ARMA(p,q) y ARIMA(p,d,q) seleccionados para cada componente (ver (A.1), (A.2) y (A.3)). En los residuos de la componente Norte de Iquique, el menor valor AIC y BIC lo obtuvo el modelo ARIMA(2,1,1). Con respecto a los residuos de la componente Este de Iquique, el menor valor AIC y BIC lo obtuvo el modelo ARMA(2,2), por lo que serían los mejores modelos que se ajustan sobre los residuos. Finalmente en la serie de tiempo componente Altura de Iquique, el menor valor AIC y BIC lo obtuvo el modelo ARIMA(1,1,1), por lo que sería el mejor modelo que se ajusta sobre los datos.

Tabla 5.6. Modelos seleccionados para cada componente, con sus respectivos criterios de información AIC y BIC.

Serie de tiempo	Modelo	AIC	BIC
Residuos componente Norte	ARIMA(2,1,1)	-11821.48	-11801.49
Residuos componente Este	ARMA(2,2)	-11656.25	-11626.25
componente Altura	ARIMA(1,1,1)	-9315.193	-9300.198

En la Tabla 5.7 se presenta el contraste de significación de cada modelo seleccionado. Se puede apreciar que en el modelo ARIMA(2,1,1) los tres parámetros resultan significativos para el modelo. En el modelo ARMA(2,2) los cuatro parámetros resultan significativos para el modelo y finalmente en el modelo ARIMA(1,1,1) los dos parámetros resultan significativos para el modelo. Esto por tener un p-valor menor que un nivel de significación de $\alpha = 0.05$.

Tabla 5.7. Contraste de significación de los parámetros del modelo ARIMA(2,1,1), ARMA(2,2) y ARIMA(1,1,1).

Parámetro	Estimación	Error estándar	z value	$Pr(>\|z\|)$
Residuos componente Norte: ARIMA(2,1,1)				
ar1	0.3239	0.0483	6.70470	< 0.00001
ar2	0.1206	0.0390	3.09290	0.001982
ma1	-0.8263	0.0360	-22.9748	< 0.00001
Residuos componente Este: ARMA(2,2)				
ar1	1.5826	0.1552	10.1950	< 0.00001
ar2	-0.5916	0.1507	-3.92670	0.000086
ma1	-1.2698	0.1641	-7.73610	< 0.00001
ma2	0.3518	0.1267	2.77570	0.005508
Serie componente Altura: ARIMA(1,1,1)				
ar1	0.2466	0.0489	5.04880	< 0.00001
ma1	-0.7688	0.0339	-22.6823	< 0.00001

En la Figura 5.11, 5.12 y 5.13 se presentan los ajustes del modelo ARIMA(2,1,1), ARMA(2,2) y ARIMA(1,1,1).

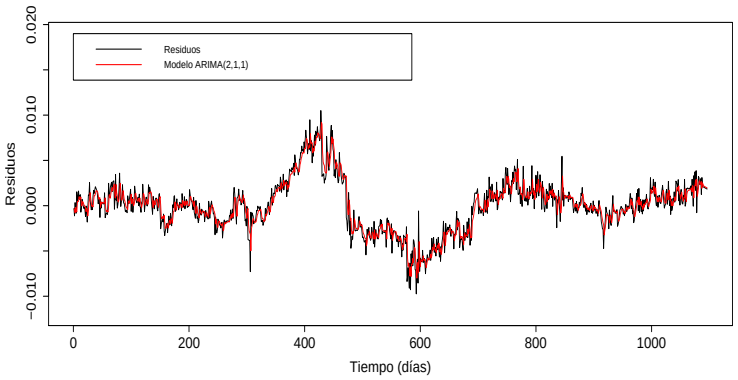


Figura 5.11. Ajuste del modelo ARIMA(2,1,1) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Norte de Iquique.

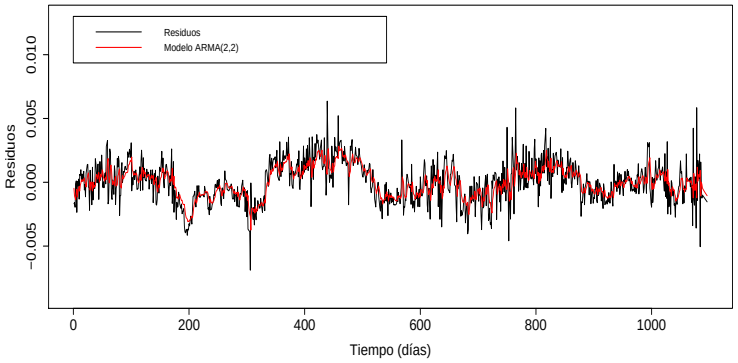


Figura 5.12. Ajuste del modelo ARMA(2,2) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Este de Iquique.

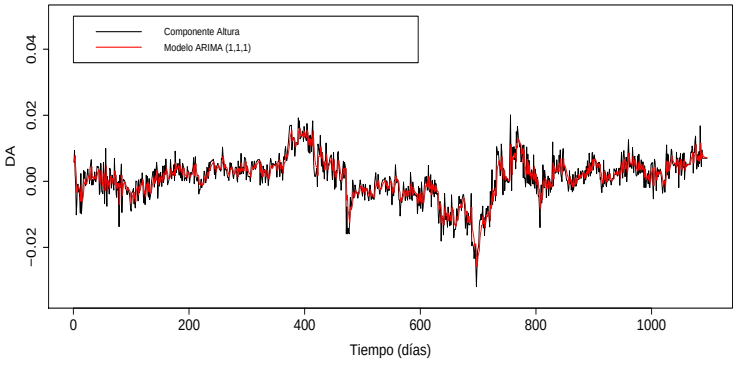
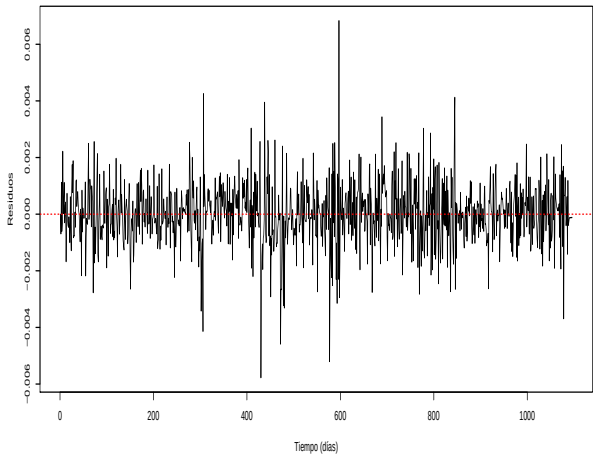


Figura 5.13. Ajuste del modelo ARIMA(1,1,1) sobre la serie de tiempo componente Altura de Iquique.

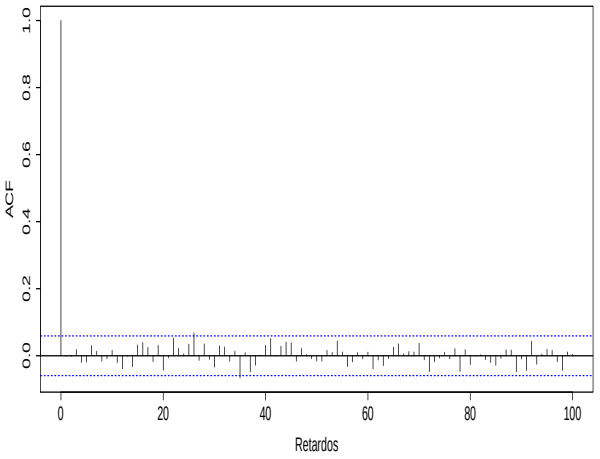
Se observa a simple vista que los tres modelos se ajustan de buena manera sobre cada serie de tiempo. Sin embargo, para determinar si los modelo realmente son adecuados para la representación de los datos, se deben analizar los residuos.

5.4.1.3. Diagnóstico del modelo

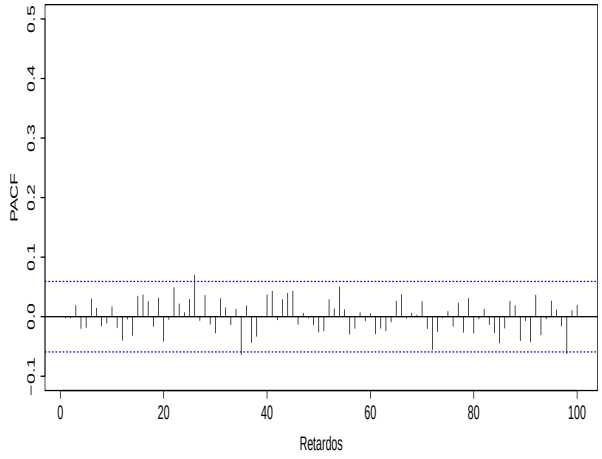
A continuación se presenta gráficamente el diagnóstico de cada modelo seleccionado sobre la serie de tiempo componente Altura y los residuos de la componente Norte y Este de Iquique.



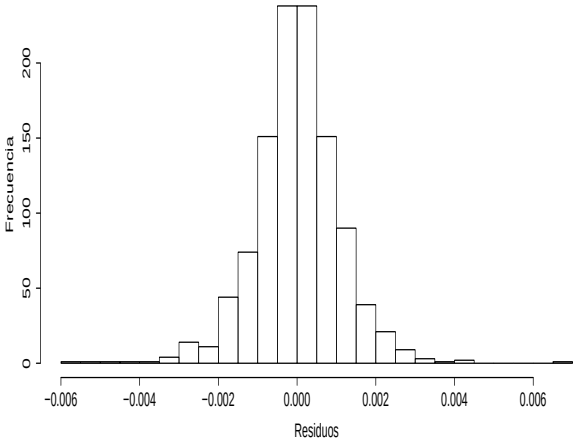
(a) Residuos



(b) ACF

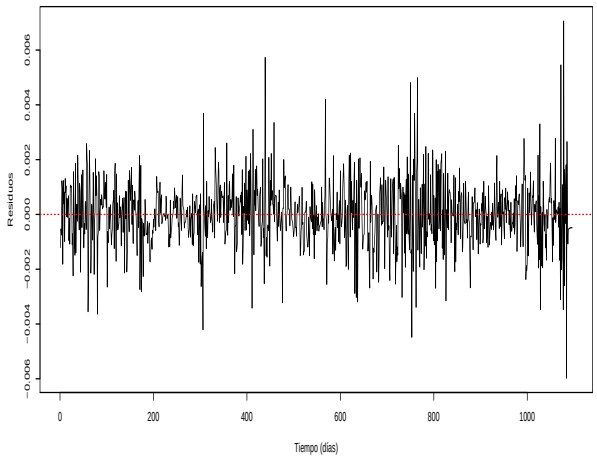


(c) PACF

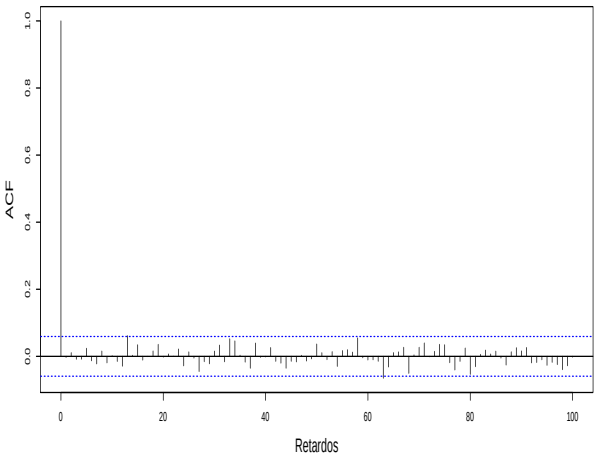


(d) Distribución de frecuencias

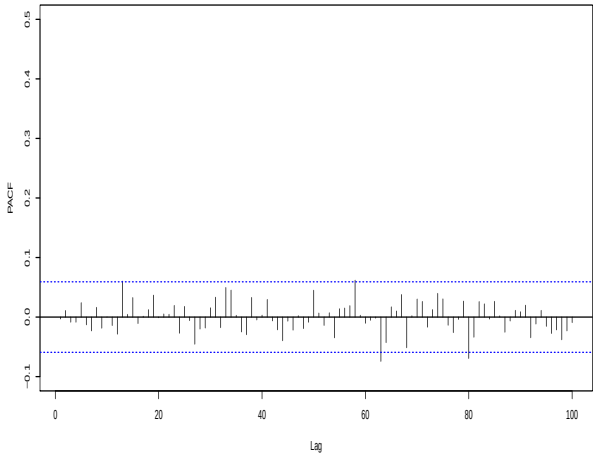
Figura 5.14. Gráfico de dispersión de los residuos del modelo ARIMA(2,1,1) con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Iquique.



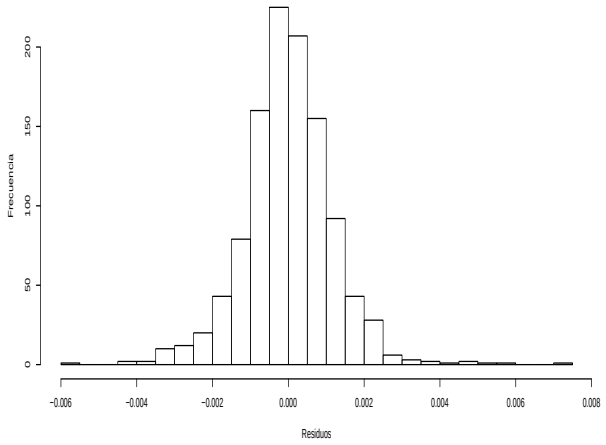
(a) Residuos



(b) ACF

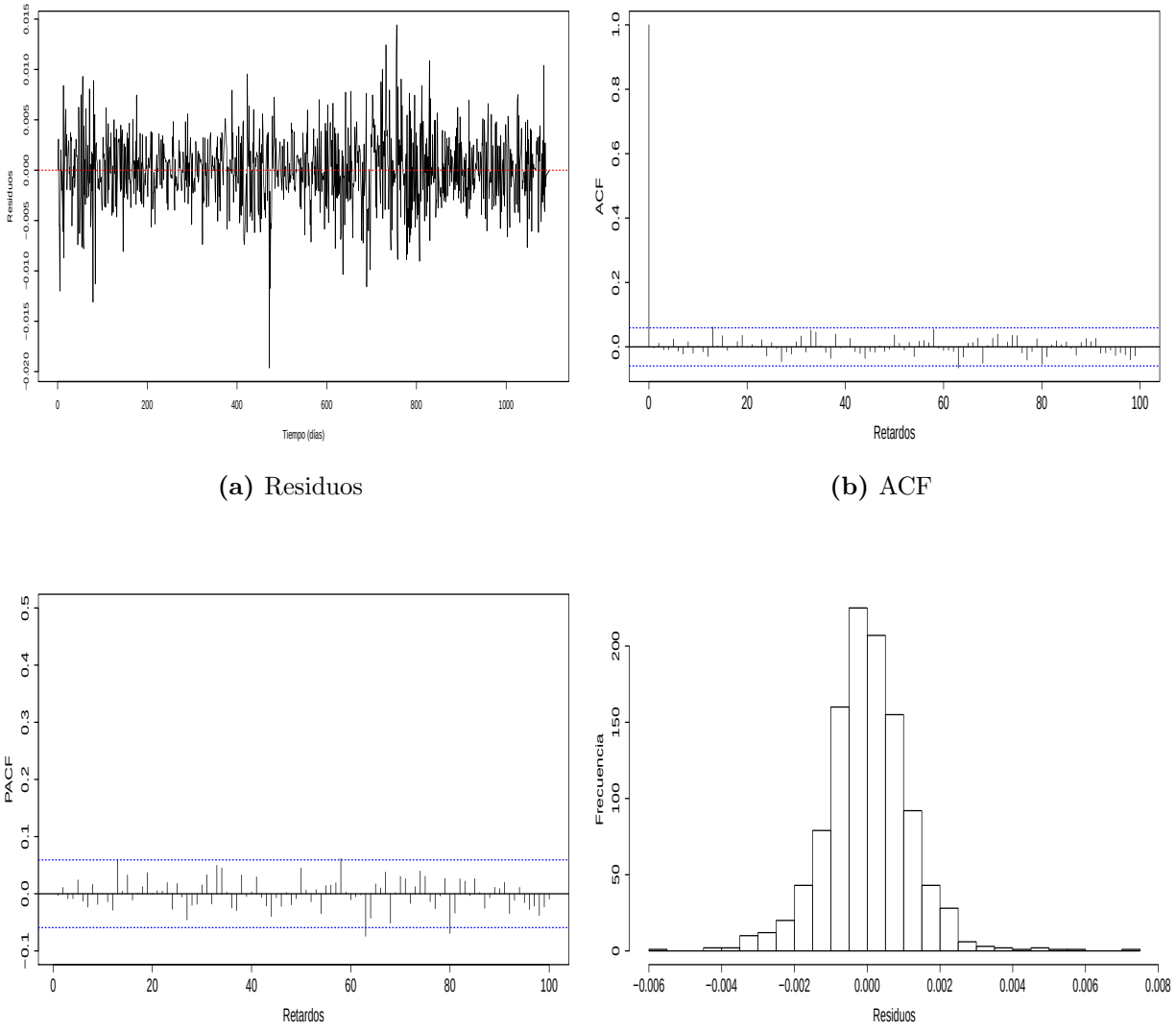


(c) PACF



(d) Distribución de frecuencias

Figura 5.15. Gráfico de dispersión de los residuos del modelo ARMA(2,2) con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Iquique.



(a) Residuos

(b) ACF

(c) PACF

(d) Distribución de frecuencias

Figura 5.16. Gráfico de dispersión de los residuos del modelo ARIMA(1,1,1) con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Iquique.

En la Figura 5.14a, 5.15a y 5.16a se observa que los residuos de los modelos ARIMA(2,1,1), ARMA(2,2) y ARIMA(1,1,1) presentan una media constante y una varianza con respecto a la media también constante. En la Figura 5.14b, 5.15b y 5.16b se aprecia el correlograma ACF de los residuos para los tres modelos, donde se puede notar la presencia de independencia entre las observaciones, ya que las autocorrelaciones se mantienen dentro de los intervalos de confianza. Finalmente en la Figura 5.14d, 5.15d y 5.16d se observa a simple vista que los residuos de los tres modelos se aproximan a una distribución normal.

A continuación se presentan en la Tabla 5.8 los resultados obtenidos del test Ljung-Box para comprobar independencia entre los residuos de cada modelo.

Tabla 5.8. Test Ljung-Box sobre los residuos de cada modelo.

Componente	Residuos	Test Ljung-Box p-valor	Resultado
Norte	Modelo ARIMA(2,1,1)	0.8582	Independientes
Este	Modelo ARMA(2,2)	0.9487	Independientes
Altura	Modelo ARIMA(1,1,1)	0.1565	Independientes

De la tabla 5.8 de deduce que los residuos de cada modelo presentan independencia entre sí.

Finalmente los modelos para los residuos de cada componente, quedan expresados de la siguiente manera:

Componente Norte: modelo ARIMA(2,1,1)

$$\varepsilon_t = 0.323901\varepsilon_{t-1} + 0.120635\varepsilon_{t-2} + \omega_t - (-0.826313)\omega_{t-1}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

Componente Este: modelo ARMA(2,2)

$$\varepsilon_t = 1.58260\varepsilon_{t-1} + -0.59155\varepsilon_{t-2} + \omega_t - (-1.26980)\omega_{t-1} - (0.35177)\omega_{t-2}, \omega_t \sim RB(0, \sigma^2)$$

Componente Altura: modelo ARIMA(1,1,1)

$$\varepsilon_t = 0.246643\varepsilon_{t-1} + \omega_t - (-0.768753)\omega_{t-1}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

5.4.1.4. Pronóstico

En esta sección se proceden a realizar dos tipos de pronósticos. El primer pronóstico es con respecto a la parte determinística (regresión lineal) y el segundo con respecto a la parte estocástica (modelos ARMA y ARIMA). Es importante recordar que los datos en estudio corresponder a enero del 2010 hasta diciembre del 2012, por ende el pronóstico se realiza sobre los primeros cuatro días de enero del 2013.

Pronóstico determinístico

Se procede con el pronóstico obtenido de la estimación sobre la tendencia. En la Figura 5.17 y 5.18 se representan gráficamente con color plomo los intervalos de

confianza, es decir, los límites de predicción y con color azul las predicciones.

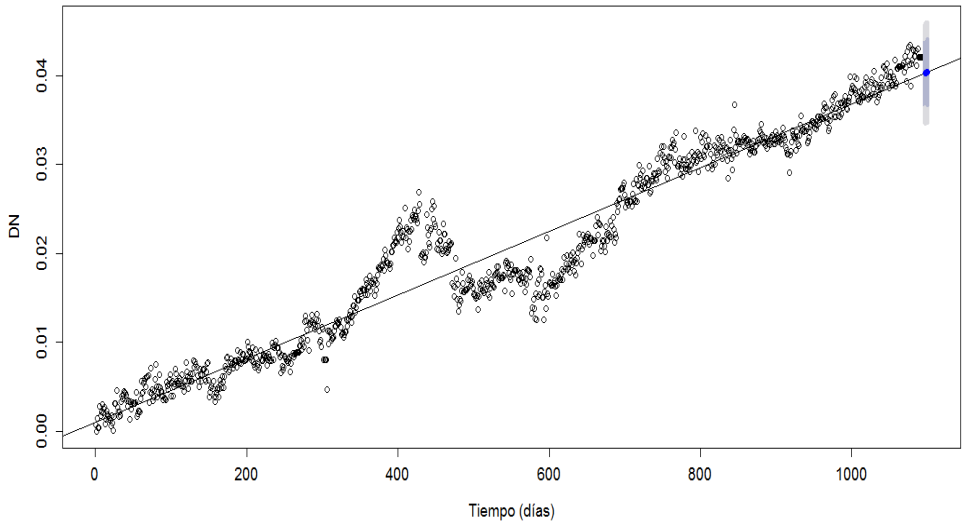


Figura 5.17. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Norte de Iquique.

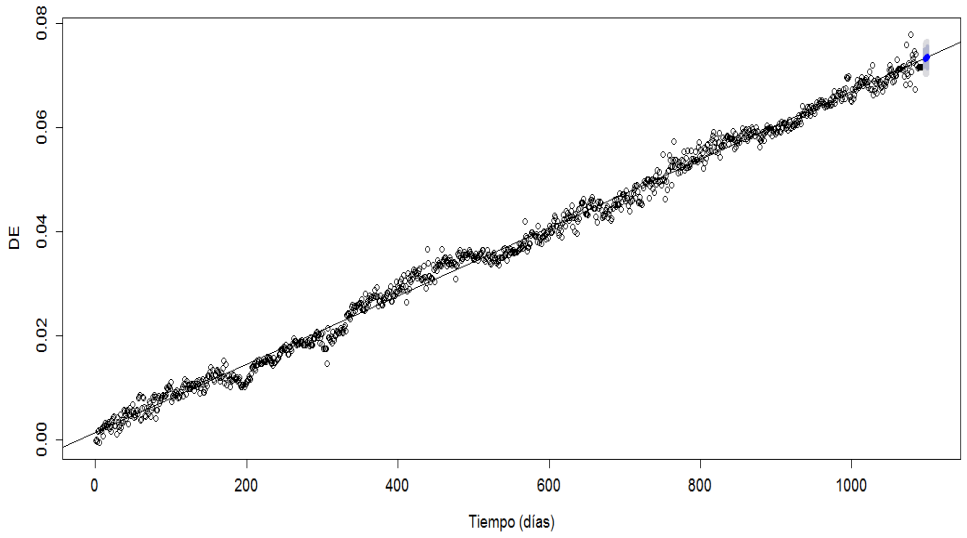


Figura 5.18. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Este de Iquique.

A continuación se presentan en la Tabla 5.9 los resultados obtenidos de las predicciones.

Tabla 5.9. Predicciones obtenidas sobre los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Norte y Este de Iquique.

Día	Mes	Año	Pronóstico componente Norte	Pronóstico componente Este
1	Enero	2013	0.04025	0.07325
2	Enero	2013	0.04029	0.07331
3	Enero	2013	0.04032	0.07338
4	Enero	2013	0.04036	0.07344

Pronóstico estocástico

Se procede con el pronóstico sobre los residuos obtenidos de la regresión lineal, con respecto a la serie de tiempo componente Norte y Este de Iquique. También se realiza el pronóstico sobre la serie de tiempo componente Altura de Iquique que no fue estimada (por no presentar tendencia).

Las predicciones obtenidas con respecto a la serie de tiempo componente altura y los residuos obtenidos en la regresión lineal sobre la componente Norte y Este de Iquique, se presenta gráficamente en la Figura 5.19, 5.20 y 5.21. Se pueden observar con color azul las predicciones para los primeros cuatro días de enero del 2013 y con color plomo los intervalos de confianza, es decir, los limites de las predicciones.

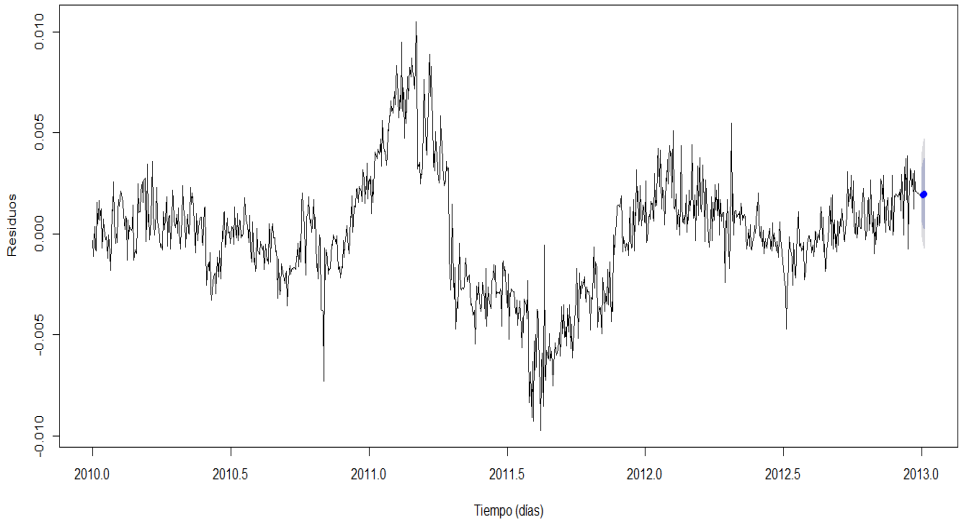


Figura 5.19. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(2,1,1) sobre los residuos de la componente Norte de Iquique.

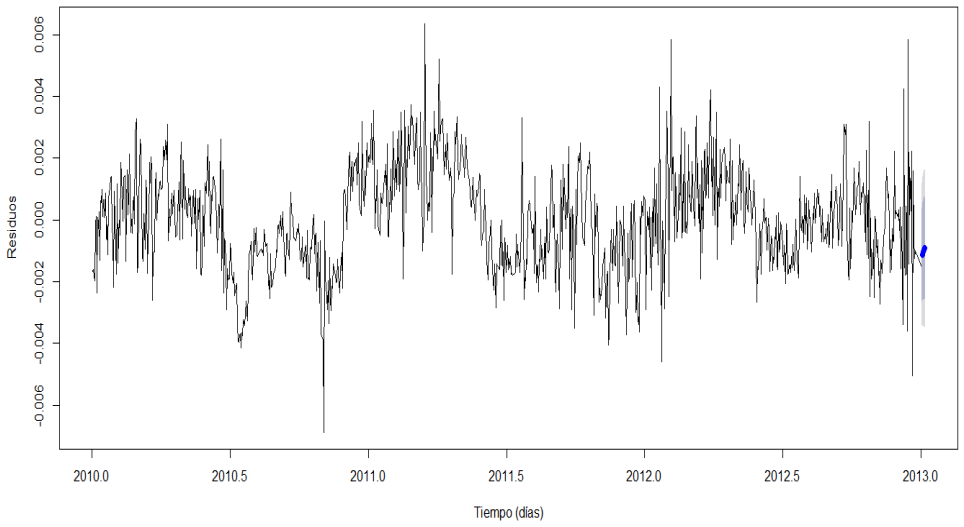


Figura 5.20. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(2,2) sobre los residuos de la componente Este de Iquique.

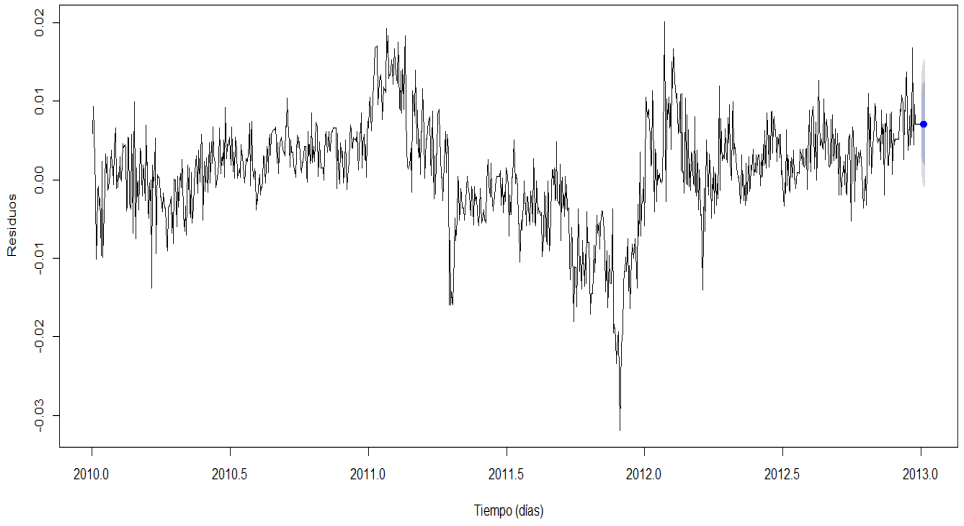


Figura 5.21. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(1,1,1) sobre la componente Altura de Iquique.

A continuación se presentan en la Tabla 5.10 los resultados obtenidos de las predicciones sobre la componente Altura y los residuos de la componente Norte y Este de Iquique.

Tabla 5.10. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto a la componente Altura y los residuos de la componente Norte y Este de Iquique.

Día	Mes	Año	Pronóstico residuos DN Modelo ARIMA(2,1,1)	Pronóstico residuos DE Modelo ARMA(2,2)	Pronóstico DA Modelo ARIMA(1,1,1)
1	Enero	2013	0.00192	-0.00112	0.00708
2	Enero	2013	0.00194	-0.00103	0.00710
3	Enero	2013	0.00196	-0.00097	0.00710
4	Enero	2013	0.00197	-0.00092	0.00711

Pronóstico general: parte determinística más parte estocástica

En esta sección se obtiene el pronóstico general de cada componente, es decir, la suma de la parte determinística más la parte estocástica.

En la Figura 5.22 y 5.23 se representan gráficamente las predicciones generales para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte y Este de Iquique. Se pueden observar con color rojo las predicciones para los primeros cuatro días de enero del 2013.

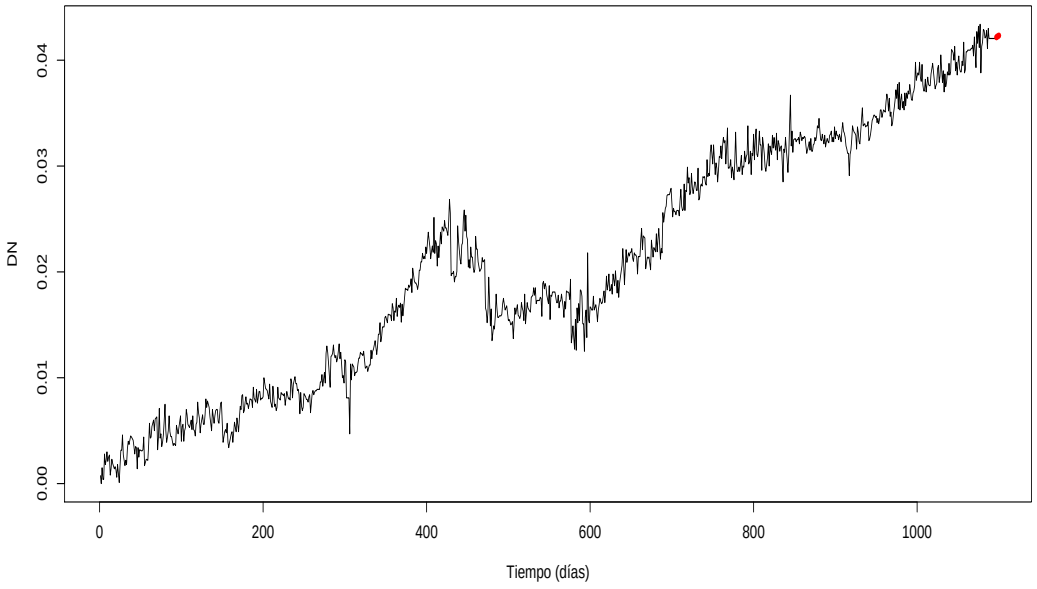


Figura 5.22. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Norte de Iquique.

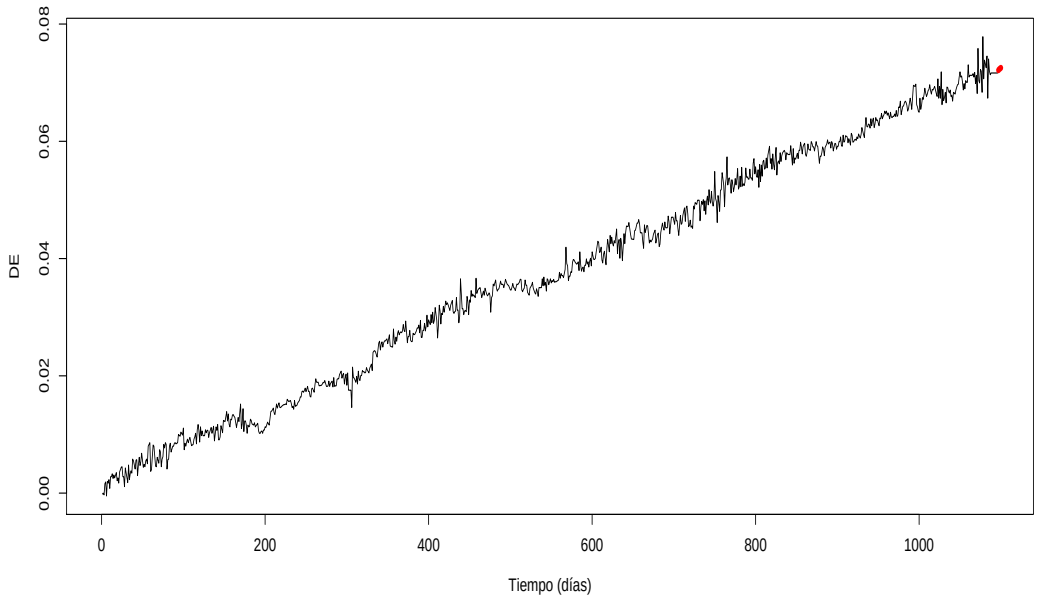


Figura 5.23. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Este de Iquique.

A continuación se presentan en la Tabla 5.11 las predicciones generales, para los primeros cuatro días de enero del 2013.

Tabla 5.11. Predicciones generales obtenidas para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte y Este de Iquique.

Día	Mes	Año	Predicción parte determinística	Predicción parte estocástica	Predicción general
Componente Norte					
1	Enero	2013	0.04025	0.00192	0.04217
2	Enero	2013	0.04029	0.00194	0.04223
3	Enero	2013	0.04032	0.00196	0.04228
4	Enero	2013	0.04036	0.00197	0.04232
Componente Este					
1	Enero	2013	0.07325	-0.00112	0.07213
2	Enero	2013	0.07331	-0.00103	0.07228
3	Enero	2013	0.07338	-0.00097	0.07241
4	Enero	2013	0.07344	-0.00092	0.07252

Predicciones finales y módulo del vector desplazamiento

Obtenidas las predicciones finales sobre las series de tiempo de la componente Norte, Este y Altura de Iquique es necesario calcular el módulo del vector desplazamiento.

Recordemos la ecuación 1.2.3:

$$A = \sqrt{A_x^2 + A_y^2 + A_z^2}$$

donde $A_x = DN$, $A_y = DE$ y $A_z = DA$.

Con las predicciones finales obtenidas para la componente Norte y Este de Iquique en la Tabla 5.11 y componente Altura de Iquique en la Tabla 5.10, se realiza el cálculo del módulo del vector desplazamiento. En la Tabla 5.12 se obtienen los resultados.

Tabla 5.12. Predicciones finales y módulo del vector desplazamiento obtenidos para los primeros cuatro días de enero del 2013, con respecto a la componente Norte, Este y Altura de Iquique.

Día	Mes	Año	Predicción final DN	Predicción final DE	Predicción final DA	Módulo
1	Enero	2013	0.04217	0.07213	0.00708	0.08385
2	Enero	2013	0.04223	0.07228	0.00709	0.08401
3	Enero	2013	0.04228	0.07241	0.00710	0.08415
4	Enero	2013	0.04232	0.07252	0.00711	0.08427

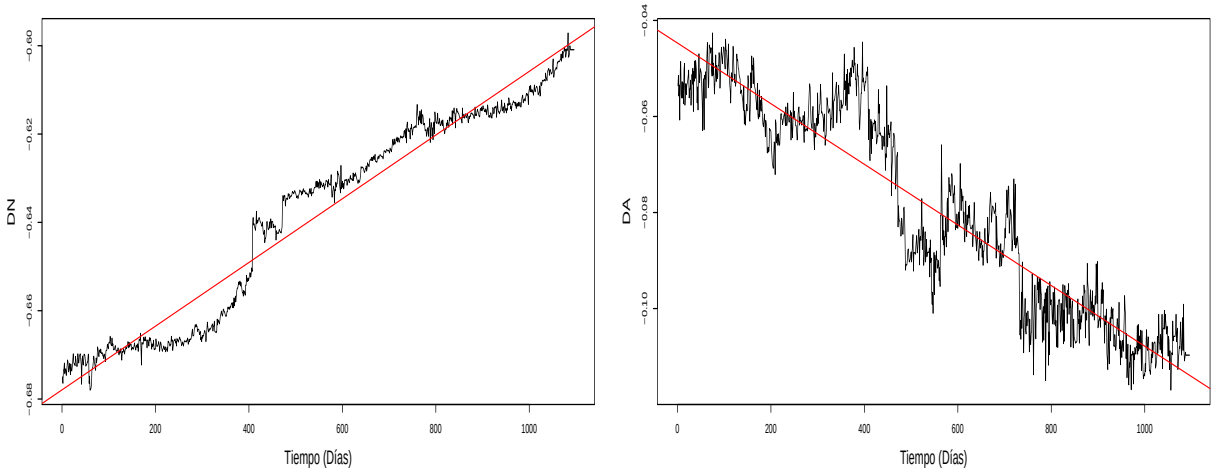
De la Tabla 5.12 se puede deducir que el desplazamiento resultante es un vector de longitud 0.08385 cm para el primer día, 0.08401 cm para el segundo día, 0.08415 cm para el tercer día y 0.08427 cm para el cuarto día.

5.4.2. Estimación sobre las series de tiempo en Concepción

En la Figura 5.3a, 5.4a y 5.4c se observa que las series de tiempo presentan una tendencia muy marcada, por consiguiente se procede con la aplicación de la estimación de la tendencia a través de regresión lineal y suavizamiento spline (dependiendo de la ocasión).

5.4.2.1. Estimación de la tendencia

En la Figura 5.24a y 5.24b se aprecia a simple vista un buen ajuste de la recta sobre las series de tiempo componente Norte y Altura de Concepción.



(a) Componente Norte **(b)** Componente Altura

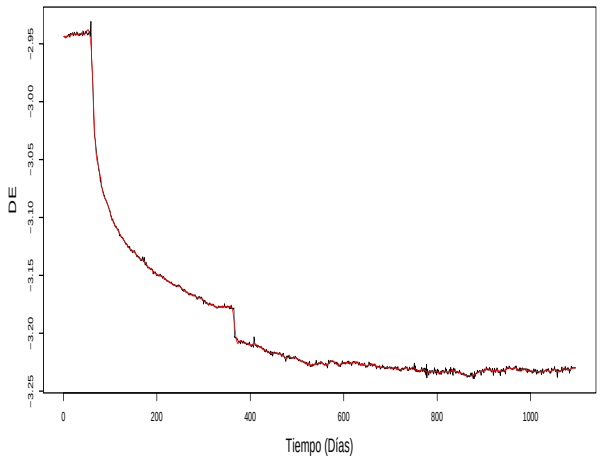
Figura 5.24. Ajuste de la recta sobre las componente Norte y Altura de Concepción.

En la tabla 5.13 se determina según los resultados obtenidos que existe un muy buen ajuste de la recta para la componente Norte y Altura de Concepción, ya que el coeficiente de determinación R^2 presenta un porcentaje sobre el 87 %, lo que indica que la recta se ajusta de muy buena manera sobre los datos. En el caso de la pendiente (β_1), la componente Norte de Concepción resulta positiva y la componente Altura de Concepción negativa.

Tabla 5.13. Resultados obtenidos de la estimación de la tendencia, sobre la componentes Norte y Altura de Concepción.

Componente	R^2	β_0	β_1	desv. estándar
Norte	95.35 %	-0.6779	0.00007	0.0051
Altura	87.68 %	-0.0447	-0.00006	0.0075

En la Figura 5.25 se observa el suavizamiento spline sobre la serie de tiempo componente Este de Concepción. En la tabla 5.14 se presenta el parámetro de suavizamiento, con el correspondiente valor de λ.



(a) Componente Este

Figura 5.25. Suavizamiento spline sobre las componente Este.

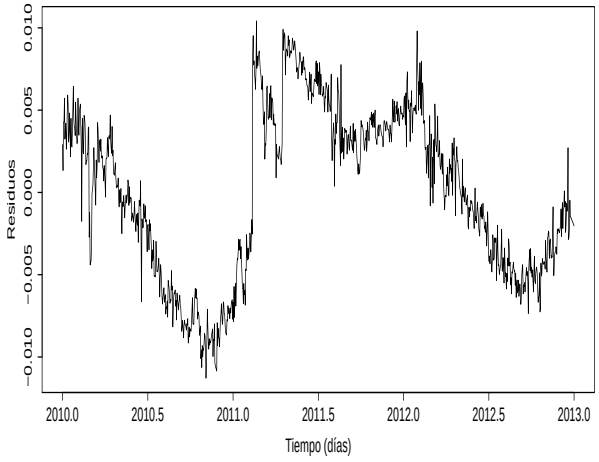
En la Tabla 5.14 se presentan los resultados obtenidos del suavizamiento spline aplicado sobre la componente Este de Concepción

Tabla 5.14. Resultados obtenidos de la estimación de la tendencia a través del suavizamiento spline sobre la componente Este de Concepción.

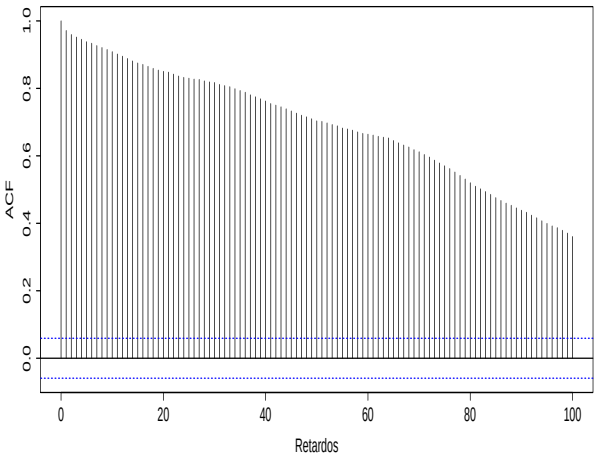
Componente	Parámetro de suavizamiento	λ
Este	0.0678	0.0000000053

Residuos obtenidos de la regresión lineal y suavizamiento spline

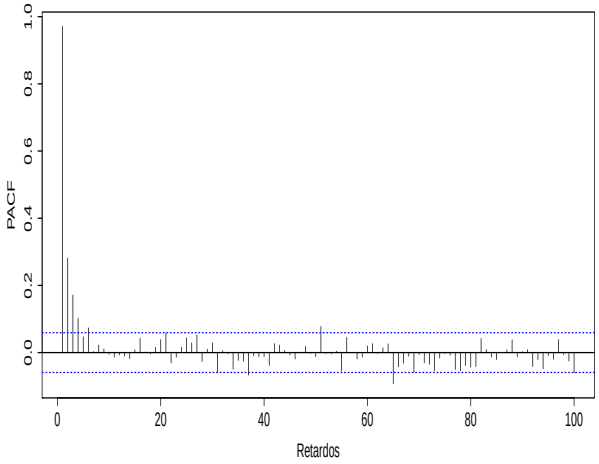
Para la estimación de modelos sobre la componente Norte, Este y Altura de Concepción, se utilizarán los residuos obtenidos en cada regresión lineal y suavizamiento spline (analizadas anteriormente).



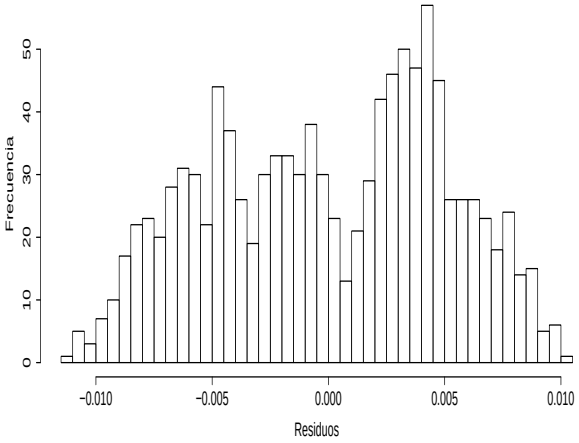
(a) Componente Norte



(b) ACF

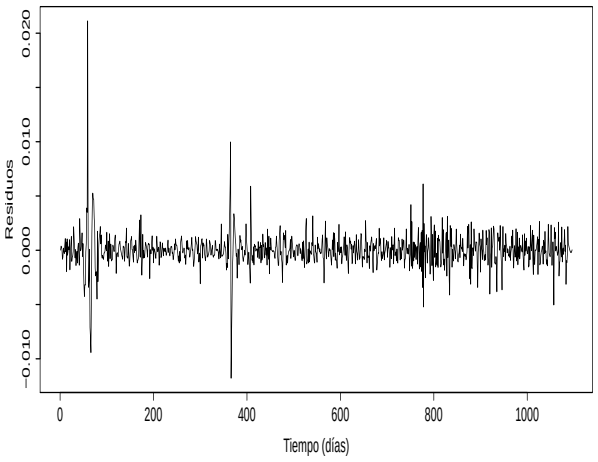


(c) PACF

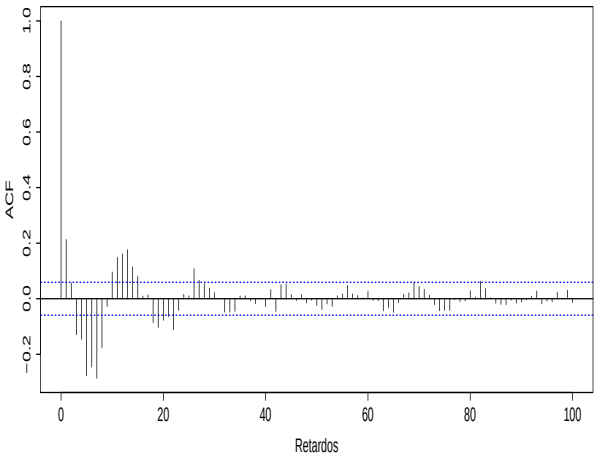


(d) Distribución de frecuencias

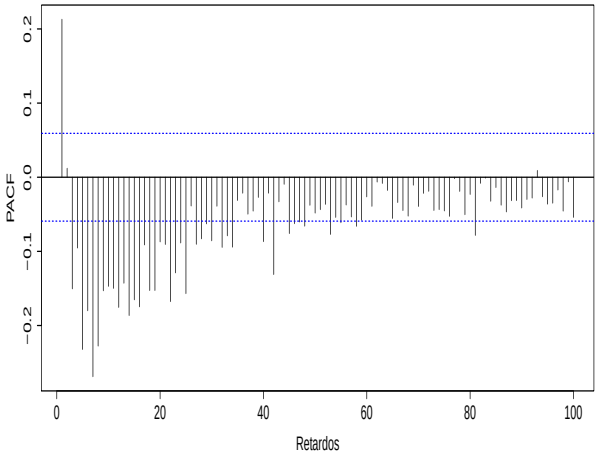
Figura 5.26. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Concepción.



(a) Componente Este

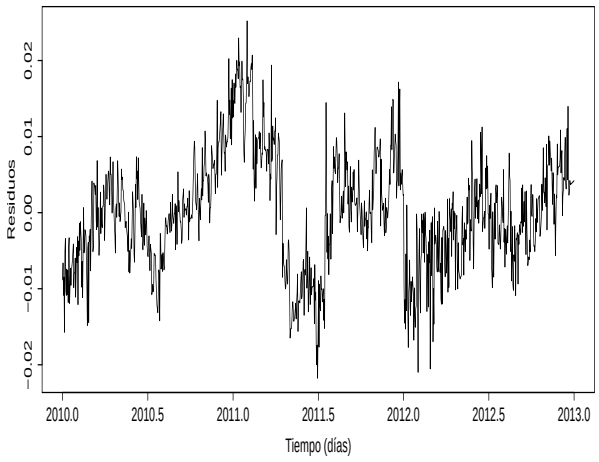


(b) ACF

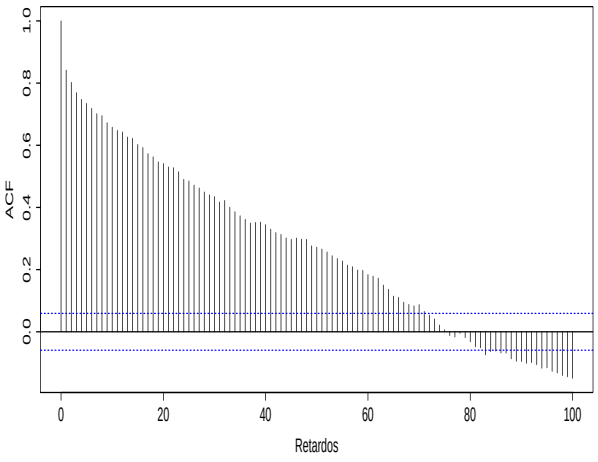


(c) PACF

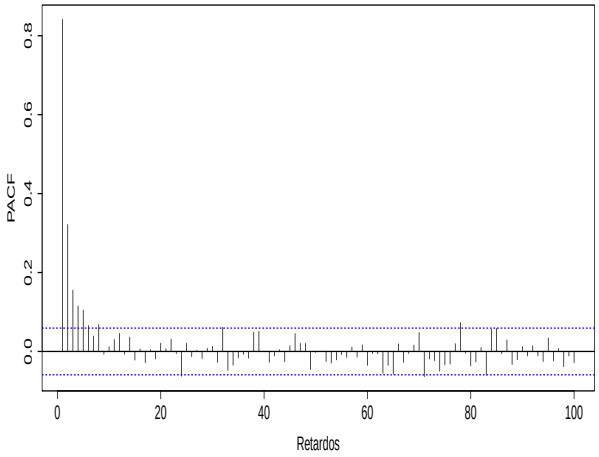
Figura 5.27. Gráfico de los residuos obtenidos en el suavizamiento spline, con sus respectivos correlogramas ACF y PACF, para la componente Este de Concepción.



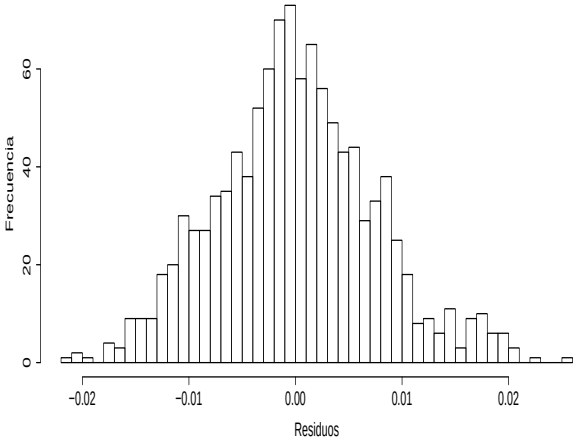
(a) Componente Altura



(b) ACF



(c) PACF



(d) Distribución de frecuencias

Figura 5.28. Gráfico de los residuos obtenidos en la regresión lineal, con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Concepción.

En la Figura 5.26b, 5.27b y 5.28b se observa que el ACF de los residuos no presentarían independencia, ya que las autocorrelaciones se encuentran fuera de los intervalos de confianza. Al observar la Figura 5.26d se puede determinar que los residuos no siguen un comportamiento normal. Sin embargo de la Figura 5.28d se podría pensar que los residuos se aproximarían a un comportamiento normal. Finalmente sobre los residuos de la Figura 5.26a, 5.27a y 5.28a, se debe determinar la existencia de estacionariedad, por lo que se procede con la aplicación del test ADF y KPSS. Los resultados se presentan en la Tabla 5.15.

Tabla 5.15. Resultados de los test ADF y KPSS sobre los residuos de la componentes Norte, Este y Altura de Concepción.

Residuos	Test ADF p-valor	Resultado	Test KPS p-valor	Resultado
Componente Norte	0.61	Serie de tiempo no estacionaria	0.01	Serie de tiempo no estacionaria
Componente Este	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria
Componente Altura	0.05	Serie de tiempo estacionaria	0.04	Serie de tiempo no estacionaria

De los resultados obtenidos en la Tabla 5.15 se puede deducir que los residuos de la componente Norte de Concepción no son estacionarios. En los residuos de la componente Altura de Concepción se presenta una contradicción pero se considerando el test KPSS, ya que al observar la Figura 5.28b se logra apreciar que la función de autocorrelación simple (ACF) decae lentamente hacia cero, indicando presencia tendencial y no de estacionariedad. Finalmente los residuos de la componente Este de Concepción son estacionarios.

5.4.2.2. Estimación de la componente residual

A continuación se identifican los modelos a utilizar sobre los residuos obtenidos del suavizamiento spline de la componente Este de Concepción y de la regresión lineal de la componente Norte y Altura de Concepción.

Identificación de modelos

Como fue mostrado previamente, los residuos de la componente Norte y Altura de Concepción no son estacionarios, por lo que se aplica diferenciación para cumplir con la condición de estacionariedad. Sin embargo, los residuos de la componente Este de Concepción resultan estacionarios, por lo que no necesitan diferenciación. A continuación se presenta en la Tabla 5.16 los resultados obtenidos del test ADF y KPSS, respecto a la aplicación de la primera diferenciación.

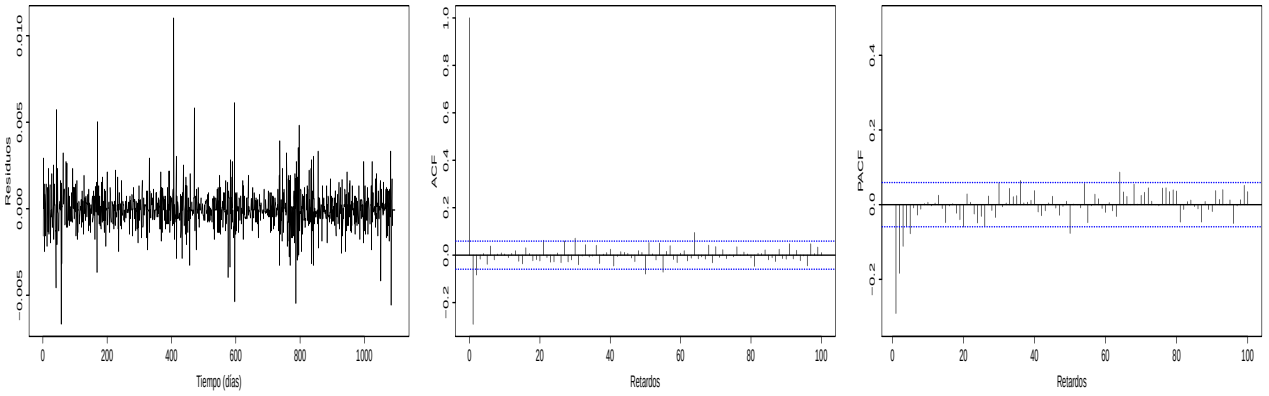
Tabla 5.16. Resultados de los test ADF y KPSS, aplicados sobre la primera diferenciación de los residuos de la componente Norte y Altura de Concepción.

Serie de tiempo	Test ADF p-valor	Resultado	Test KPSS p-valor	Resultado
Residuos componente Norte	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria
Residuos componente Altura	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria

De la Tabla 5.16 se logra apreciar que los dos test afirman que la primera diferenciación hace que los residuos de la componente Norte y Altura se vuelvan estacionarios, por consiguiente en ambos casos es aceptable para que se ajusten modelos

ARIMA(p,d,q), con d=1.

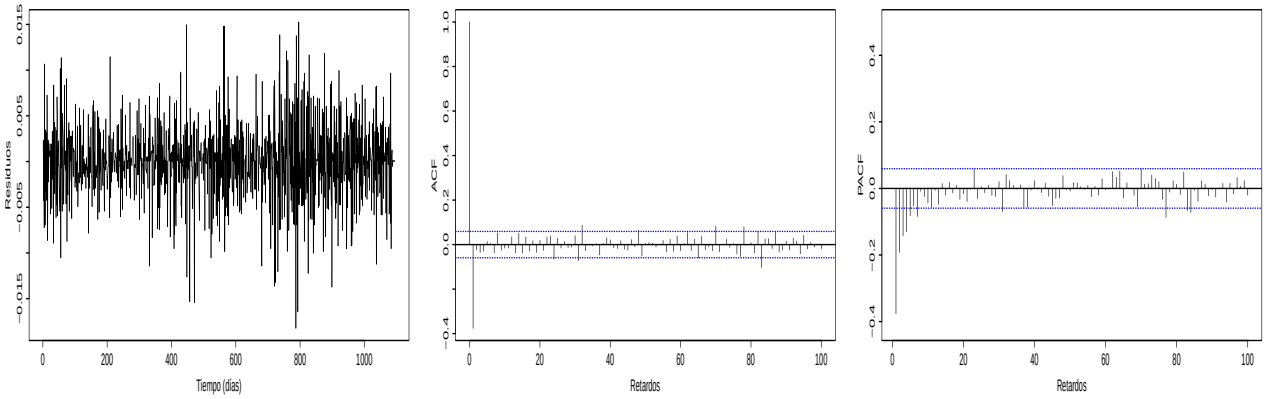
En la Figura 5.29a y 5.29d se observa que los gráficos de la primera diferenciación presentan una media constante, ya que oscilan alrededor del cero y una variabilidad con respecto a esa media también constante. En los correlogramas de la Figura 5.29b, 5.29c, 5.29e y 5.29f no se logra apreciar una clara identificación del orden del modelo, por lo que se ajustan distintos modelos ARIMA(p,d,q).



(a) Primera diferenciación: residuos componente Norte

(b) ACF

(c) PACF



(d) Primera diferenciación: residuos componente Altura

(e) ACF

(f) PACF

Figura 5.29. Gráfico de la primera diferenciación, con sus respectivos correlogramas ACF y PACF sobre los residuos de la componente Norte y Altura de Concepción.

Con respecto a los residuos de la componente Este se aprecia en la Figura 5.27a que los residuos presentan una media constante, ya que oscilan alrededor del cero y una variabilidad con respecto a esa media también constante. En los correlogramas

de la Figura 5.27b y 5.27c no se logra apreciar una clara identificación del orden del modelo, por lo que se ajustan distintos modelos ARMA(p,q).

En la Tabla 5.17 se presentan los modelos ARMA(p,q) y ARIMA(p,d,q) seleccionados para cada componente (ver (B.1), (B.2) y (B.3)). En los residuos de la componente Norte de Concepción, el menor valor AIC y BIC lo obtuvo el modelo ARIMA(1,1,1), por lo que sería el mejor modelo que se ajusta sobre los residuos. Con respecto a los residuos de la componente Este de Concepción, el menor valor AIC y BIC lo obtuvo el modelo ARMA(2,6). Sin embargo los residuos de este modelo no cumplen con el supuesto de independencia, no así el modelo ARMA(4,7), que sí cumple con este supuesto. Para identificar el mejor modelo se procede con la validación y pronóstico de ambos modelos en las siguientes secciones. Finalmente en los residuos de la componente Altura de Concepción, el menor valor AIC y BIC lo obtuvo el modelo ARIMA(2,1,1), por lo que sería el mejor modelo que se ajusta sobre los residuos.

Tabla 5.17. Modelos seleccionados para cada componente, con sus respectivos criterios de información AIC y BIC .

Serie de tiempo	Modelo	AIC	BIC
Residuos componente Norte	ARIMA(1,1,1)	-11766.17	-11751.18
Residuos componente Este	ARMA(2,6)	-11904.35	-11854.35
Residuos componente Este	ARMA(4,7)	-11891.08	-11826.09
Residuos componente Altura	ARIMA(2,1,1)	-9135.208	-9115.214

En la Tabla 5.18 se presenta el contraste de significación de los cuatro modelos seleccionados. Se puede apreciar que en el modelo ARIMA(1,1,1) los dos parámetros resultan significativos para el modelo. En el modelo ARMA(2,6) los ocho parámetros resultan significativos para el modelo y en el modelo ARMA(4,7) se aprecia que tres de los once parámetros resultan significativos para el modelo. Finalmente en el modelo ARIMA(2,1,1) los tres parámetros resultan significativos para el modelo. Los parámetros resultan significativos por tener un p-valor menor que un nivel de significación de $\alpha = 0.05$.

Tabla 5.18. Contraste de significación de los parámetros del modelo ARIMA(1,1,1), ARMA(2,6), ARMA(4,7) y ARIMA(2,1,1).

Parámetro	Estimación	Error estándar	z value	Pr(> |z|)
Residuos componente Norte: ARIMA(1,1,1)				
ar1	0.2656	0.0634	4.1899	0.00003
ma1	-0.6440	0.0503	-12.8131	< 0.00001
Residuos componente Este: ARMA(2,6)				
ar1	1.7065	0.0112	152.64570	< 0.00001
ar2	-0.9017	0.0108	-83.346300	< 0.00001
ma1	-2.1587	0.0015	-1408.3263	< 0.00001
ma2	1.1435	0.0053	214.80020	< 0.00001
ma3	0.0268	0.0103	2.5931000	0.009512
ma4	0.2117	0.0137	15.495800	< 0.00001
ma5	-0.2751	0.0152	-18.141100	< 0.00001
ma6	0.0521	0.0081	6.4779000	< 0.00001
Residuos componente Este: ARMA(4,7)				
ar1	0.8320	0.1764	4.7181	< 0.00001
ar2	0.4004	0.1986	2.0162	0.04378
ar3	-0.4141	0.2252	-1.8389	0.06593
ar4	-0.2203	0.1820	-1.2103	0.22618
ma1	-1.2794	0.2276	-5.6224	< 0.00001
ma2	-0.4947	0.2827	-1.7496	0.08018
ma3	0.5299	0.3117	1.7002	0.08909
ma4	0.4225	0.2325	1.8173	0.06918
ma5	-0.0954	0.0713	-1.3366	0.18137
ma6	0.0127	0.0594	0.2142	0.83042
ma7	-0.0951	0.0495	-1.9239	0.05436
Residuos componente Altura: ARIMA(2,1,1)				
ar1	0.2881	0.0517	5.57530	< 0.00001
ar2	0.0861	0.0409	2.10430	0.03535
ma1	-0.8069	0.0404	-19.9573	< 0.00001

En la Figura 5.30, 5.31, 5.32 y 5.33 se presentan los ajustes del modelo ARIMA(1,1,1), ARMA(2,6), ARMA(4,7) y ARIMA(2,1,1). Se observa a simple vista que los cuatro modelos se ajustan de buena manera sobre los residuos. Sin embargo, para determinar si los modelo realmente son adecuados para la representación de los datos, se deben analizar los residuos. En el caso de los modelos ARMA(2,6) y ARMA(4,7), para determinar cuál de los dos modelo es el adecuado para la representación de los datos, se deben analizar los errores del pronóstico.

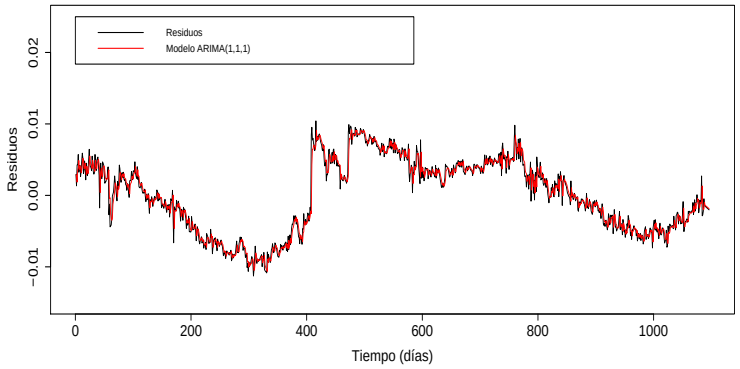


Figura 5.30. Ajuste del modelo ARIMA(1,1,1) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Norte de Concepción.

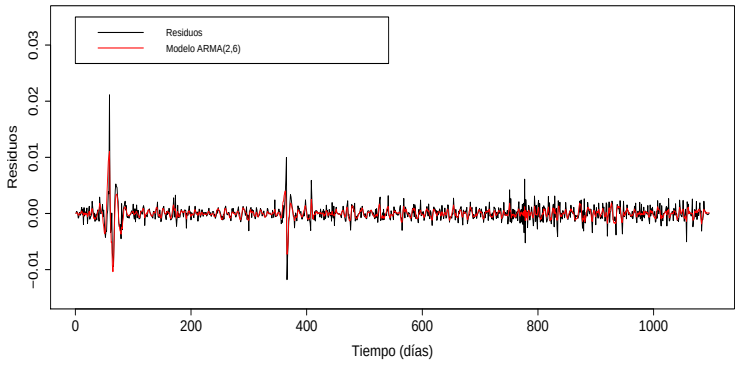


Figura 5.31. Ajuste del modelo ARMA(2,6) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Este de Concepción.

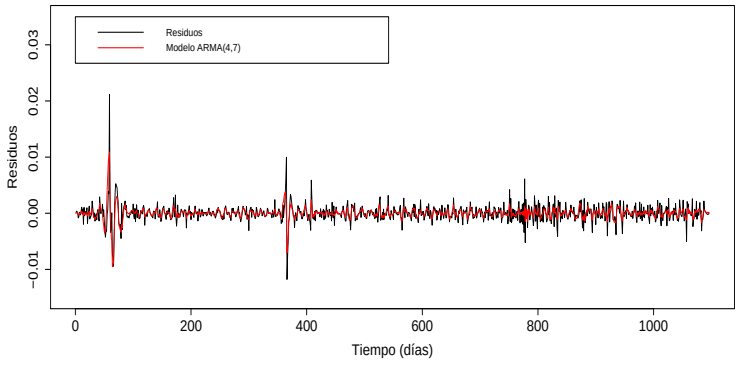


Figura 5.32. Ajuste del modelo ARMA(4,7) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Este de Concepción.

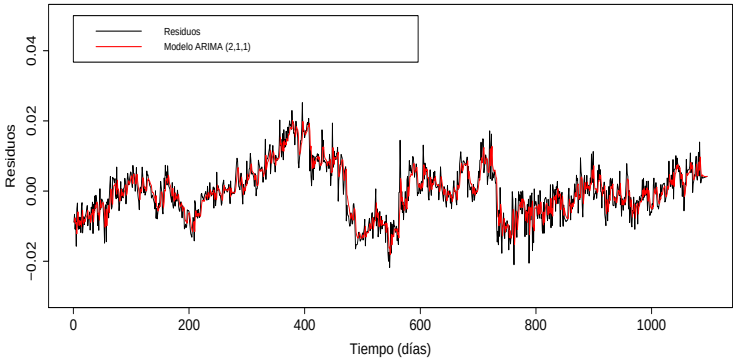
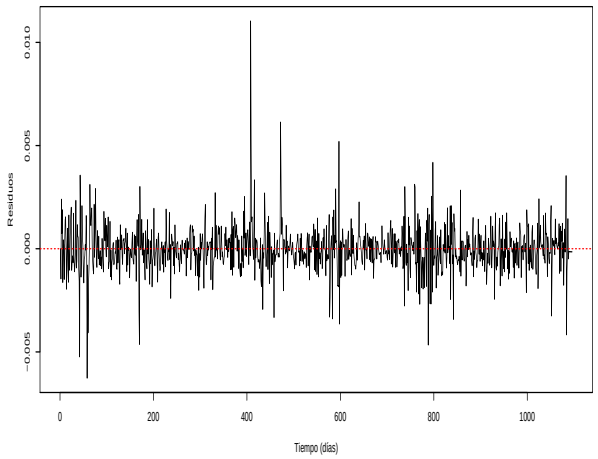


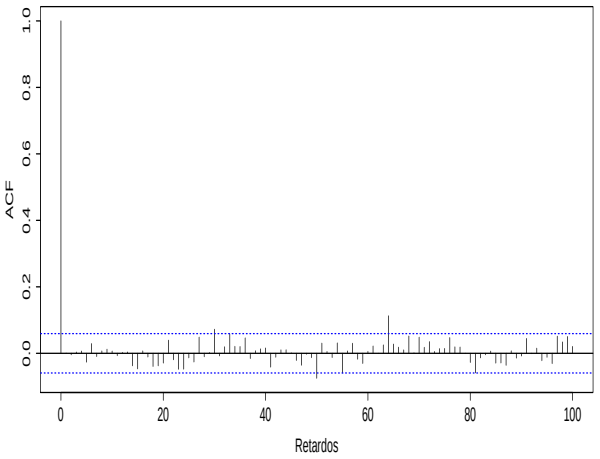
Figura 5.33. Ajuste del modelo ARIMA(2,1,1) sobre los residuos obtenidos de la regresión lineal, en la serie de tiempo componente Altura de Concepción.

5.4.2.3. Diagnóstico del modelo

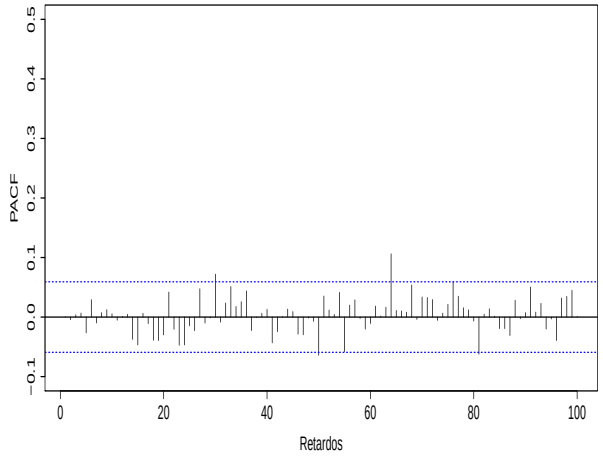
En la Figura 5.34a, 5.35a, 5.36a y 5.37a se observa que los residuos de los modelos ARIMA(1,1,1), ARMA(2,6), ARMA(4,7) y ARIMA(2,1,1) presentan una media constante y una varianza con respecto a la media también constante. En la Figura 5.34b, 5.36b y 5.37b se apreciar el correlograma ACF de los residuos para los tres modelos, donde se puede notar la presencia de independencia de las observaciones, ya que las autocorrelaciones se mantienen dentro de los intervalos de confianza. Sin embargo en la Figura 5.35b al observar el correlograma de los residuos se sospecha de la presencia de independencia entre las observaciones, ya que no todas las autocorrelaciones se mantienen dentro de los intervalos de confianza. Finalmente en la Figura 5.34d, 5.35d, 5.36d y 5.37d se observa a simple vista que los residuos de los cuatro modelos se aproximan a una distribución normal.



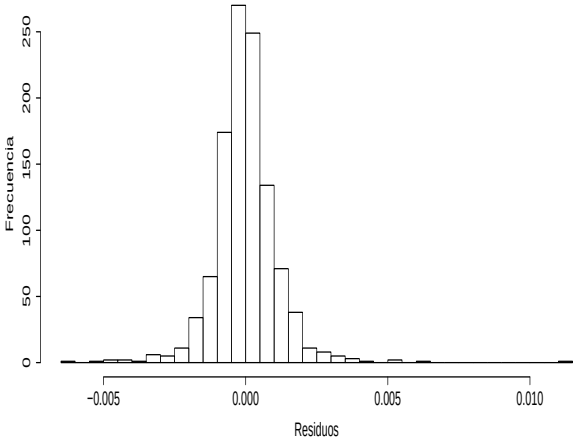
(a) Residuos



(b) ACF

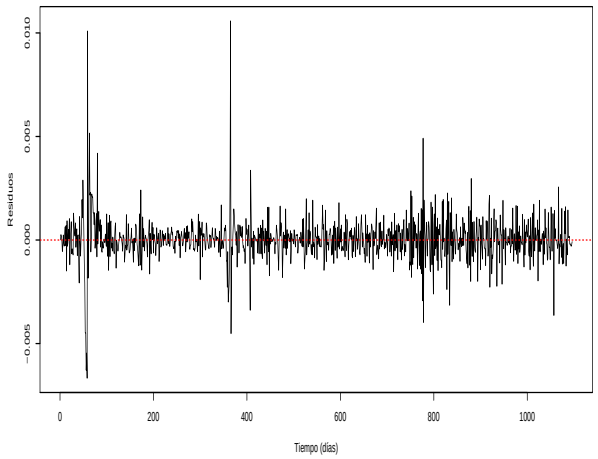


(c) PACF

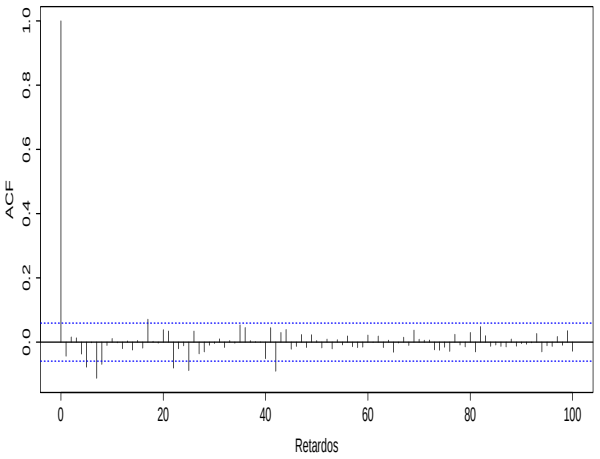


(d) Distribución de frecuencias

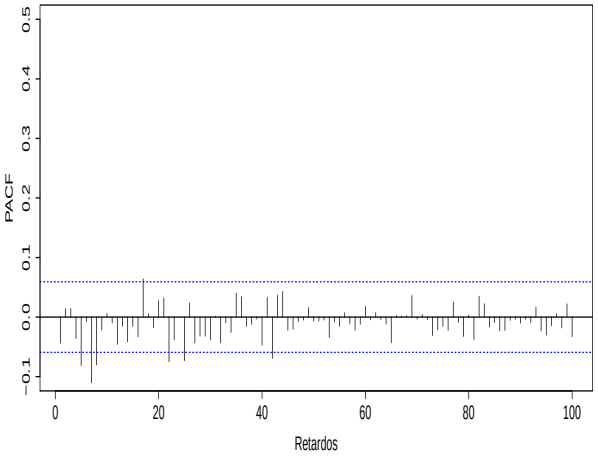
Figura 5.34. Gráfico de dispersión de los residuos del modelo ARIMA(1,1,1), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Concepción.



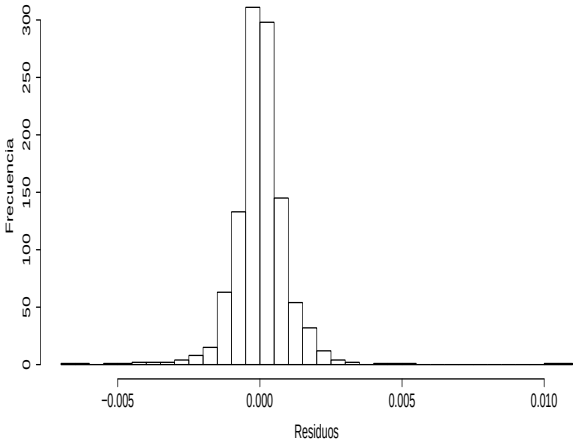
(a) Residuos



(b) ACF

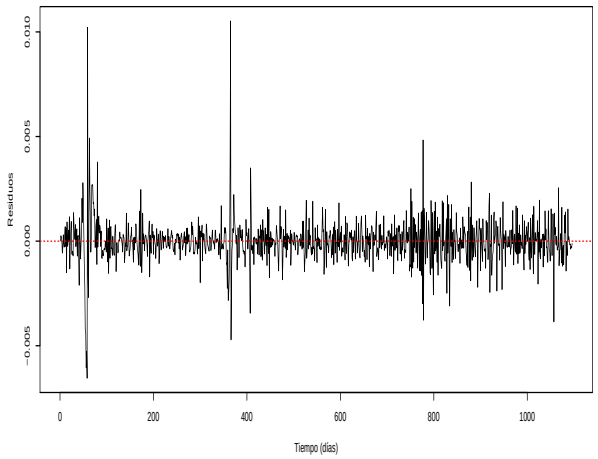


(c) PACF

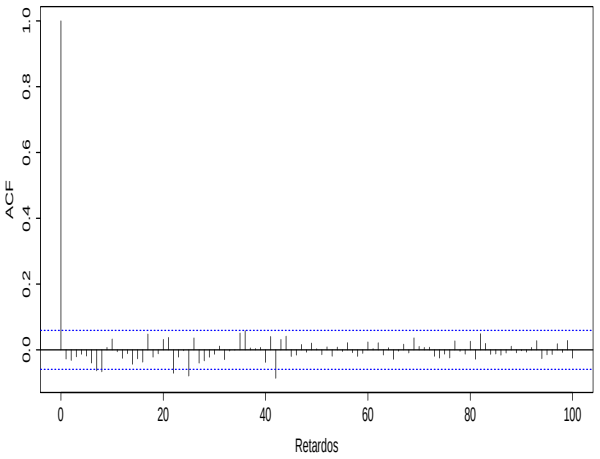


(d) Distribución de frecuencias

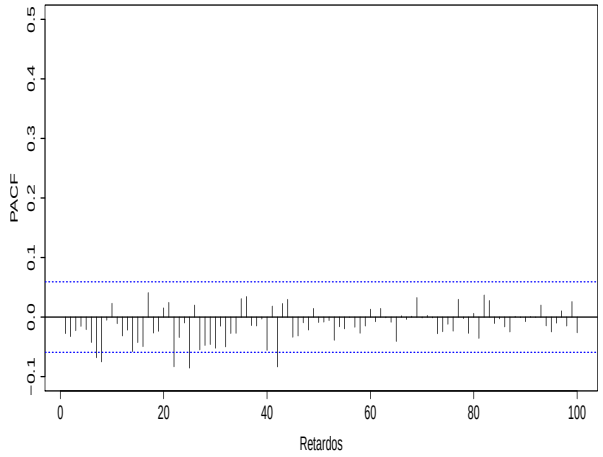
Figura 5.35. Gráfico de dispersión de los residuos del modelo ARMA(2,6), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Concepción.



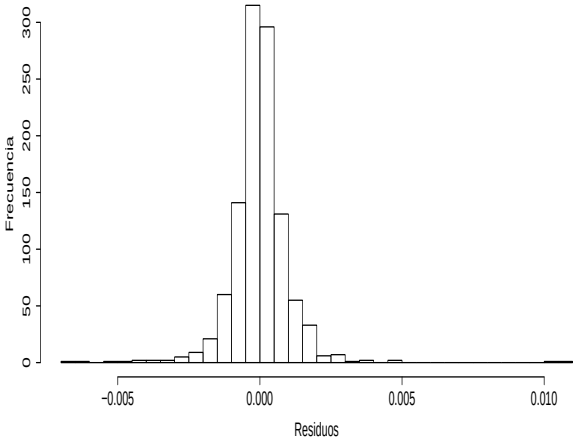
(a) Residuos



(b) ACF

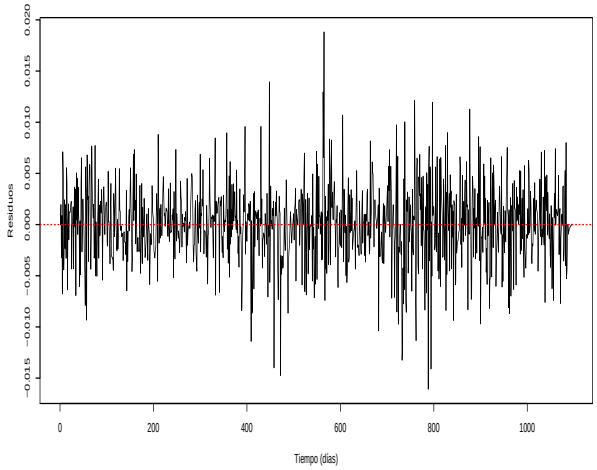


(c) PACF

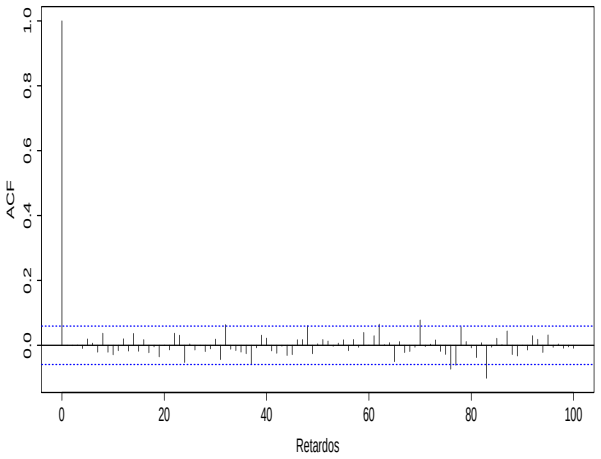


(d) Distribución de frecuencias

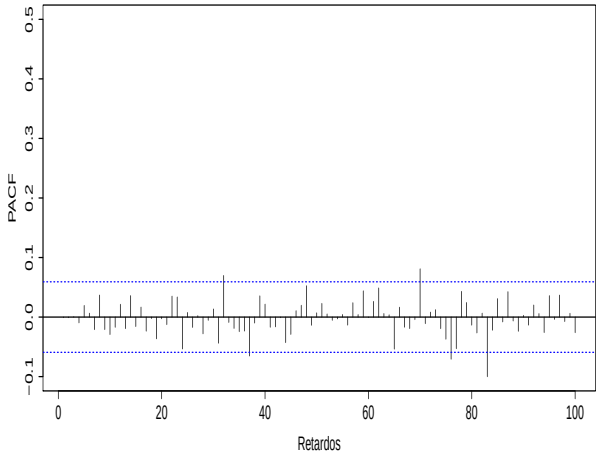
Figura 5.36. Gráfico de dispersión de los residuos del modelo ARMA(4,7), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Concepción.



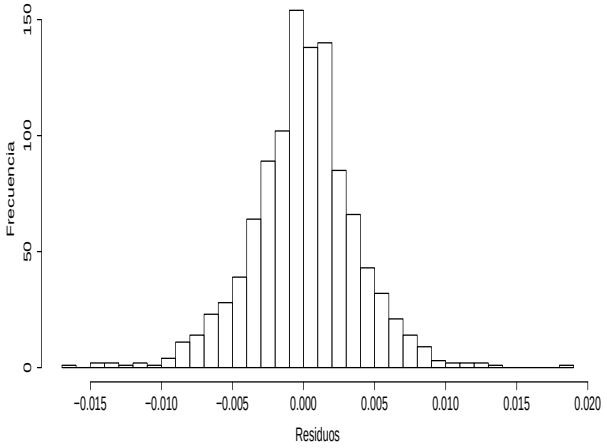
(a) Residuos



(b) ACF



(c) PACF



(d) Distribución de frecuencias

Figura 5.37. Gráfico de dispersión de los residuos del modelo ARIMA(2,1,1), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Concepción.

A continuación se presentan en la Tabla 5.19 los resultados obtenidos del test Ljung-Box para comprobar independencia entre los residuos de cada modelo.

Tabla 5.19. Test Ljung-Box sobre los residuos de cada modelo.

Componente	Residuos	Test Ljung-Box p-valor	Resultado
Norte	Modelo ARIMA(1,1,1)	0.9556	Independientes
Este	Modelo ARMA(2,3)	0.0066	No son independientes
Este	Modelo ARMA(4,7)	0.2031	Independientes
Altura	Modelo ARIMA(2,1,1)	0.9804	Independientes

Finalmente los modelos para cada componente quedan expresados de la siguiente manera:

Componente Norte: modelo ARIMA(1,1,1)

$$\varepsilon_t = 0.265626\varepsilon_{t-1} + \omega_t - (-0.644047)\omega_{t-1}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

Componente Este: modelo ARMA(2,6)

$$\varepsilon_t = 1.706500\varepsilon_{t-1} + -0.901690\varepsilon_{t-2} + \omega_t - (-2.158700)\omega_{t-1} - (1.143500)\omega_{t-2} - (0.026808)\omega_{t-3} \\ -(0.211720)\omega_{t-4} - (-0.275060)\omega_{t-5} - (0.052148)\omega_{t-6}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

Componente Este: modelo ARMA(4,7)

$$\varepsilon_t = 0.8322020\varepsilon_{t-1} + 0.400430\varepsilon_{t-2} + -0.414080\varepsilon_{t-3} + -0.220270\varepsilon_{t-4} + \omega_t - (-1.279400)\omega_{t-1}$$

$$-(-0.494690)\omega_{t-2} - (0.529920)\omega_{t-3} - (0.422520)\omega_{t-4} - (-0.095359)\omega_{t-5} - (0.012711)\omega_{t-6} \\ -(-0.095148)\omega_{t-7}\ ,\ \ \omega_t \sim RB(0\sigma^2)$$

Componente Altura: modelo ARIMA(2,1,1)

$$\varepsilon_t = 0.288089\varepsilon_{t-1} + 0.086090\varepsilon_{t-2} + \omega_t - (-0.806948)\omega_{t-1}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

5.4.2.4. Pronóstico

En esta sección se proceden a realizar dos tipos de pronósticos. El primer pronóstico es con respecto a la parte determinística (regresión lineal y suavizamiento spline) y el segundo pronóstico es con respecto a la parte estocástica (modelos ARMA y ARIMA).

Pronóstico determinístico

A continuación se procede con el pronóstico obtenido de la estimación sobre la tendencia. En la Figura 5.38, 5.39 y 5.40 se representan gráficamente con color plomo los intervalos de confianza, es decir, los límites de predicción y con color azul las predicciones.

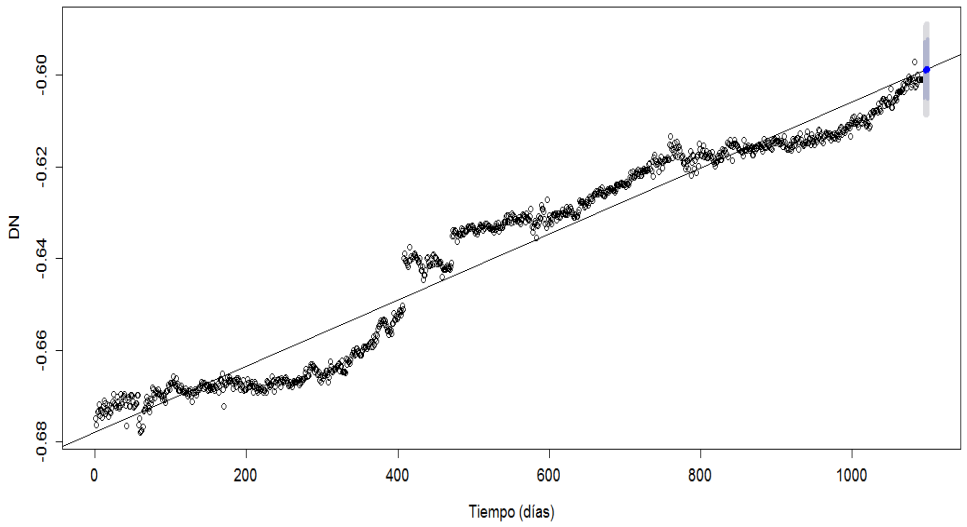


Figura 5.38. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Norte de Concepción.

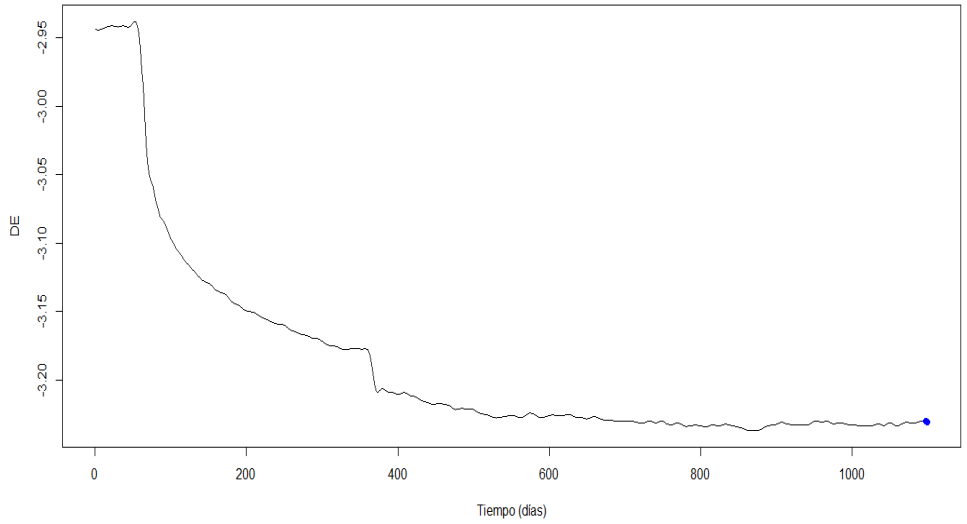


Figura 5.39. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Este de Concepción.

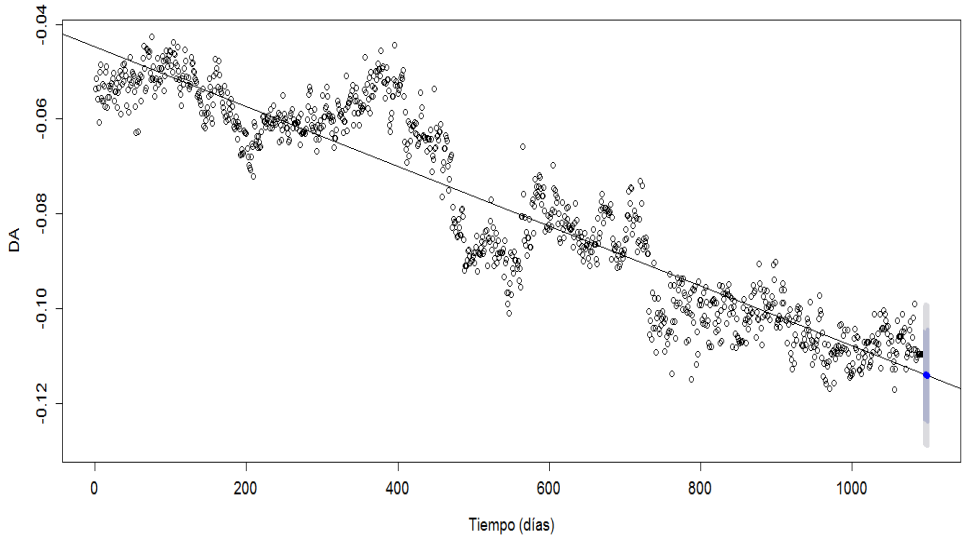


Figura 5.40. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto a la regresión lineal sobre la componente Altura de Concepción.

A continuación se presentan en la Tabla 5.20 los resultados obtenidos de las predicciones.

Tabla 5.20. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Este de Concepción y la regresión lineal sobre la componente Norte y Altura de Concepción.

Día	Mes	Año	Pronóstico componente Norte	Pronóstico componente Este	Pronóstico componente Altura
1	Enero	2013	-0.59881	-3.23004	-0.11396
2	Enero	2013	-0.59873	-3.23017	-0.11402
3	Enero	2013	-0.59866	-3.23030	-0.11408
4	Enero	2013	-0.59859	-3.23041	-0.11415

Pronóstico de la componente residual

Las predicciones con respecto a los residuos obtenidos en la regresión lineal sobre la componente Norte y Altura de Concepción se presenta gráficamente en la Figura 5.41 y 5.44. Con respecto a los residuos obtenidos en el suavizamiento spline sobre la componente Este de Concepción se proponen dos modelos, por lo que se debe escoger el mejor. En la sección 4.5.1 se presentó el método del error de pronóstico RMSE. Con el fin de seleccionar el mejor modelo, se procederá a realizar el pronóstico para diciembre del 2012 con los dos modelos seleccionados (ARMA(2,6) y ARMA(4,7)) y se determinará el mejor mediante el RMSE. En la Figura 5.42 y 5.43 se presentan las predicciones para los 31 días de diciembre del 2012.

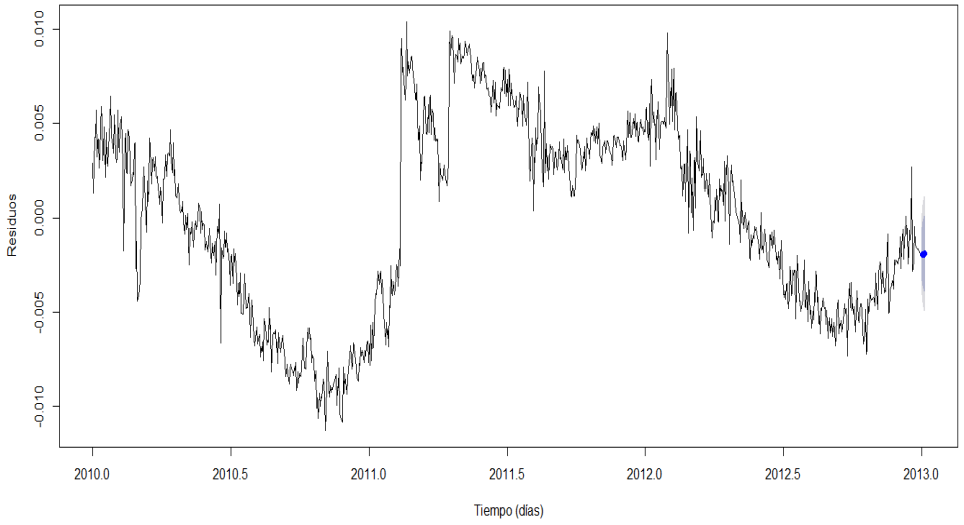


Figura 5.41. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(1,1,1) sobre los residuos de la componente Norte de Concepción.

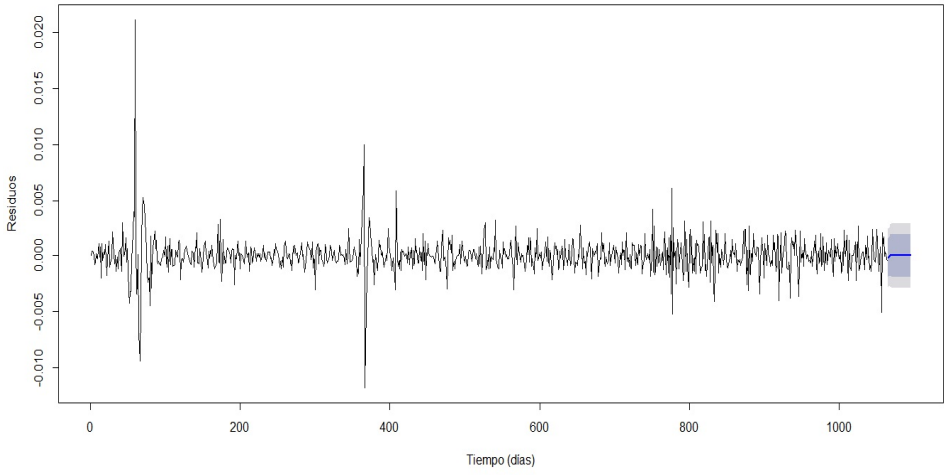


Figura 5.42. Gráfico de predicción para el mes de diciembre del 2012, utilizando el modelo ARMA(2,6) sobre los residuos de la componente Este de Concepción.

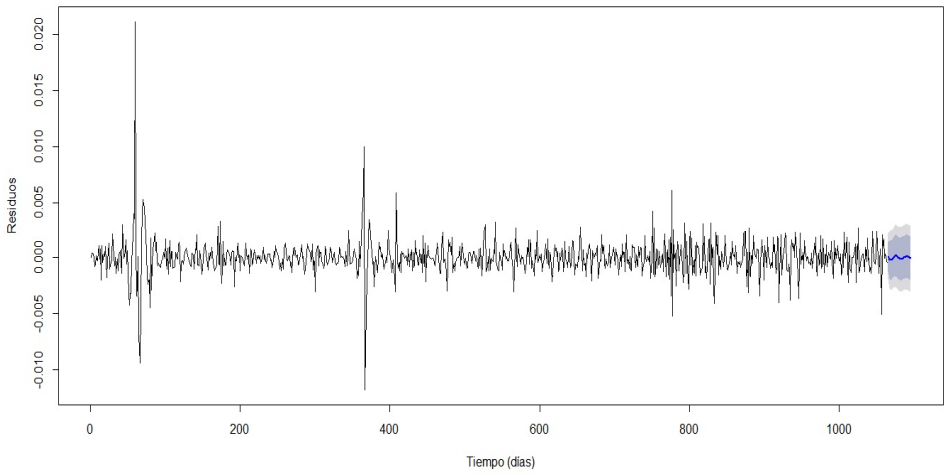


Figura 5.43. Gráfico de predicción para el mes de diciembre del 2012, utilizando el modelo ARMA(4,7) sobre los residuos de la componente Este de Concepción.

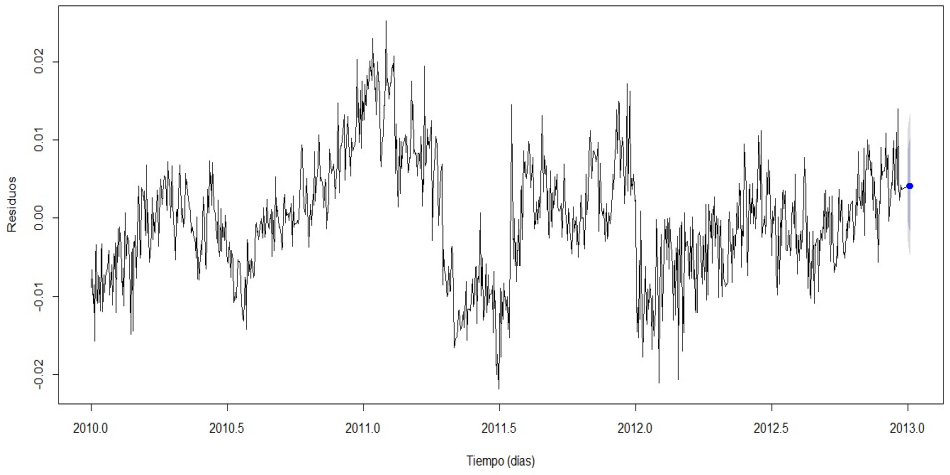


Figura 5.44. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARIMA(2,1,1) sobre los residuos de la componente Altura de Concepción.

En los gráficos de la Figura 5.42 y 5.43 se percibe a simple vista una diferencia, pero no se logra identificar el mejor modelo, por lo que se procede con el cálculo de la medida del error de pronóstico RMSE. En la Tabla 5.21 se presentan los resultados obtenidos del RMSE, para los modelos ARMA(2,6) y ARMA(4,7).

Tabla 5.21. Resultados obtenidos del cálculo RMSE, para los modelos ARMA(2,6) y ARMA(4,7)

Error de pronóstico	ARMA(2,6)	ARMA(4,7)
RMSE	0.0012	0.0011

El modelo con mayor precisión cuando se pronostique será aquel que presente el valor más pequeño de la medida del error RMSE. En esta ocasión el modelo ARMA(4,7) es el que obtiene el menor valor, pero también es importante destacar que la diferencia no es muy grande entre los dos modelos, por lo que es mejor escoger el modelo ARMA(2,6) ya que es más parsimonioso, es decir, tiene menos parámetros. Finalmente se concluye que el modelo ARMA(2,6) es el más adecuado para realizar predicciones sobre los residuos del suavizamiento spline de la componente Este de Concepción.
A continuación se proceden a representar gráficamente las predicciones obtenidas para los primeros cuatro días de de enero del 2013, utilizando el modelo ARMA(2,6).

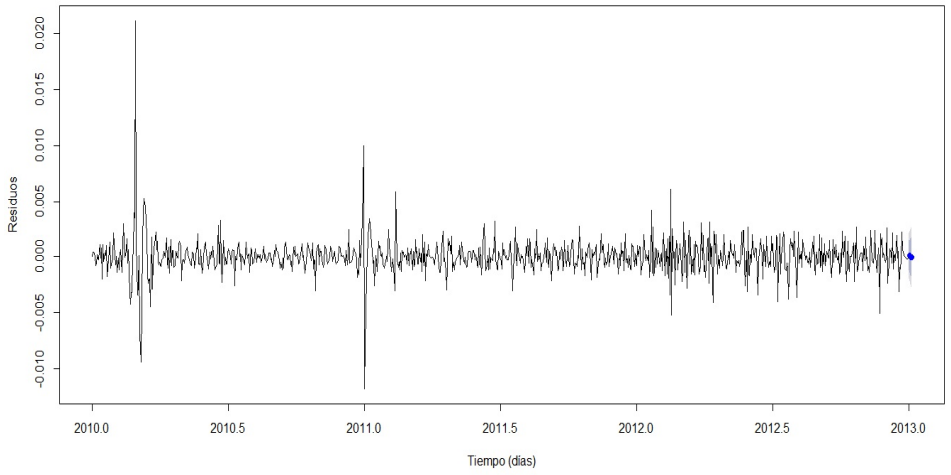


Figura 5.45. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(2,6) sobre los residuos de la componente Este de Concepción.

En la tabla 5.22 se presentan los resultados obtenidos de las predicciones sobre los residuos de la componente Norte, Este y Altura de Concepción.

Tabla 5.22. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto a los residuos de la componente Norte, Este y Altura de Concepción.

Día	Mes	Año	Pronóstico residuos DN Modelo ARIMA(1,1,1)	Pronóstico residuos DE Modelo ARMA(2,6)	Pronóstico residuos DA Modelo ARIMA(2,1,1)
1	Enero	2013	-0.001947	0.000080	0.004194
2	Enero	2013	-0.001926	0.000002	0.004199
3	Enero	2013	-0.001921	-0.000050	0.004200
4	Enero	2013	-0.001920	-0.000058	0.004201

Pronóstico general

En la Figura 5.46, 5.47 y 5.48 se representan gráficamente las predicciones generales para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte, Este y Altura de Concepción.

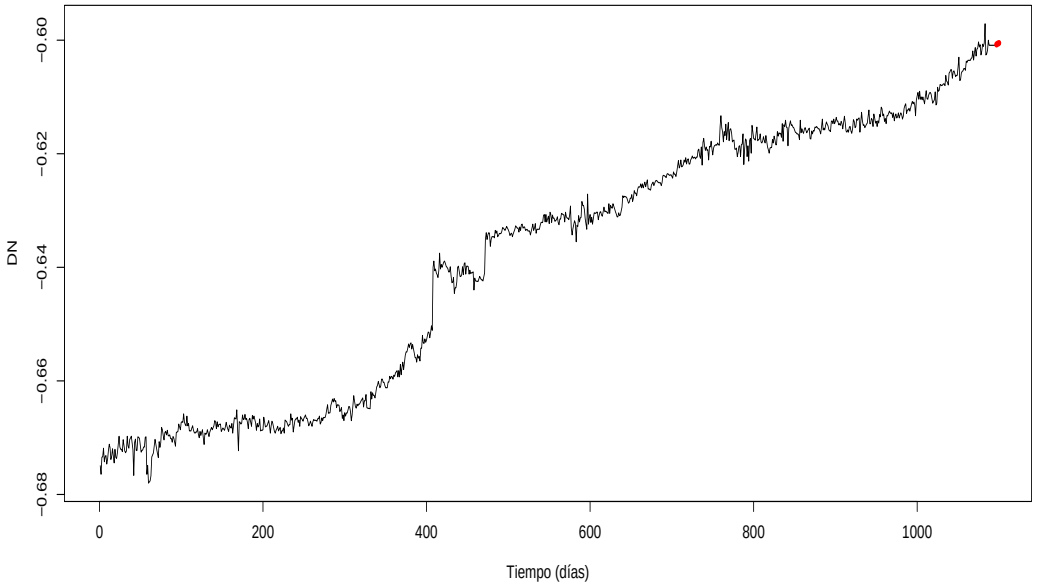


Figura 5.46. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Norte de Concepción.

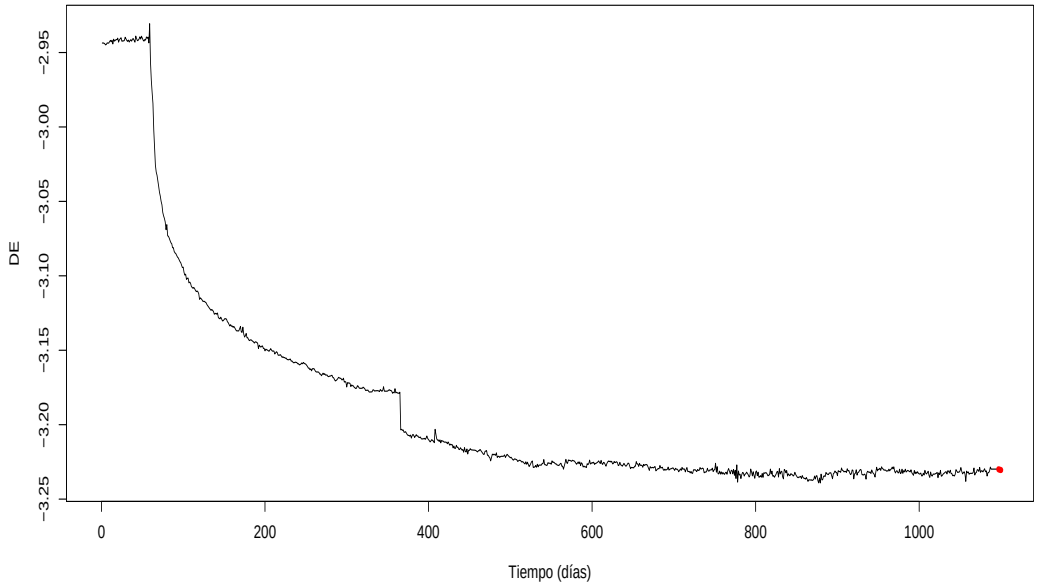


Figura 5.47. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Este de Concepción.

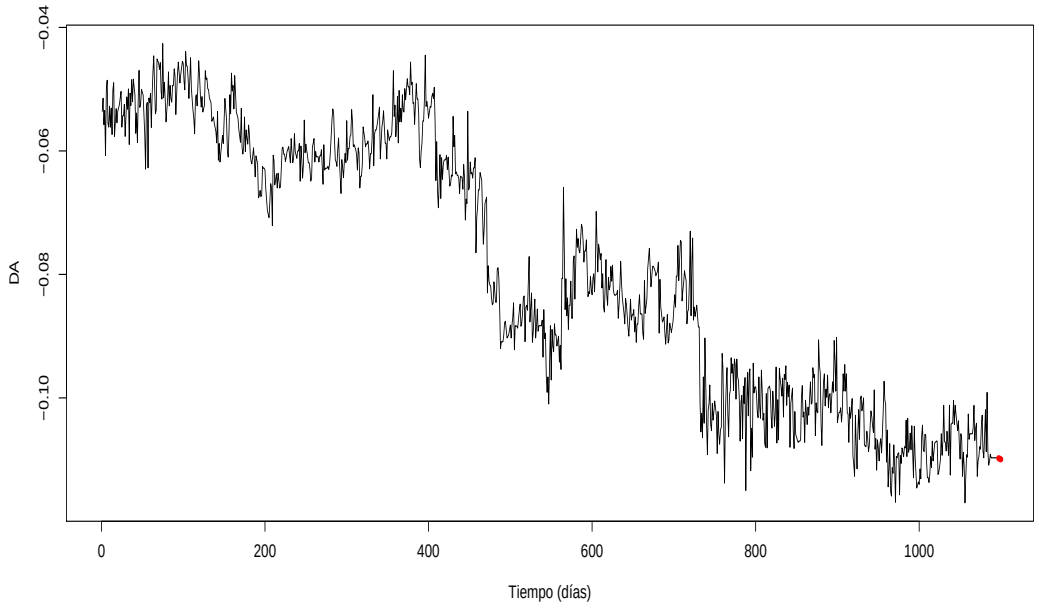


Figura 5.48. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a la serie de tiempo componente Altura de Concepción.

En la Tabla 5.23 se presentan las predicciones generales, para los primeros cuatro días de enero del 2013.

Tabla 5.23. Predicciones generales obtenidas para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte, Este y Altura de Concepción.

Día	Mes	Año	Predicción parte determinística	Predicción parte estocástica	Predicción general
Componente Norte					
1	Enero	2013	-0.59881	-0.001947	-0.60076
2	Enero	2013	-0.59873	-0.001926	-0.60066
3	Enero	2013	-0.59866	-0.001921	-0.60058
4	Enero	2013	-0.59859	-0.001920	-0.60051
Componente Este					
1	Enero	2013	-3.23004	0.000080	-3.22996
2	Enero	2013	-3.23017	0.000002	-3.23017
3	Enero	2013	-3.23030	-0.000050	-3.23035
4	Enero	2013	-3.23041	-0.000058	-3.23047
Componente Altura					
1	Enero	2013	-0.11396	0.004194	-0.10977
2	Enero	2013	-0.11402	0.004199	-0.10982
3	Enero	2013	-0.11408	0.004200	-0.10988
4	Enero	2013	-0.11415	0.004201	-0.10995

Predicciones finales y módulo del vector desplazamiento

Con las predicciones finales obtenidas para la componente Norte, Este y Altura de Concepción en la Tabla 5.23, se realiza el cálculo del módulo del vector desplazamiento. En la Tabla 5.24 se obtienen los resultados.

Tabla 5.24. Predicciones finales y módulo del vector desplazamiento obtenidos para los primeros cuatro días de enero del 2013, con respecto a la componente Norte, Este y Altura de Concepción.

Día	Mes	Año	Predicción final DN	Predicción final DE	Predicción final DA	Módulo
1	Enero	2013	-0.60076	-3.22996	-0.10977	3.28719
2	Enero	2013	-0.60066	-3.23017	-0.10982	3.28738
3	Enero	2013	-0.60058	-3.23035	-0.10988	3.28754
4	Enero	2013	-0.60051	-3.23047	-0.10995	3.28765

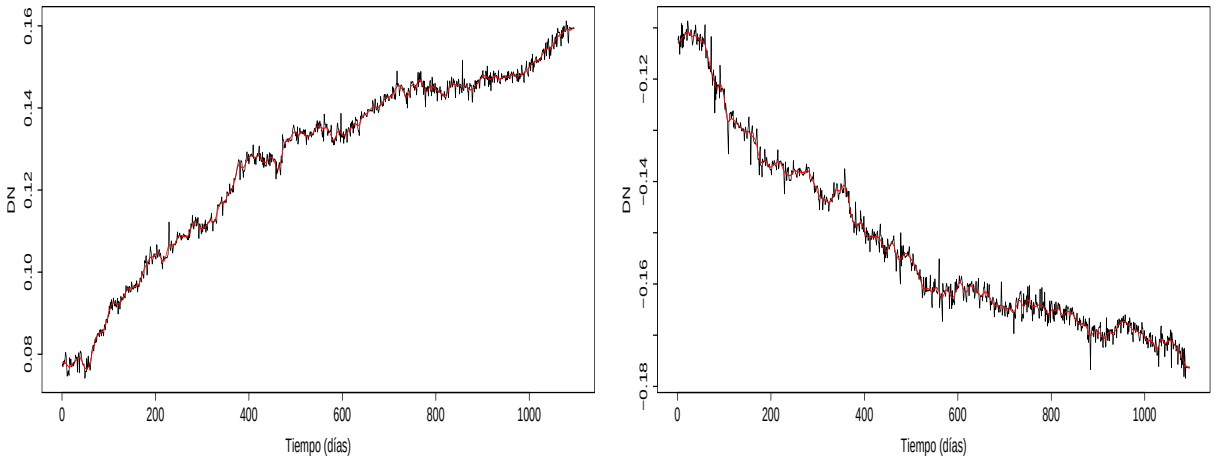
De la tabla 5.24 se puede deducir que el desplazamiento resultante es un vector de longitud 3.28719 cm para el primer día, 3.28738 cm para el segundo día, 3.28754 cm para el tercer día y 3.28765 cm para el cuarto día.

5.4.3. Estimación sobre las series de tiempo en Llafenco

En la Figura 5.5a y 5.6a se observa que las series de tiempo presentan una tendencia muy marcada, por consiguiente se procede con la aplicación de la estimación de la tendencia a través de suavizamiento spline.

5.4.3.1. Estimación de la tendencia

En la Figura 5.49a y 5.49b se observa el suavizamiento spline sobre la serie de tiempo de la componente Norte y Este de Llafenco.



(a) Componente Norte **(b)** Componente Este

Figura 5.49. Suavizamiento spline sobre las componente Norte y Altura de Llafenco.

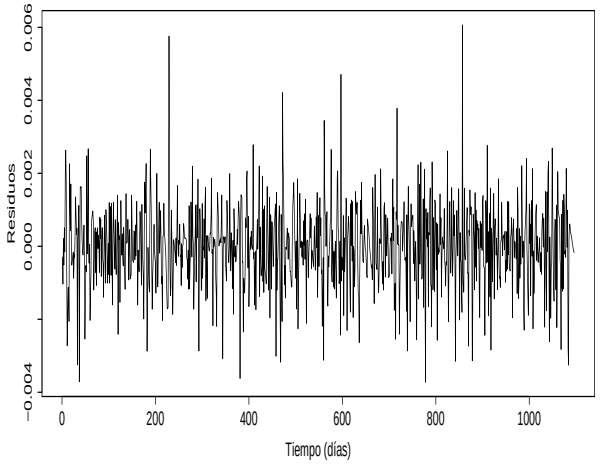
En Tabla 5.25 se presentan los parámetros de suavizamiento, con los correspondientes valores de λ, en la componente Norte y Este de Llafenco.

Tabla 5.25. Resultados obtenidos de la estimación de la tendencia sobre la componentes Norte y Este de Llafenco.

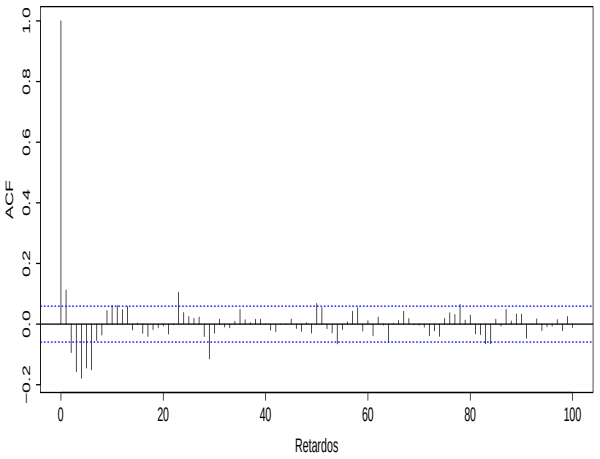
Componente	Parámetro de suavizamiento	λ
Norte	0.2006	0.000000049
Este	0.2164	0.000000063

Residuos obtenidos del suavizamiento spline

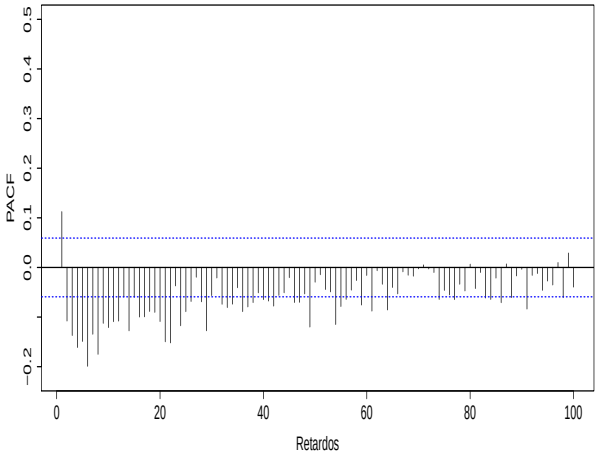
Para la estimación de modelos sobre la componente Norte y Este de Llafenco, se utilizarán los residuos obtenidos del suavizamiento spline (analizadas anteriormente).



(a) Componente Norte

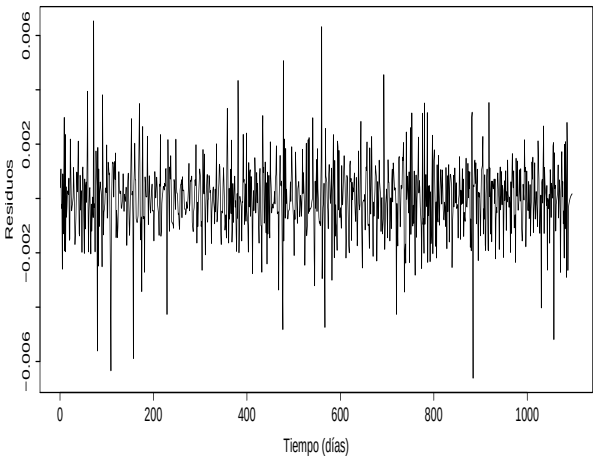


(b) ACF

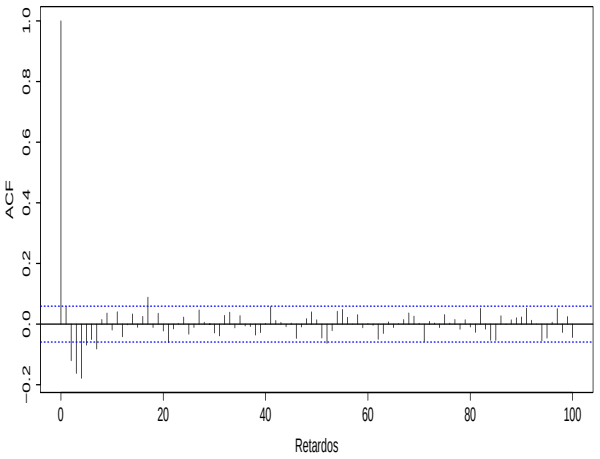


(c) PACF

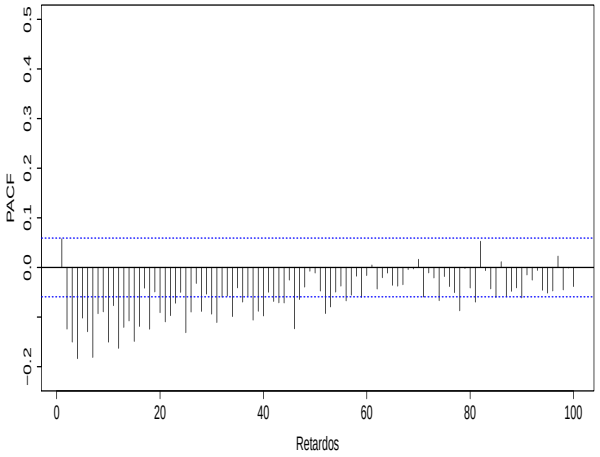
Figura 5.50. Gráfico de los residuos obtenidos en el suavizamiento spline, con sus respectivos gráficos ACF y PACF, para la componente Norte de Llafenco.



(a) Componente Este



(b) ACF



(c) PACF

Figura 5.51. Gráfico de los residuos obtenidos en el suavizamiento spline, con sus respectivos gráficos ACF y PACF, para la componente Este de Llafenco.

En la Figura 5.50b y 5.51b se observa que el ACF de los residuos no presentarían independencia, ya que las autocorrelaciones se encuentran fuera de los intervalos de confianza. Finalmente sobre los residuos de la Figura 5.50a y 5.51a, se debe determinar la existencia de estacionariedad, por lo que se procede con la aplicación del test ADF y KPSS. Los resultados se presentan en la Tabla 5.26.

Tabla 5.26. Resultados de los test ADF y KPSS sobre los residuos de la componentes Norte y Este de Llafenco.

Residuos	Test ADF p-valor	Resultado	Test KPS p-valor	Resultado
Componente Norte	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria
Componente Este	0.01	Serie de tiempo estacionaria	0.10	Serie de tiempo estacionaria

De los resultados obtenidos en la Tabla 5.26 se puede deducir que los residuos de la componente Norte y Este de Llafenco son estacionarios.

5.4.3.2. Estimación de la componente residual

A continuación se identifican los modelos a utilizar sobre la serie de tiempo componente Altura de Llafenco y los residuos obtenidos del suavizamiento spline de la componente Norte y Este de Llafenco.

Identificación de modelos

Como fue mostrado previamente, los residuos de la componente Norte, Este y la serie de tiempo componente Altura de Llafenco son estacionarios, por lo que no necesitan diferenciación. Con respecto a los residuos de la componente Norte y Este de Llafenco se aprecia en la Figura 5.50a y 5.51a que los residuos presentan una media constante, ya que oscilan alrededor del cero y una variabilidad con respecto a esa media también constante. En la serie de tiempo componente Altura de Llafenco presentada en la Figura 5.6c se puede apreciar que las observaciones presentan una media constante y una variabilidad con respecto a esa media también constante. En los correlogramas de la Figura 5.50b, 5.50c, 5.51b, 5.51b y 5.6d no se logra apreciar una clara identificación del orden del modelo, por lo que se ajustan distintos modelos ARMA(p,q).

En la Tabla 5.27 se presentan los modelos ARMA(p,q) seleccionados para cada componente (ver (C.1), (C.2) y (C.3)). En los residuos de la componente Norte de Llafenco, el menor valor AIC y BIC lo obtuvo el modelo ARMA(3,3), pero los residuos del modelo no presentan independencia. Sin embargo el modelo ARMA(3,4) presenta criterios de información cercanos con el modelo ARMA(3,3) y además los residuos de este modelo podrían presentar independencia, por lo que se escoge como el modelo que se ajusta de mejor manera sobre los residuos. Con respecto a los residuos de la componente Este de Llafenco, el menor valor AIC y BIC lo obtuvo el modelo ARMA(4,7). Finalmente en la componente Altura de Llafenco, el menor

valor AIC y BIC lo obtuvo el modelo ARMA(2,2), por lo que sería el mejor modelo que se ajusta sobre los datos.

Tabla 5.27. Modelos seleccionados para cada componente, con sus respectivos criterios de información AIC y BIC .

Serie de tiempo	Modelo	AIC	BIC
Residuos componente Norte	ARMA(3,4)	-12342.29	-12297.30
Residuos componente Este	ARMA(2,6)	-11904.35	-11854.35
Componente Altura	ARMA(2,2)	-8422.074	-8392.078

En la Tabla 5.28 se presenta el contraste de significación de los tres modelos seleccionados. Se puede apreciar que en el modelo ARMA(3,4) los siete parámetros resultan significativos para el modelo, en el modelo ARMA(4,7) los ocho parámetros resultan significativos para el modelo y finalmente en el modelo ARMA(2,2) ocho de los diez parámetros resultan significativos para el modelo. Esto por tener un p-valor menor que un nivel de significación de $\alpha = 0.05$.

Tabla 5.28. Contraste de significación de los parámetros del modelo ARMA(3,4), ARMA(4,7) y ARMA(2,2).

Parámetro	Estimación	Error estándar	z value	$Pr(>\|z\|)$
Residuos componente Norte: ARMA(3,4)				
ar1	0.6632	0.0352	18.8377	< 0.00001
ar2	0.6820	0.0638	10.6924	< 0.00001
ar3	-0.6907	0.0440	-15.6906	< 0.00001
ma1	-1.0155	0.0159	-64.0181	< 0.00001
ma2	-0.9267	0.0530	-17.4806	< 0.00001
ma3	0.9030	0.0509	17.7286	< 0.00001
ma4	0.0396	0.0138	2.86680	0.004146
Residuos componente Este: ARMA(4,7)				
ar1	0.3877	0.0403	9.62140	< 0.00001
ar2	0.5229	0.0682	7.66590	< 0.00001
ar3	0.4621	0.0426	10.8514	< 0.00001
ar4	-0.7449	0.0327	-22.7549	< 0.00001
ma1	-0.8089	0.0260	-31.1439	< 0.00001
ma2	-0.8504	0.0776	-10.9555	< 0.00001
ma3	-0.4363	0.0692	-6.30120	< 0.00001
ma4	1.1938	0.0501	23.8085	< 0.00001
ma5	0.1485	0.0516	2.87850	0.003995
ma6	-0.0201	0.0443	-0.45290	0.650637
ma7	-0.2260	0.0290	-7.79360	< 0.00001
Serie componente Altura: ARMA(2,2)				
ar1	1.5567	0.0999	15.5753	< 0.00001
ar2	-0.5640	0.0972	-5.80190	< 0.00001
ma1	-1.0939	0.1093	-10.0104	< 0.00001
ma2	0.1853	0.0820	2.25940	0.02386

En la Figura 5.52, 5.53 y 5.54 se presentan los ajustes del modelo ARMA(3,4), ARMA(4,7) y ARMA(2,2). Se observa a simple vista que los tres modelos se ajustan

de buena manera sobre los residuos. Sin embargo, para determinar si los modelo realmente son adecuados para la representación de los datos, se deben analizar los residuos.

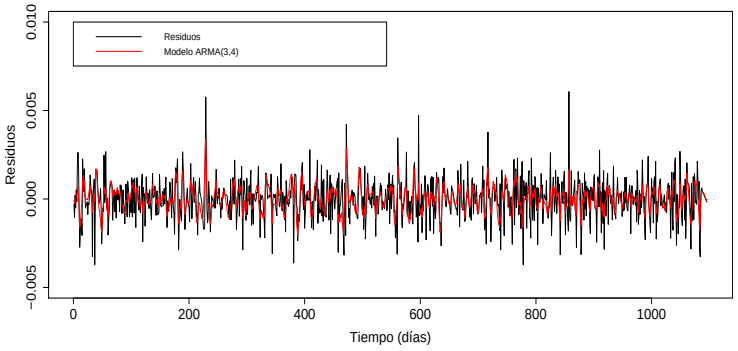


Figura 5.52. Ajuste del modelo ARMA(3,4) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Norte de Llafenco.

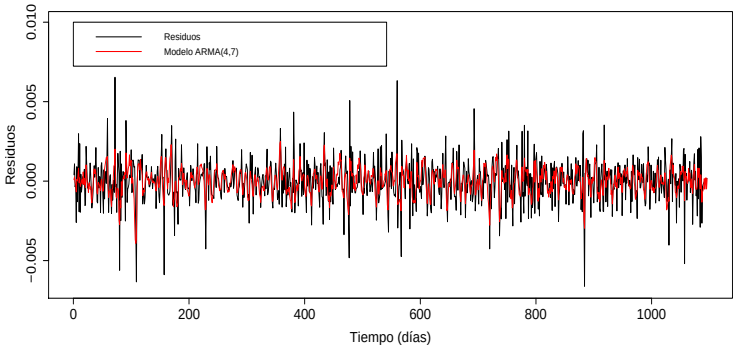


Figura 5.53. Ajuste del modelo ARMA(4,7) sobre los residuos obtenidos del suavizamiento spline, en la serie de tiempo componente Este de Llafenco.

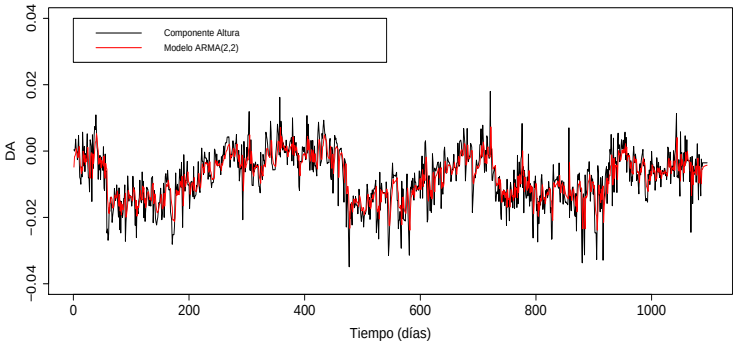
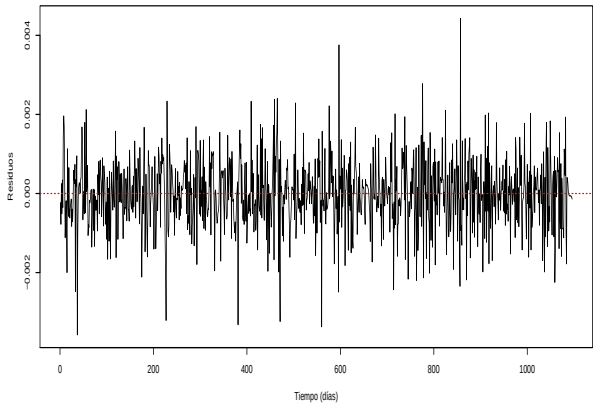


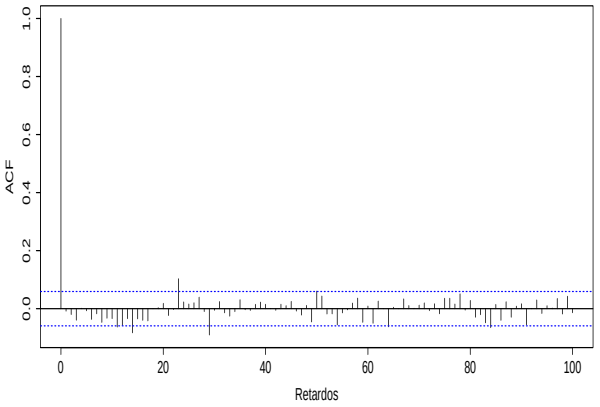
Figura 5.54. Ajuste del modelo ARMA(2,2) sobre los residuos de la serie de tiempo componente Altura de Llafenco.

5.4.3.3. Diagnóstico del modelo

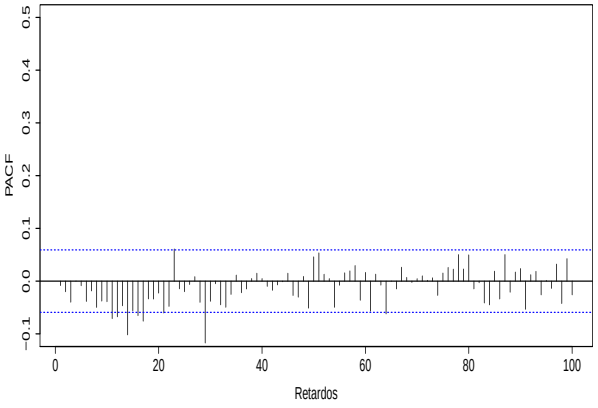
En la Figura 5.55a, 5.56a y 5.57a se observa que los residuos de los modelos ARMA(3,4), ARMA(4,7) y ARMA(2,2) presentan una media constante y una varianza con respecto a la media también constante. En la Figura 5.55b, 5.56b y 5.57b se aprecia el correlograma ACF de los residuos para los tres modelos, notándose en la Figura 5.56b y 5.57b la presencia de independencia entre las observaciones, ya que las autocorrelaciones se mantienen dentro de los intervalos de confianza. Sin embargo en la Figura 5.55b se sospecha de la presencia de independencia ya que no todas las autocorrelaciones se mantienen dentro de los intervalos de confianza. Finalmente en la Figura 5.55d, 5.56d y 5.57d se observa a simple vista que los residuos de los tres modelos se aproximan a una distribución normal.



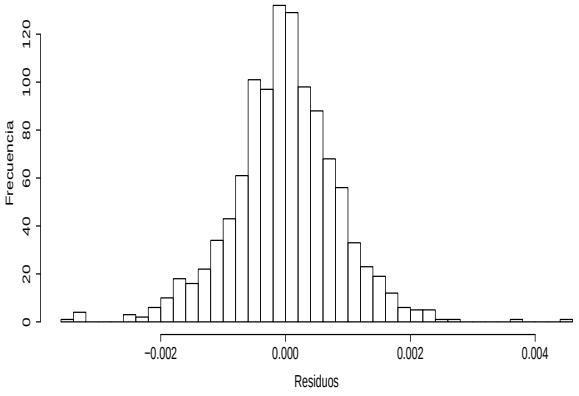
(a) Residuos



(b) ACF

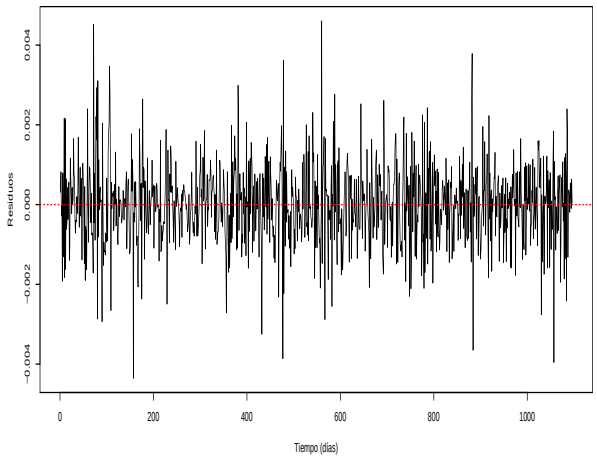


(c) PACF

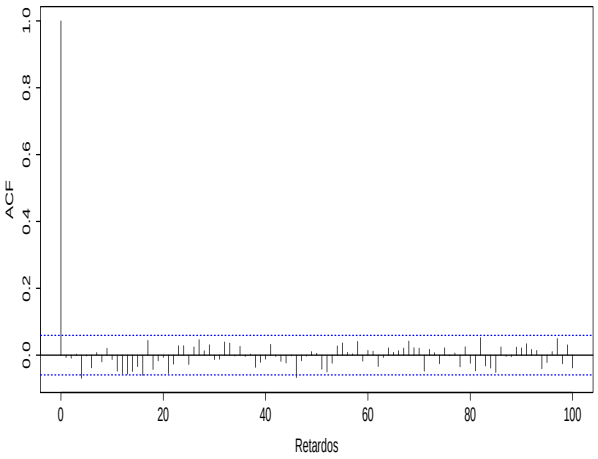


(d) Distribución de frecuencias

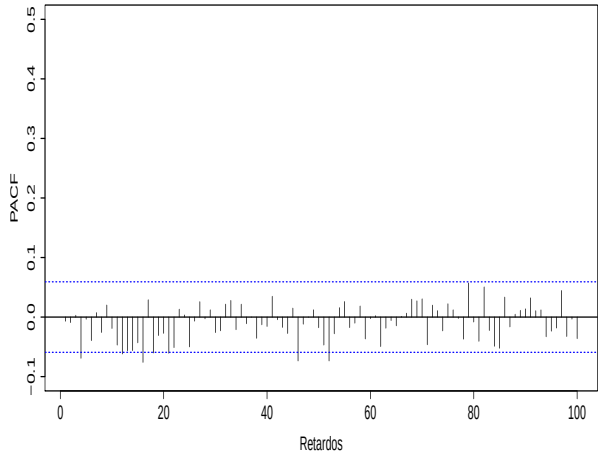
Figura 5.55. Gráfico de dispersión de los residuos del modelo ARMA(3,4), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Norte de Llafenco .



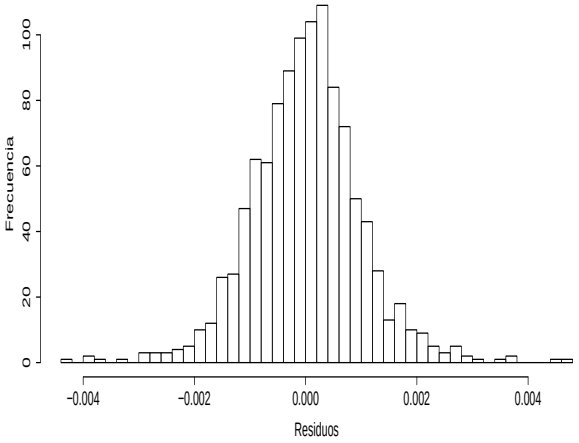
(a) Residuos



(b) ACF

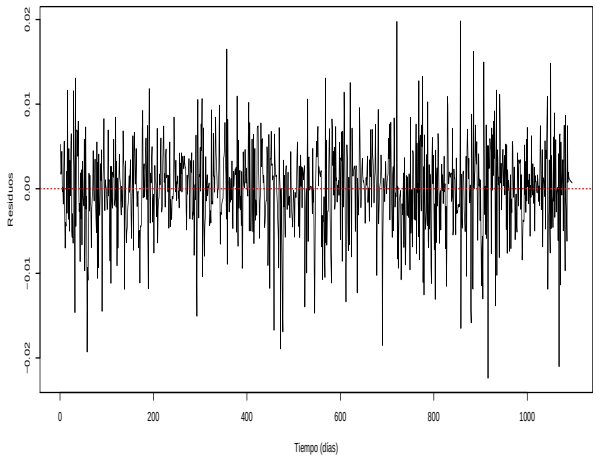


(c) PACF

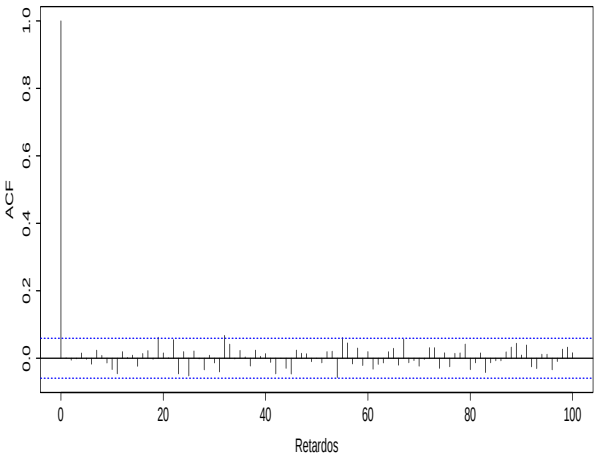


(d) Distribución de frecuencias

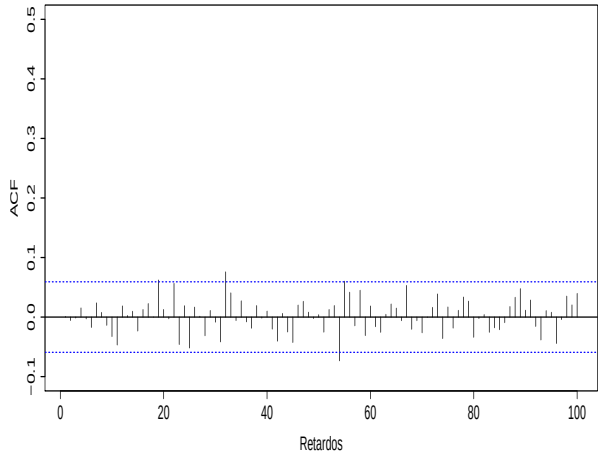
Figura 5.56. Gráfico de dispersión de los residuos del modelo ARMA(4,7), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Este de Llafenco.



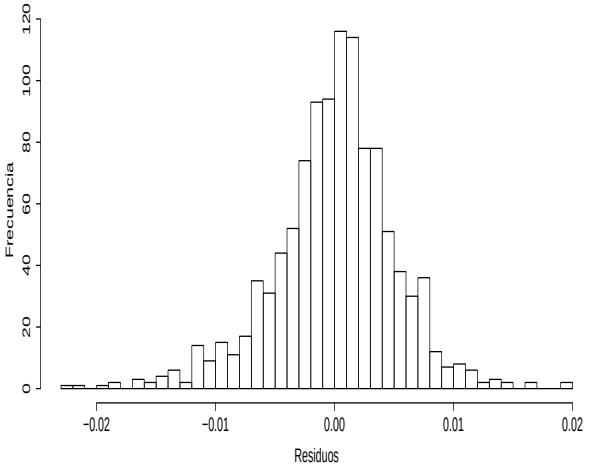
(a) Residuos



(b) ACF



(c) PACF



(d) Distribución de frecuencias

Figura 5.57. Gráfico de dispersión de los residuos del modelo ARMA(2,2), con sus respectivos correlogramas ACF, PACF y distribución de frecuencias, para la componente Altura de Llafenco.

A continuación se presentan en la Tabla los resultados obtenidos del test Ljung-Box para comprobar independencia entre los residuos de cada modelo.

Tabla 5.29. Test Ljung-Box sobre los residuos de cada modelo.

Componente	Residuos	Test Ljung-Box p-valor	Resultado
Norte	Modelo ARMA(3,4)	0.0485	No son independientes
Este	Modelo ARMA(4,7)	0.0662	Independientes
Altura	Modelo ARMA(2,2)	0.9308	Independientes

Finalmente los modelos para cada componente quedan expresados de la siguiente manera:

Componente Norte: modelo ARMA(3,4)

$$\varepsilon_t = 0.66321\varepsilon_{t-1} + 0.68197\varepsilon_{t-2} + -0.69066\varepsilon_{t-3} + \omega_t - (-1.015500)\omega_{t-1} - (-0.92668)\omega_{t-2} - (0.90299)\omega_{t-3} - (0.039618)\omega_{t-4}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

Componente Este: modelo ARMA(4,7)

$$\varepsilon_t = 0.38774\varepsilon_{t-1} + 0.52293\varepsilon_{t-2} + 0.46207\varepsilon_{t-3} + (-0.74493)\varepsilon_{t-4} + \omega_t - (-0.80890)\omega_{t-1}$$

$$-(-0.85043)\omega_{t-2} - (-0.43626)\omega_{t-3} - (1.193800)\omega_{t-4} - (0.14851)\omega_{t-5} - (-0.020077)\omega_{t-6} -(-0.22602)\omega_{t-7}\ ,\ \ \omega_t \sim RB(0, \sigma^2)$$

Componente Altura: modelo ARMA(2,2)

$$\varepsilon_t = 1.5567279\varepsilon_{t-1}+(-0.5639820)\varepsilon_{t-2}+\omega_t-(-1.0939278)\omega_{t-1}-(0.1853378)\omega_{t-2}, \omega_t \sim RB(0, \sigma^2)$$

5.4.3.4. Pronóstico

En esta sección se proceden a realizar dos tipos de pronósticos. El primer pronóstico es con respecto a la parte determinística (suavizamiento spline) y el segundo pronóstico es con respecto a la parte estocástica (modelos ARMA y ARIMA).

Pronóstico determinístico

A continuación se procede con el pronóstico obtenido de la estimación de la tendencia. En la Figura 5.58 y 5.59 se representan gráficamente las predicciones obtenidas.

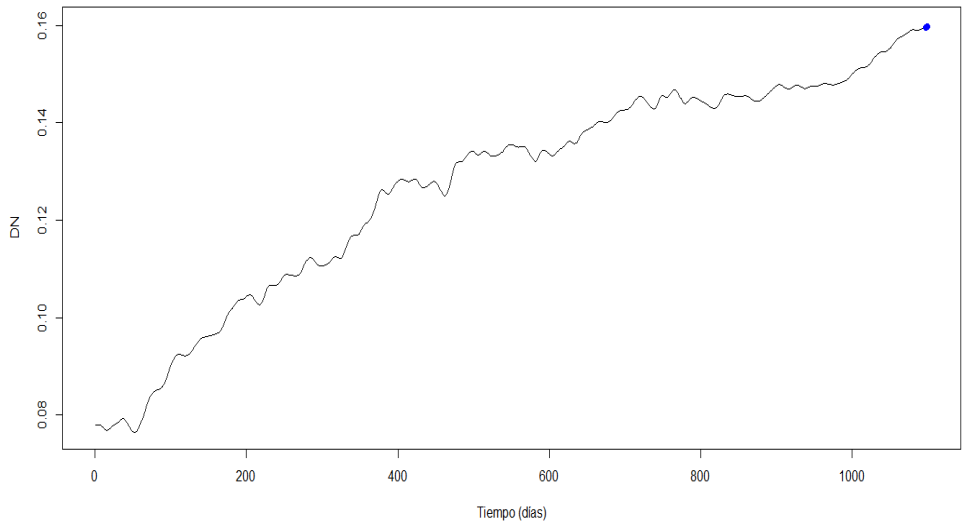


Figura 5.58. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Norte de Llafenco.

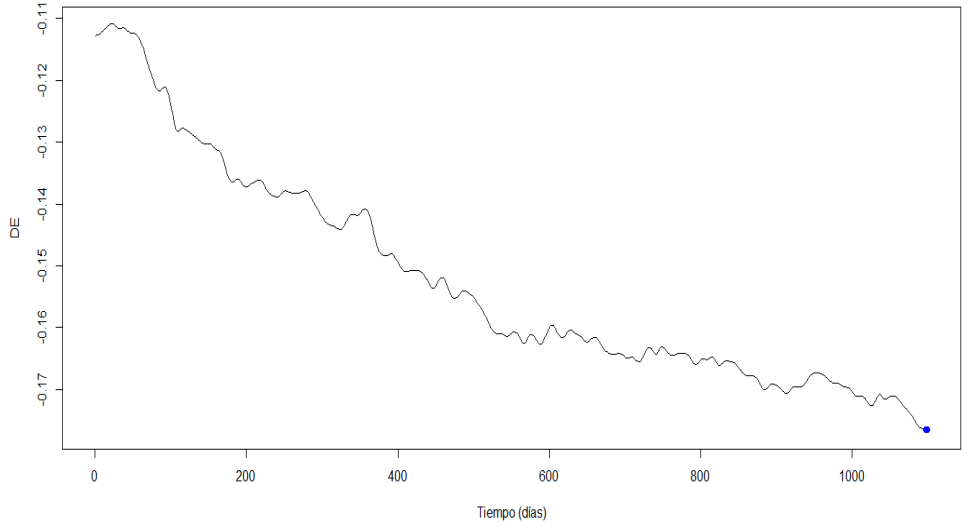


Figura 5.59. Gráfico de predicción para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Este de Llafenco.

A continuación se presentan en la Tabla 5.30 los resultados obtenidos de las predicciones.

Tabla 5.30. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto al suavizamiento spline sobre la componente Norte y Este de Llafenco.

Día	Mes	Año	Pronóstico componente Norte	Pronóstico componente Este
1	Enero	2013	0.15962	-0.17644
2	Enero	2013	0.15966	-0.17647
3	Enero	2013	0.15970	-0.17649
4	Enero	2013	0.15974	-0.17651

Pronóstico estocástico

Las predicciones obtenidas sobre la serie de tiempo componente Altura de Llafenco y los residuos del suavizamiento spline sobre la componente Norte y Este de Llafenco, se presenta gráficamente en la Figura 5.60, 5.61 y 5.62.

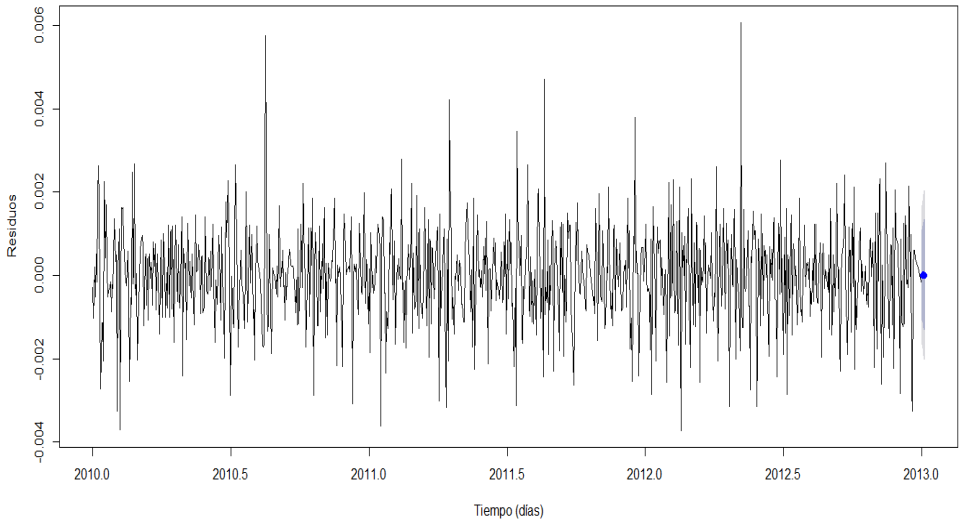


Figura 5.60. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(3,4) sobre los residuos de la componente Norte de Llafenco.

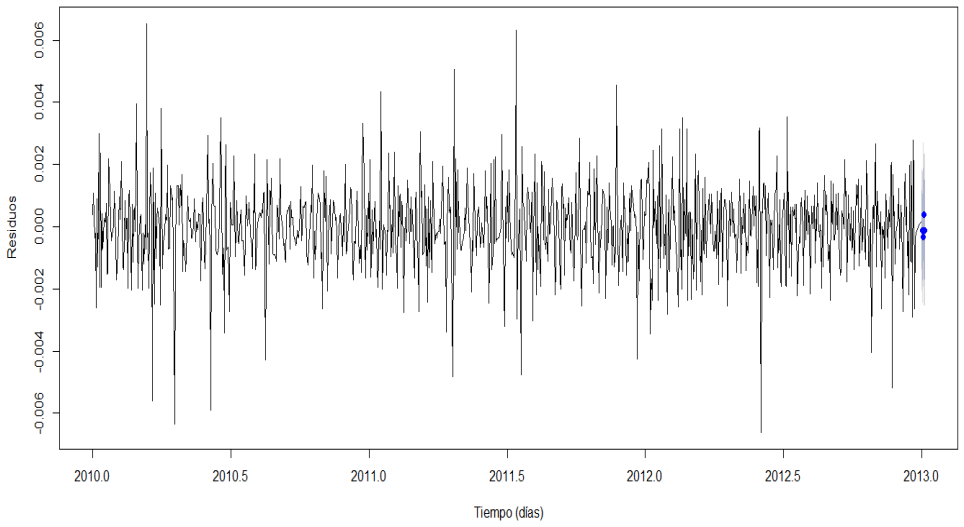


Figura 5.61. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(4,7) sobre los residuos de la componente Este de Llafenco.

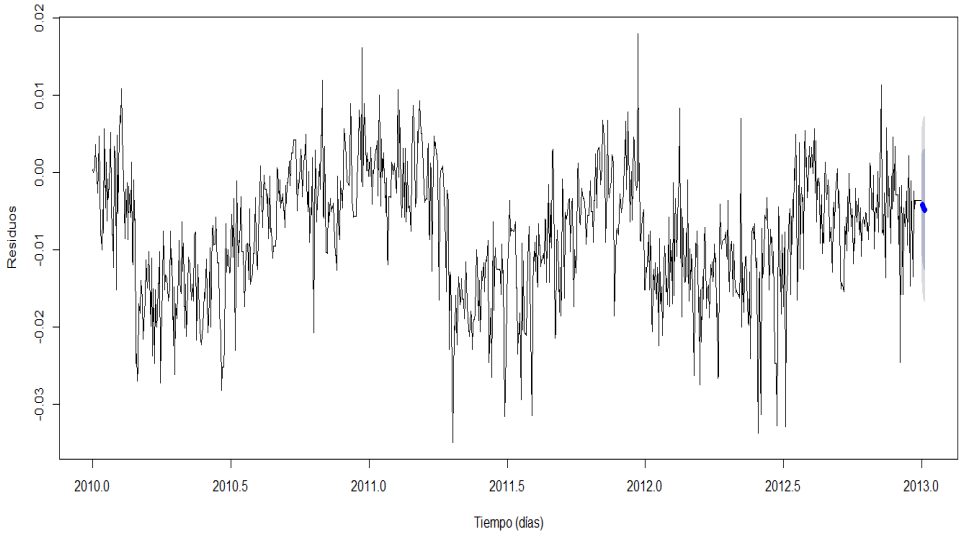


Figura 5.62. Gráfico de predicción para los primeros cuatro días de enero del 2013, utilizando el modelo ARMA(2,2) sobre la componente Altura de Llafenco.

En la Tabla 5.31 se presentan los resultados obtenidos de las predicciones sobre la componente Altura de Llafenco y los residuos de la componente Norte y Este de Llafenco.

Tabla 5.31. Predicciones obtenidas para los primeros cuatro días de enero del 2013, con respecto a la componente Altura de Llafenco y los residuos de la componente Norte y Este de Llafenco.

Día	Mes	Año	Pronóstico residuos DN Modelo ARMA(3,4)	Pronóstico residuos DE Modelo ARMA(4,7)	Pronóstico DA Modelo ARMA(2,2)
1	Enero	2013	-0.000018	-0.000111	-0.004217
2	Enero	2013	0.000001	-0.000341	-0.004486
3	Enero	2013	-0.000006	0.000385	-0.004661
4	Enero	2013	0.000008	-0.000106	-0.004782

Pronóstico general

En la Figura 5.63 y 5.64 se representan gráficamente las predicciones generales para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte y Este de Llafenco.

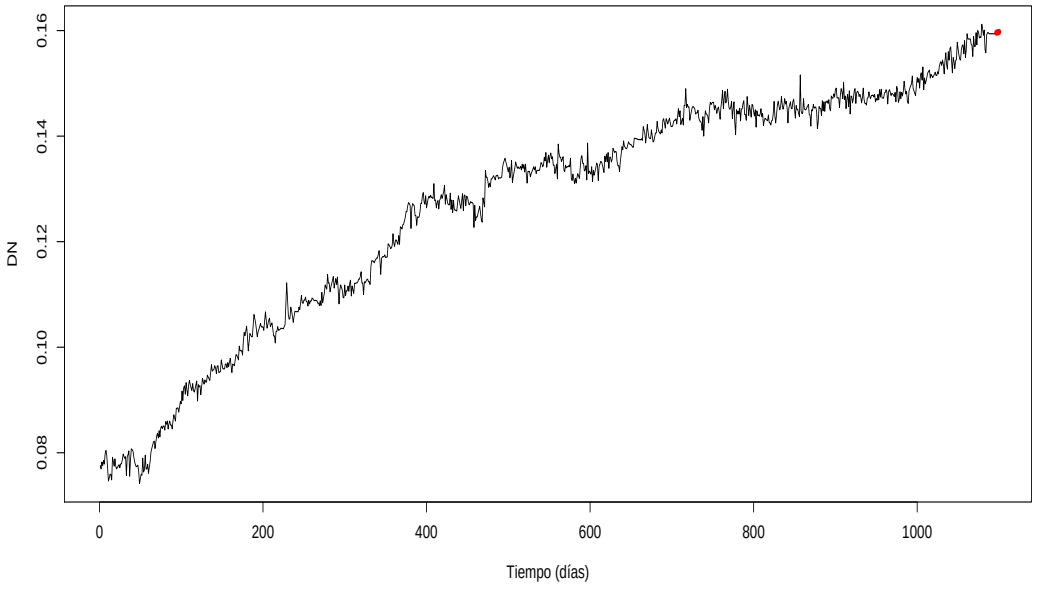


Figura 5.63. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte de Llafenco.

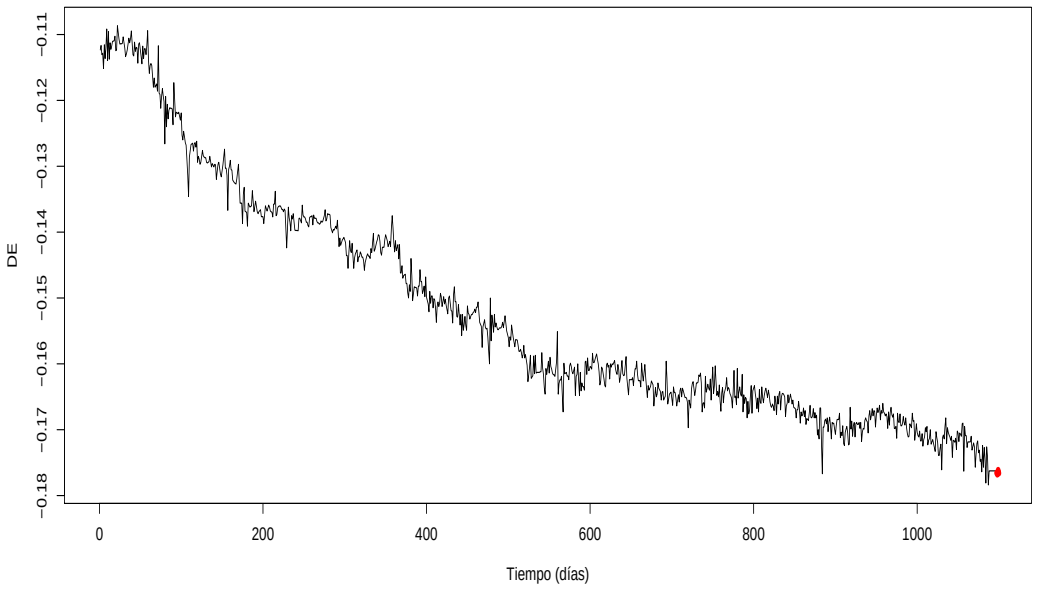


Figura 5.64. Gráfico de predicción general para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Este de Llafenco.

En la Tabla 5.32 se presentan las predicciones generales, para los primeros cuatro días de enero del 2013.

Tabla 5.32. Predicciones generales obtenidas para los primeros cuatro días de enero del 2013, con respecto a las series de tiempo componente Norte y Este de Llafenco.

Día	Mes	Año	Predicción parte determinística	Predicción parte estocástica	Predicción general
Componente Norte					
1	Enero	2013	0.15962	-0.000018	0.15960
2	Enero	2013	0.15966	0.000001	0.15966
3	Enero	2013	0.15970	-0.000006	0.15969
4	Enero	2013	0.15974	0.000008	0.15975
Componente Este					
1	Enero	2013	-0.17644	-0.000111	-0.17655
2	Enero	2013	-0.17647	-0.000341	-0.17681
3	Enero	2013	-0.17649	0.000385	-0.17611
4	Enero	2013	-0.17651	-0.000106	-0.17662

Predicciones finales y módulo del vector desplazamiento

Con las predicciones finales obtenidas para la componente Norte y Este de Llafenco en la Tabla 5.32 y componente Altura de Llafenco en la Tabla 5.31, se realiza el cálculo del módulo del vector desplazamiento. En la Tabla 5.33 se obtienen los resultados.

Tabla 5.33. Predicciones finales y módulo del vector desplazamiento obtenidos para los primeros cuatro días de enero del 2013, con respecto a la componente Norte, Este y Altura de Llafenco.

Día	Mes	Año	Predicción final DN	Predicción final DE	Predicción final DA	Módulo
1	Enero	2013	0.15960	-0.17655	-0.00422	0.23803
2	Enero	2013	0.15966	-0.17681	-0.00449	0.23827
3	Enero	2013	0.15969	-0.17611	-0.00466	0.23778
4	Enero	2013	0.15975	-0.17662	-0.00478	0.23820

De la Tabla 5.33 se puede deducir que el desplazamiento resultante es un vector de longitud 0.23803 cm para el primer día, 0.23827 cm para el segundo día, 0.23778 cm para el tercer día y 0.23820 cm para el cuarto día.

Capítulo 6

Conclusiones

En el presente trabajo de tesis se logra determinar, con respecto a las componentes Norte, Este y Altura de cada estación de monitoreo un desplazamiento diario ya sea creciente o decreciente en algunos casos, que se torna significativo. El terremoto del 27 de febrero del 2010 produjo movimiento y fricción entre las dos placas tectónicas ubicadas a lo largo del territorio chileno. Esto se puede comprobar observando los gráficos de las componentes en cada estación de monitoreo (ver Figura 5.1, 5.2, 5.3, 5.4, 5.5 y 5.6), donde además se logra aprecia claramente un mayor desplazamiento de la corteza terrestre en la componente Este de Concepción. Recordemos que el terremoto del 27 de febrero del 2013 afecto mayoritariamente la zona centro sur del país.

Como la mayoría de las series de tiempo en estudio presentaron una tendencia muy marcada se aplicaron métodos de estimación determinísticos (regresión lineal y suavizamiento spline), de las cuales se obtuvieron predicciones. A los residuos obtenidos de estas regresiones se aplicaron 2 tipos de modelos estocásticos, con el propósito de modelar de mejor manera los residuos y obtener de éstos buenas predicciones. Los modelos ajustados sobre los residuos fueron los modelos lineales ARMA y ARIMA, donde la utilización de cada uno de éstos dependió del comportamiento que seguían los residuos, es decir, los modelos ARMA se utilizaron cuando los residuos presentaban estacionariedad y los modelos ARIMA cuando los residuos no presentaban estacionariedad. Con la suma de las predicciones obtenidas por la parte determinística y estocástica en cada serie de tiempo en estudio, se obtuvieron predicciones generales sobre el desplazamiento de la corteza terrestre. Recordemos que también se obtuvieron predicciones sólo de la parte estocástica, como en el caso de la componente Altura en Iquique y Llafenco.

Finalmente se puede deducir con respecto a las predicciones obtenidas, que Con-

cepción presenta un mayor desplazamiento de la corteza terrestre respecto a las otras estaciones de monitoreo, durante los primeros cuatro días de enero del 2013, ya que al comparar las Tablas 5.12, 5.24 y 5.33 (respecto al módulo del vector desplazamiento) la que presenta valores mayores es la estación de monitoreo Concepción. En la presente tesis las observaciones son diarias pero sería mejor obtener datos que sean mensuales o anuales, para predecir en un futuro próximo donde podría haber un mayor desplazamiento de la corteza terrestre. Es importante mencionar que los modelos utilizados en la presente tesis podrían ser usados para futuros trabajos, en los cuales se utilicen datos ya sean mensuales o anuales.

Apéndice A

Estimación de Modelos: Iquique

A.1. Componente Norte

Tabla A.1. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre los residuos obtenidos de la regresión lineal, de la serie de tiempo componente Norte.

Modelo	AIC	BIC
ARIMA(0,1,1)	-11802.83	-11792.84
ARIMA(1,1,0)	-11739.33	-11729.33
ARIMA(1,1,1)	-11814.63	-11799.63
ARIMA(2,1,0)	-11748.32	-11733.33
ARIMA(2,1,2)	-11819.63	-11794.64
ARIMA(2,1,1)	-11821.48	-11801.49
ARIMA(1,1,2)	-11820.72	-11800.73
ARIMA(3,1,1)	-11819.69	-11794.70
ARIMA(1,1,3)	-11819.83	-11794.84
ARIMA(3,1,2)	-11817.84	-11787.85
ARIMA(2,1,3)	-11817.21	-11787.22
ARIMA(3,1,3)	-11817.36	-11782.37
ARIMA(4,1,3)	-11820.10	-11780.11
ARIMA(4,1,4)	-11818.00	-11773.02

A.2. Componente Este

Tabla A.2. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos de la regresión lineal de la serie de tiempo componente Este.

Modelo	AIC	BIC
ARMA(1,1)	-11649.69	-11629.69
ARMA(1,2)	-11652.93	-11627.93
ARMA(1,3)	-11655.23	-11625.23
ARMA(2,1)	-11654.18	-11629.18
ARMA(2,2)	-11656.25	-11626.25
ARMA(2,3)	-11654.44	-11619.44
ARMA(3,1)	-11656.13	-11626.14
ARMA(3,2)	-11654.47	-11619.47
ARMA(3,3)	-11652.44	-11612.44

A.3. Componente Altura

Tabla A.3. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre la serie de tiempo componente Altura.

Modelo	AIC	BIC
ARIMA(0,1,1)	-9293.898	-9283.901
ARIMA(1,1,0)	-9177.461	-9167.464
ARIMA(1,1,1)	-9315.193	-9300.198
ARIMA(2,1,0)	-9255.441	-9240.446
ARIMA(2,1,2)	-9317.453	-9292.460
ARIMA(2,1,1)	-9313.385	-9293.391
ARIMA(1,1,2)	-9313.502	-9293.508
ARIMA(3,1,1)	-9313.479	-9288.487
ARIMA(1,1,3)	-9313.302	-9288.309
ARIMA(3,1,2)	-9314.430	-9284.439
ARIMA(2,1,3)	-9317.206	-9287.215
ARIMA(3,1,3)	-9317.722	-9282.732
ARIMA(4,1,3)	-9319.571	-9279.583
ARIMA(4,1,4)	-9319.152	-9274.165

Apéndice B

Estimación de Modelos: Concepción

B.1. Componente Norte

Tabla B.1. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre los residuos obtenidos de la regresión lineal de la serie de tiempo componente Norte.

Modelo	AIC	BIC
ARIMA(0,1,1)	-11752.88	-11742.89
ARIMA(1,1,0)	-11708.13	-11698.14
ARIMA(1,1,1)	-11766.17	-11751.18
ARIMA(2,1,0)	-11743.47	-11728.47
ARIMA(2,1,2)	-11762.19	-11737.19
ARIMA(2,1,1)	-11764.20	-11744.21
ARIMA(1,1,2)	-11764.20	-11744.20
ARIMA(3,1,1)	-11762.21	-11737.22
ARIMA(1,1,3)	-11762.80	-11737.81
ARIMA(3,1,2)	-11760.72	-11730.73
ARIMA(2,1,3)	-11760.31	-11730.32
ARIMA(3,1,3)	-11758.62	-11723.63
ARIMA(4,1,3)	-11757.90	-11717.92
ARIMA(4,1,4)	-11761.39	-11716.40

B.2. Componente Este

En esta ocasión se aumenta el orden de los modelos ARMA(p,q), ya que existen complicaciones para lograr identificar el modelo que consiga que los residuos sean independientes.

Tabla B.2. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos del suavizamiento spline, de la serie de tiempo componente Este.

Modelo	AIC	BIC
ARMA(1,1)	-11127.61	-11107.61
ARMA(1,2)	-11140.70	-11115.70
ARMA(1,3)	-11370.67	-11340.68
ARMA(1,4)	-11376.81	-11341.82
ARMA(1,5)	-11395.16	-11355.17
ARMA(1,6)	-11393.17	-11348.18
ARMA(1,7)	-11392.37	-11342.38
ARMA(2,1)	-11135.76	-11110.76
ARMA(2,2)	-11846.58	-11816.58
ARMA(2,3)	-11854.76	-11819.77
ARMA(2,4)	-11367.49	-11327.49
ARMA(2,5)	11896.85	-11851.86
ARMA(2,6)	-11904.35	-11854.35
ARMA(2,7)	-11878.66	-11823.66
ARMA(3,1)	-11484.72	-11454.72
ARMA(3,2)	-11860.64	-11825.65
ARMA(3,3)	-11839.15	-11799.16
ARMA(3,4)	-11863.23	-11818.24
ARMA(3,5)	-11836.92	-11786.93
ARMA(3,6)	-11877.31	-11822.31
ARMA(3,7)	-11909.04	-11849.05
ARMA(4,1)	-11895.94	-11486.8
ARMA(4,2)	-11844.08	11804.08
ARMA(4,3)	-11841.12	-11796.13
ARMA(4,4)	-11849.71	-11799.72
ARMA(4,5)	-11886.49	-11831.49
ARMA(4,6)	-11895.94	-11835.94
ARMA(4,7)	-11891.08	-11826.09

B.3. Componente Altura

Tabla B.3. Resultados de los distintos modelos ARIMA(p,d,q), aplicados sobre los residuos obtenidos de la regresión lineal, de la serie de tiempo componente Altura.

Modelo	AIC	BIC
ARIMA(0,1,1)	-9117.960	-9107.963
ARIMA(1,1,0)	-9037.297	-9027.300
ARIMA(1,1,1)	-9132.865	-9117.870
ARIMA(2,1,0)	-9076.168	-9061.172
ARIMA(2,1,2)	-9133.211	-9108.219
ARIMA(2,1,1)	-9135.208	-9115.214
ARIMA(1,1,2)	-9134.940	-9114.946
ARIMA(3,1,1)	-9133.214	-9108.221
ARIMA(1,1,3)	-9133.199	-9108.207
ARIMA(3,1,2)	-9133.915	-9103.924
ARIMA(2,1,3)	-9133.921	-9103.930
ARIMA(3,1,3)	-9131.887	-9096.898
ARIMA(4,1,3)	-9129.855	-9089.867
ARIMA(4,1,4)	-9127.774	-9082.788

Apéndice C

Estimación de Modelos: Llafenco

C.1. Componente Norte

Tabla C.1. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos del suavizamiento spline, de la serie de tiempo componente Norte.

Modelo	AIC	BIC
ARMA(1,1)	-11821.32	-11801.32
ARMA(1,2)	-12066.01	-12041.01
ARMA(1,3)	-12084.99	-12054.99
ARMA(1,4)	-12093.17	-12058.18
ARMA(2,1)	-12106.39	-12081.4
ARMA(2,2)	-12306.69	-12276.69
ARMA(2,3)	-12331.29	-12296.29
ARMA(2,4)	-12090.72	-12050.72
ARMA(3,1)	-12136.52	-12106.52
ARMA(3,2)	-12289.05	-12254.06
ARMA(3,3)	-12340.69	-12300.7
ARMA(3,4)	-12342.29	-12297.3
ARMA(4,1)	-12160.18	-12125.19
ARMA(4,2)	-12143.06	-12103.07
ARMA(4,3)	-12340.68	-12295.68
ARMA(4,4)	-12338.91	-12288.92

C.2. Componente Este

En esta ocasión se aumenta el orden de los modelos ARMA(p,q), ya que existen complicaciones para lograr identificar el modelo que consiga que los residuos sean independientes.

Tabla C.2. Resultados de los distintos modelos ARMA(p,q), aplicados sobre los residuos obtenidos del suavizamiento spline, de la serie de tiempo componente Este.

Modelo	AIC	BIC
ARMA(1,1)	-11476.23	-11456.23
ARMA(1,2)	-11731.82	-11706.82
ARMA(1,3)	-11744.48	-11714.48
ARMA(1,4)	-11757.31	-11722.32
ARMA(1,5)	-11755.68	-11715.68
ARMA(1,6)	-11921.16	-11876.17
ARMA(1,7)	-11923.48	-11873.49
ARMA(2,1)	-11759.72	-11734.72
ARMA(2,2)	-11968.43	-11938.44
ARMA(2,3)	-11936.51	-11901.51
ARMA(2,4)	-11882.86	-11842.87
ARMA(2,5)	-11756.30	-11711.31
ARMA(2,6)	-11753.5	-11703.5
ARMA(2,7)	-11922.92	-11867.93
ARMA(3,1)	-11781.64	-11751.64
ARMA(3,2)	-11943.63	-11908.63
ARMA(3,3)	-11949.46	-11909.47
ARMA(3,4)	-11929.62	-11884.62
ARMA(3,5)	-11949.35	-11899.36
ARMA(3,6)	-11978.72	-11923.73
ARMA(3,7)	-11977.19	-11917.19
ARMA(4,1)	-11795.51	-11760.52
ARMA(4,2)	-11975.24	-11935.24
ARMA(4,3)	-11913.08	-11868.08
ARMA(4,4)	-11981.03	-11931.04
ARMA(4,5)	-11979.95	-11924.95
ARMA(4,6)	-11974.38	-11914.39
ARMA(4,7)	-12007.29	-11942.29

C.3. Componente Altura

Tabla C.3. Resultados de los distintos modelos ARMA(p,q), aplicados sobre la serie de tiempo componente Altura.

Modelo	AIC	BIC
ARMA(1,1)	-8395.846	-8375.848
ARMA(1,2)	-8411.377	-8386.380
ARMA(1,3)	-8417.122	-8387.125
ARMA(2,1)	-8419.776	-8394.779
ARMA(2,2)	-8422.074	-8392.078
ARMA(2,3)	-8420.185	-8385.190
ARMA(3,1)	-8421.631	-8391.634
ARMA(3,2)	-8421.631	-8385.182
ARMA(3,3)	-8418.06	-8378.065

Bibliografía

Barrientos, A. F., Olaya, J., y González, V. M. (2007). Un modelo spline para el pronóstico de la demanda de energía eléctrica. *Revista Colombiana de Estadística*, 30(2):187–202.

Bowerman, B. L. y O'Connell, R. T. (2007). *Pronósticos, series de tiempo y regresión: un enfoque aplicado*. Cengage Learning Editores. De México, 4 edition.

Bowman, A. W. y Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford.

Bustamante, A. M. (2013). Series de tiempo: Una aplicación a registros hidrométricos en una cuenca del estado de oaxaca.

Casquero, María José Presno y Menéndez, A. J. L. (2001). Tratamiento estadístico de series con cambios estructurales: un caso de estudio. *RAE: Revista Asturiana de Economía*, (22):123–141.

Chatfield, C. (2003). *The Analysis of Time Series, An Introduction*. Chapman and Hall/CRC, Blvd, Boca Raton, Florida 33431. United States of America, 6 edition. The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*, 74(368):829–836.

del Alcázar, J. A. y Quiróz, J. H. (1996). Modelos de series de tiempo para el pronóstico de precios de minerales. *Revista de Análisis Económico (UDAPE)*.

del Carmen García, M., Castellana, N., Rapelli, C., Koegel, L., y Catalano, M. (2014). Criterios de información y predictivos para la selección de un modelo lineal mixto. *SaberEs*, (6).

Gasser, T. y Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, pages 171–185.

Gonzáles, K., Froger, J.-L., Rivera Porras, M. A., y Audin, L. (2006). Deformación co-sísmica producida por el sismo mb= 5.4 del 01 de octubre de 2005 (carumas-moquegua), detectada por interferometría radar-insar.

Green, P. J. y Silverman, B. W. (2000). *Nonparametric regression and generalized linear models*. Chapman and Hall, New York.

Gujarati, D. N. (1997). Econometría. *Santa Fé de Bogotá, Colombia: McGraw-Hill*.

Härdle, W. (1992). *Applied nonparametric regression*. Econometric Society Monographs. Cambridge, UK: Cambridge University Press, 333 p.

Hildebrand, D. K. et al. (1998). *Estadística aplicada a la administración y a la economía*. Serie de probabilidad y estadística. Addison Wesley Longman., Boulevard de las Cataratas No. 3, Col Jardines del Pedregal, México, DF., 3 edition. Esta versión en español es la única autorizada con 4000 ejemplares.

Isern, G. y Cuesta, C. (2013). Eleccion del parametro de suavizado Óptimo en regresiones p-spline. un estudio por simulacion.

issuu (2012). *GPS*. Accedido en 09-03-2017 a `http://https://issuu.com/johnphillipbuena/docs/1.gps_background`.

Julio, J. M., Mera, S. J., y Hérault, A. R. (2002). La curva spot (cero cupón) estimation con splines cúbicos suavizados, usos y ejemplos. Technical report, Banco de la Republica de Colombia.

Li, J., Miyashita, K., Kato, T., y Miyazaki, S. (2000). Gps time series modeling by autoregressive moving average method: Application to the crustal deformation in central japan. *Earth, planets and space*, 52(3):155–162.

Madariaga, R. (1998). Sismicidad de chile. *Física de la Tierra*, (10):221.

Meza Santa Cruz, L. A. (2013). Regresión no paramétrica utilizando spline para la suavización de la estructura de la mortalidad en el perú.

Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory of Probability y Its Applications*, 9(3):497–500.

Parra, H., Maturana, R., Piña, V., y Báez, J. C. (2010). *Red Geodésica Nacional SIRGAS – Chile*. Accedido en 01-02-2017 a `http://www.sirgas.org/fileadmin/docs/Boletines/Bol15/34_Parra_et_al_Reporte_Chile.pdf`.

Pérez, J. G. G. (2014). *Pronóstico de demanda y precios de energía en un mercado eléctrico*. PhD thesis, Universidad Michoacana de San Nicolás de Hidalgo.

Priestley, M. y Chao, M. (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 385–392.

Reina, J. y Olaya, J. (2013). Ajuste de curvas mediante métodos no paramétricos para estudiar el comportamiento de contaminación del aire por material particulado pm10 (curve fitting nonparametric methods for studying behavior from air pollution pm10). *Revista EIA*, 9(18):19–31.

Rincón, L. (2012). Introducción a los procesos estocásticos. *Departamento de Matemáticas, Facultad de Ciencias UNAM*.

Ruegg, J. (1994). Deformación de la corteza terrestre y terremotos: aplicación al estudio del ciclo sísmico en el norte de chile. *Física de la Tierra*, (6):202.

Shumway, R. H. y Stoffer, D. S. (2010). *Time series analysis and its applications: with R examples*. Springer Science y Business Media.

Tarbuck, E. J., Lutgens, F. K., y Tasa, D. (2005). Ciencias de la tierra. *Una introduc*.

Tipler, P. A. y Mosca, G. (2005). *Física para la ciencia y la tecnología*, volume 1. Reverté, 5 edition.

Villavicencio, J. (2010). Introducción a series de tiempo. *Obtenido de Sitio web del Instituto de Estadísticas de Puerto Rico: http://www. estadisticas. gobierno. pr/iepr/LinkClick. aspx*.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 359–372.

Wegman, E. J. y Wright, I. W. (1983). Splines in statistics. *Journal of the American Statistical Association*, 78(382):351–365.