



**FACULTAD DE MATEMÁTICAS**  
PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE

*Consultoria Estadística III - EPG3383*

Proyecto de Magíster:

Predicción de temperaturas con modelos SARIMAX para  
prevenir el efecto de heladas en agricultura.

Autor:

Matías Gabriel Mayol Suárez

Profesora:

Maria Constanza Prado Stuardo

Fecha:

Segundo semestre 2020

# Índice

<b>I Resumen</b>	<b>1</b>
<b>II Introducción y Objetivos</b>	<b>2</b>
II.a Objetivos Generales . . . . .	3
II.b Objetivos Específicos . . . . .	3
<b>III Metodología</b>	<b>4</b>
III.a Antecedentes . . . . .	4
III.b Modelos SARIMAX . . . . .	4
III.c Validación y Selección del Modelos . . . . .	5
<b>IV Información por Estación</b>	<b>7</b>
<b>V Estaciones por Región</b>	<b>8</b>
<b>VI Las Bases de Datos</b>	<b>9</b>
VI.a Estaciones de Estudio . . . . .	9
VI.b Descarga y Descompresión de Información . . . . .	9
VI.c Construcción y Descripción de las Bases de Datos . . . . .	9
<b>VII Software</b>	<b>9</b>
<b>VIII Resultados</b>	<b>10</b>
VIII.a Chillán . . . . .	10
VIII.a.1 Modelo 1 . . . . .	10
VIII.a.2 Modelo 2 . . . . .	11
VIII.a.3 Modelo 3 . . . . .	13
VIII.a.4 Modelo 4 . . . . .	14
VIII.b Pichilemu . . . . .	15
VIII.b.1 Modelo 1 . . . . .	15
VIII.b.2 Modelo 2 . . . . .	17
VIII.b.3 Modelo 3 . . . . .	19
VIII.b.4 Modelo 4 . . . . .	20
VIII.c Manquehue . . . . .	21
VIII.c.1 Modelo 1 . . . . .	22
VIII.c.2 Modelo 2 . . . . .	23
VIII.c.3 Modelo 3 . . . . .	24
VIII.c.4 Modelo 4 . . . . .	26

<b>IX Validación</b>	<b>28</b>
IX.a Chillán . . . . .	28
IX.a.1 Estructura de Correlación . . . . .	28
IX.a.2 Normalidad . . . . .	28
IX.a.3 Homocedasticidad . . . . .	28
IX.a.4 AIC . . . . .	28
IX.a.5 Valor Absoluto del Error . . . . .	28
IX.b Pichilemu . . . . .	28
IX.b.1 Estructura de Correlación . . . . .	28
IX.b.2 Normalidad . . . . .	29
IX.b.3 Homocedasticidad . . . . .	29
IX.b.4 AIC . . . . .	29
IX.b.5 Valor Absoluto del Error . . . . .	29
IX.c Manquehue . . . . .	29
IX.c.1 Estructura de Correlación . . . . .	29
IX.c.2 Normalidad . . . . .	29
IX.c.3 Homocedasticidad . . . . .	30
IX.c.4 AIC . . . . .	30
IX.c.5 Valor Absoluto del Error . . . . .	30
<b>X Conclusión</b>	<b>31</b>
<b>XI Anexos</b>	<b>32</b>
XI.a Anexo I: Descarga . . . . .	32
XI.b Anexo II: Descompresión . . . . .	33
<b>Antecedentes Bibliográficos</b>	<b>34</b>

# **I. Resumen**

Con la intención de proteger cultivos agrícolas, este proyecto se realiza para buscar en qué condiciones es posible anticipar heladas. Para esto se construye diversos modelos que predicen la temperatura con una antelación de 40 minutos.

Las bases de datos utilizadas corresponden a estaciones meteorológicas de Pichilemu, Chillán y Manquehue, entre enero del año 2018 y diciembre del año 2019, construidas con información publicada por la Dirección Meteorológica de Chile.

Se utiliza modelos SARIMAX de la temperatura para hacer las predicciones, estos modelos son escogidos porque permiten modelar efectos estacionales e históricos. Para ajustar los modelos se utiliza el 70% de la información, el 30% restante se utiliza para validar y para analizar el valor absoluto de los errores de las predicciones.

Los modelos generan resultados favorables en lo que se refiere la magnitud del error, pero no han podido ser validados. Esto se debe a que, si bien en general el valor del error no tiene una magnitud muy alta en la mayoría de los casos, presenta comportamientos con fenomenologías que no se ha podido modelar completamente para este proyecto.

## II. Introducción y Objetivos

En este proyecto se pretende obtener un modelo predictivo de la temperatura que permita anticiparse lo suficiente a heladas para amortiguar el daño que sufren diversos cultivos por bajas temperaturas.

En [*Snyder, Melo-Abreu (2010)*] se explica lo siguiente:

La agrometeorología trata de las interacciones entre los factores meteorológicos e hidrológicos con la agricultura en el sentido más amplio, incluyendo la horticultura, la ganadería y la producción forestal. Su objetivo es estudiar y definir dichas interacciones, y luego aplicar el conocimiento de la atmósfera para el uso práctico en la agricultura. El principal papel de la agrometeorología moderna es asegurar que los investigadores, planificadores, y agricultores dispongan de datos, herramientas y del conocimiento necesario para enfrentarse a los problemas de la producción agrícola provocados por las variaciones meteorológicas y climáticas.

La palabra "helada" hace referencia a la producción de cristales de hielo en las superficies, pero es utilizada normalmente a fin de describir cuándo los cultivos y otras plantas experimentan daño por congelación. Las heladas se han definido en dos ámbitos, "advectiva" y "radiativa". Las primeras se asocian a incursiones a gran escala de aire frío con una atmósfera con viento, bien mezclada y una temperatura que a menudo está por debajo de cero. Mientras que las radiativas están asociadas con el enfriamiento debido a la pérdida de energía por el intercambio radiante durante las noches despejadas, en calma y con inversiones de temperatura (donde hay aire caliente a mayor altura que aire frío). Dentro de las protecciones para heladas radiativas están el uso de helicópteros o ventiladores, los cuales no son útiles para heladas advectivas, para estas se puede utilizar otros sistemas de protección, como lo es el riego.

Existen diversos métodos de protección frente a heladas, los cuales se dividen en dos, métodos pasivos y activos. Los primeros enfocados en la prevención son aplicados a lo largo del tiempo y se relacionan con técnicas biológicas o ecológicas. Mientras que los métodos activos son efectuados en un periodo corto de tiempo y requieren energía, trabajo o ambos; estos métodos incluyen la utilización de estufas, aspersores y máquinas de viento. Los daños provocados por las heladas pueden producirse en cualquier lugar en que las temperaturas descendan bajo 0 °C, y su gravedad depende de varios factores, desde el cultivo a la duración de la helada.

La predicción y el seguimiento de los valores de temperatura y otras variables meteorológicas ayuda entonces a los agricultores a decidir si necesitan utilizar métodos de protección y cuándo poner en marcha sus sistemas. Una predicción precisa de heladas y sus temperaturas mínimas puede entonces prevenir y reducir su daño asociado.

Para hacer predicciones meteorológicas se ha observado que existe diversidad de técnicas que es posible utilizar; en [*Snyder, Melo-Abreu (2010)*] se observan regresiones lineales, en [*Irigoyen (2011)*] se observan clasificadores neuronales y se observan modelos ARIMA en [*Aguado-Rodríguez et al. (2016)*]. Dado esto, en primera instancia se propone trabajar con modelos SARIMAX pues las bases de datos de las que se dispone corresponden a series de tiempo de variables meteorológicas con un periodo de

muestreo de un minuto durante dos años, lo que entrega bastante información. Un modelo SARIMAX permitiría considerar no solo efectos autorregresivos y del ruido, sino también de otras variables meteorológicas y la estacionalidad, y evitaría entrar a construir modelos de caja negra cuya fenomenología es más compleja de explicar, como ocurriría con una red neuronal.

## **II.a. Objetivos Generales**

En general este proyecto busca determinar en qué condiciones se puede anticipar el descenso de la temperatura por debajo cierto umbral, en especial de 0 °C, con la mayor antelación posible y qué variables es necesario observar para poder conocer tales condiciones.

## **II.b. Objetivos Específicos**

- Construir un modelo predictivo de la temperatura dentro de los próximos 30 a 60 minutos, especialmente que permita saber si se tendrá una temperatura por debajo de un valor determinado.
- Determinar qué variables son necesarias para el modelo.
- Determinar qué modelos son más adecuados para hacer estas predicciones.

### III. Metodología

#### III.a. Antecedentes

Para la construcción del modelo en primer lugar se buscó la información y las bases de datos que permitan hacer las simulaciones.

La Dirección General de Aeronáutica Civil disponibiliza al público tanto de las variables que analizan cómo los datos guardados por estaciones meteorológicas [9]. Se encuentra los datos de varias estaciones meteorológicas disponibles entre el 2018 y el 2019.

La información es descargada y descomprimida mediante programas construidos en bash (*"Bourne-again shell"*, interprete de lenguaje de comandos que ofrece ventajas de programación e interactividad de uso [3]). Posteriormente, en python, se arma una base de datos por estación y se realiza una visualización detallada del comportamiento de sus variables. En este punto, y durante la construcción de los modelos, se descarta un alto número de bases de estaciones por tener menos variables o por tener periodos incompletos.

Una vez seleccionadas las bases de datos a utilizar, se elaboran distintos modelos para predecir la temperatura y obtienen los resultados correspondientes.

Posteriormente se validan los modelos construidos, se revisa si son útiles y se concluye.

#### III.b. Modelos SARIMAX

Las bases de datos obtenidas contienen las variables expuestas en la sección IV, contra el tiempo, con observaciones tomadas cada 1 minuto. Las predicciones que necesitamos requieren a lo menos 30 minutos para ser de utilidad, para tener holgura y buscando no perjudicar la predicción, se decidió utilizar un lapso de 40 minutos.

Las bases de datos tienen múltiples variables observadas, sin embargo, por restricciones de tiempo dado el elevado número de combinaciones posibles, este proyecto se centra en predecir la temperatura con modelos que utilizan temperatura en instantes anteriores y comportamientos estacionales.

De este modo, la predicción de temperatura se obtiene como la suma ponderada de temperaturas anteriores, con intervalos de 40 minutos, más predictores que reflejan una estacionalidad basado en un ciclo climatológico establecido (por ejemplo, un día o una estación del año), mediante una componente con un retraso de una estación completa o una senoide cuyo ciclo es el mismo que el de la estacionalidad representada.

Para hacer las predicciones de temperatura, se plantea utilizar las variantes que sean necesarias de un modelo autorregresivo integrado de media móvil (ARIMA). En este contexto, al incorporar estacionalidad a estos modelos se los llama SARIMA y al incorporar otros regresores se los llama SARIMAX.

Como se describe en [Aguado-Rodríguez et al. (2016)]: Los modelos ARIMA sirven para predecir series de una sola variable, a partir de los valores de la variable en el pasado, tanto en corto como

en largo plazo, para variables discretas como continuas, aunque el tiempo debe ser equiespaciado y de intervalos discretos. Se requiere que la serie sea estacionaria, lo que parece restrictivo, pero la mayoría de las series se pueden transformar a estacionarias mediante diferenciaciones. Con  $z_t$  la serie de datos original y  $W_t$  la "diferencia" de la serie,  $W_t$  se calcula como:

$$W_t = z_t - z_{t-1} \quad (1)$$

Es posible diferenciar varias veces para obtener una serie estacionaria. Una vez obtenida se busca un buen modelo ARIMA y si este es adecuado, se realiza la predicción.

Para las componentes diferenciales  $d$  y  $D$  enteras no negativas, con  $B$  el operador de rezago y  $S$  la estacionalidad.  $X_t$  es un proceso  $SARIMA(p, d, q) \times (P, D, Q)_S$  si el proceso:

$$Y_t = (1 - B)^d (1 - B^S)^D X_t \quad (2)$$

es un proceso ARMA tal que:

$$\phi(B)\Phi(B^S)Y_t = \theta(B)\Theta(B^S)\varepsilon_t \quad (3)$$

$$\varepsilon_t \sim RB(0, \sigma^2) \quad (4)$$

$$\phi(Z) = 1 - \phi_1 Z - \dots - \phi_p Z^p \quad (5)$$

$$\Phi(Z) = 1 - \Phi_1 Z - \dots - \Phi_P Z^P \quad (6)$$

$$\theta(Z) = 1 - \theta_1 Z - \dots - \theta_q Z^q \quad (7)$$

$$\Theta(Z) = 1 - \Theta_1 Z - \dots - \Theta_Q Z^Q \quad (8)$$

Ahora, utilizando la descripción en [2], incorporando regresores al modelo SARIMA descrito, se obtiene una serie de tiempo estudiada  $X_{estt}$ , con un modelo SARIMAX de regresores  $X_{regt}$ , tal que:

$$X_{estt} = \beta X_{regt} + X_t \quad (9)$$

donde  $X_t$  es un proceso SARIMA.

En nuestro caso, la temperatura del instante  $t$  es representada por  $X_{estt}$  y la temperatura 40 minutos antes del instante  $t$  es  $X_{est(t-1)}$ :

### III.c. Validación y Selección del Modelos

La base de datos contiene la información climática asociada a una estación entre los años 2018 y 2019, ordenada de manera temporal, con muestras de los parámetros de la estación cada un minuto.

Para construir y validar los modelos se utiliza dos segmentos de las bases de datos, uno para ajustar y otro segmento para validar y probar el modelo, el 70 % y el 30 % respectivamente, evitando tomar muestras aleatorias para no alterar el orden de la serie de tiempo al hacer la predicción, como es explicado en [1].



Después de ajustar los modelos, al predecir la temperatura en el segmento de validación, se chequea que los residuos no tengan estructura de correlación utilizando el test de Ljung-Box, que tengan distribución normal utilizando el test de Jarque-Bera [6] y que sean homocedasticos mediante una prueba de heterocedasticidad de residuos estandarizados [8]. Por otro lado, se debe calcular el promedio del valor absoluto del error, también en el segmento de validación para conocer el error que se espera en la predicción. Si el promedio del valor absoluto del error es pequeño, existen pocos elementos en que el error sobrepase cierto umbral y los residuos son homocedasticos, se espera que el modelo sea útil al no tener un error de predicción muy alto.

El test de Ljung-Box tiene como hipótesis nula que los residuos distribuyen de manera independiente y calcula el estadístico:

$$Q = n(n+2) \sum_{k=1}^h \hat{\rho}_k^2 \quad (10)$$

donde  $n$  es el tamaño de la muestra,  $\hat{\rho}_k$  la autocorrelación de retraso  $k$ , y  $h$  el número de retardos probados. El nivel de significancia de este estadístico esta dado por su comparación con la distribución de chi-cuadrado con  $h$  grados de libertad.

Similarmente, el test de Jarque-Bera tiene como hipótesis nula que la asimetría y la curtosis de los residuos, son los de una distribución normal (0 y 3), y calcula el estadístico:

$$JB = \frac{n}{6} (S^2 + \frac{(k-3)^2}{4}) \quad (11)$$

donde  $n$  es el número de observaciones,  $S$  la asimetría y  $K$  la curtosis de la muestra. El nivel de significancia de este estadístico también este dado por su comparación con la distribución de chi-cuadrado, pero con dos grados de libertad.

La homocedasticidad se prueba mediante una comparación de la suma de errores al cuadrado de distintos segmentos de la muestra, asumiendo homocedasticidad como hipótesis nula [4]. El estadístico calculado es

$$H(h) = \frac{\sum_{t=T-h+1}^T \bar{v}_t^2}{\sum_{t=d+1}^{d+1+h} \bar{v}_t^2} \quad (12)$$

donde  $d$  corresponde al máximo entre el número de periodos perdidos esperados [5] y el número de observaciones donde no se evalúa la verosimilitud, y  $h$  es un tercio del periodo total muestreado  $T$ . El estadístico  $H(h)$  se puede comparar con una distribución  $F(h,h)$  o alternativamente a  $hH(h)$  con una chi-cuadrado con  $h$  grados de libertad.

Durante la validación se prueban modelos de distintas características buscando los mejores casos. Si se tienen varios modelos válidos y o útiles, se prevé utilizar el criterio de selección de Akaike para determinar que modelos tienen mejores características. En el criterio de selección de Akaike se seleccionan los modelos de menor AIC, cuya ecuación es:

$$AIC = 2k - \ln(L) \quad (13)$$

con  $k$  el número de parámetros del modelo estadístico y  $L$  el máximo valor de la función de verosimilitud para el modelo estimado.

## IV. Información por Estación

La información disponible en las diferentes estaciones meteorológicas es la siguiente:

Tema	Unidad	Descripción
Agua caída	mm	Precipitación acumulada de 24 horas, 6 horas y del minuto
Presión y humedad	%	Humedad relativa
Presión y humedad	hPa	Presión del sensor
Presión y humedad	hPa	Presión reducida al nivel de la estación, estándar y del mar
Radiación	$W/m^2$	Radiación Global Instantánea del minuto
Radiación	$Wh/m^2$	Radiación Global Acumulada diaria
Temperatura	$^{\circ}C$	Temperatura del aire seco
Temperatura	$^{\circ}C$	Temperatura del punto de rocío
Temperatura	$^{\circ}C$	Temperatura Máxima y Mínima 12 horas
Temperatura		Fecha y hora en UTC de la temperatura máxima y mínima
Viento	$^{\circ}$	Dirección promedio vectorial de 10 y 2 minutos
Viento	$^{\circ}$	Dirección viento instantáneo en grados (entre 0 y 360) del segundo '00'
Viento	kt	Intensidad del viento instantáneo en nudos (kt) del segundo '00'
Viento	kt	Intensidad promedio de 10 y 2 minutos

## V. Estaciones por Región

En primera instancia se seleccionó las estaciones siguientes:

Región	Código	Estación (Ubicación)
Región de La Araucanía	370067	Angol
Región de La Araucanía	380013	Manquehue, Temuco aeródromo
Región de La Araucanía	380018	Lonquimay
Región de La Araucanía	380029	La Araucanía
Región de La Araucanía	380032	Rucamanque
Región de La Araucanía	380033	Manquehue UFRO
Región de La Araucanía	380063	Victoria
Región de La Araucanía	390028	Villarrica
Región de Los Ríos	390006	Pichoy
Región de Los Ríos	390015	Isla Teja
Región de Los Ríos	390029	Panguipulli
Región de Los Ríos	390043	Corral ESSAL
Región de Ñuble	360011	Chillán aeródromo
Región de Ñuble	360019	Concepción
Región de Ñuble	360045	Chillán Quinchamalí
Región de Ñuble	360046	Chillán Mayulermo
Región de Ñuble	370036	Los Ángeles
Región del Libertador General Bernardo O'Higgins	340045	Rancagua
Región del Libertador General Bernardo O'Higgins	340066	Ñlahue
Región del Libertador General Bernardo O'Higgins	340093	Portezuelo
Región del Libertador General Bernardo O'Higgins	340115	Pichilemu
Región del Maule	340031	General Freire
Región del Maule	350028	Panguilemo
Región del Maule	360047	Retiro

## VI. Las Bases de Datos

### VI.a. Estaciones de Estudio

Las estaciones que tienen menos información incompleta o interrupciones, y en consecuencia con las que se continuara trabajando, son las siguientes:

- 340115.- Pichilemu.
- 360046.- Chillán.
- 380033.- Maquehue.

### VI.b. Descarga y Descompresión de Información

La información en las estaciones se encuentra altamente distribuida y es comprimida al ser descargada. Es por esto que se construyen dos algoritmos en lenguaje bash para descargarla y descomprimirla, estos se encuentran en los anexos siguientes: XI.a y XI.b.

### VI.c. Construcción y Descripción de las Bases de Datos

Para construir las bases de datos, se leen los archivos descomprimidos y se agrupan las variables, construyendo una base de datos por estación. Para describir las distintas bases se observa el comportamiento de sus variables en el tiempo, sus correlaciones y sus valores agrupados, entre otras características.

Esto se desarrolla mediante algoritmos escritos en python, los que se pueden observar junto a sus resultados correspondiente al descargar el archivo html en el siguiente [link](#).

## VII. Software

Como es posible observar, se trabaja en bash para recolectar la información y en python para construir las bases de datos y describir las distintas variables observadas. También se construye los modelos y obtienen las predicciones en python, en este contexto, se utiliza la librería *statsmodels* para construir los modelos SARIMAX de la temperatura.

## VIII. Resultados

Para cada una de las tres estaciones se construyen distintos modelos SARIMAX, partiendo de distintos modelos iniciales. A estos modelos iniciales se les incorpora o quita variables según las características de sus residuos: densidad espectral, autocorrelaciones, autocorrelaciones parciales y significancia, además de la convergencia del modelo.

Dentro de las variables consideradas en cada modelo para representar la temperatura, no solo se utilizan variables endógenas, correspondientes a temperaturas en instantes anteriores, sino que también se utiliza variables exógenas, correspondientes a sinusoides para representar estacionalidades del modelo. Esto se debe a que se busca representar la estacionalidad de la temperatura de manera que no esté solo sujeta a temperaturas de un periodo anterior, sino que también de la situación estacional climática correspondiente.

A continuación, se describen los distintos modelos obtenidos.

### VIII.a. Chillán

Los desarrollos para la obtención de los modelos y sus resultados se pueden observar en este [link](#).

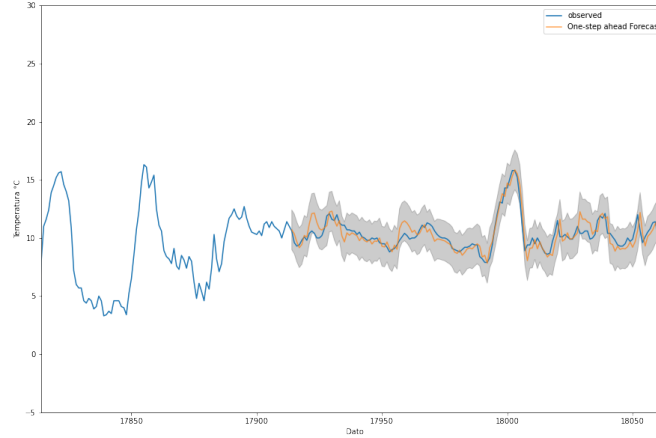
#### VIII.a.1. Modelo 1

En este caso, se parte desde un modelo de temperatura en que los únicos predictores son sinusoides con estacionalidad diaria y anual basada en los meses.

El Modelo 1 considera una componente diferencial con un retraso de un intervalo, este es un factor diferenciador respecto del comienzo del Modelo 2. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parámetro	Valor
Intercepto:	$8.50 \cdot 10^{-6}$
S:	12 (horas) o 18 intervalos
D:	0
d:	1
$\beta_1$ :	-3.72
$\beta_2$ :	1.50
$\phi_1$ :	0.13
$\Phi_1$ :	0
$\Phi_2$ :	0.98
$\Theta_1$ :	-0.01
$\Theta_2$ :	-0.89
$\sigma^2$ :	0.76

Figura 1: Resultados Modelo VIII.a.1



$$X_{regt} = \left[ \sin\left(\frac{hora * 2 * \pi}{24} + \frac{Minuto * 2 * \pi}{60}\right); \cos\left(\frac{hora * 2 * \pi}{24} + \frac{Minuto * 2 * \pi}{60}\right) \right] \quad (14)$$

Las sinusoides de estacionalidades anuales resultaron no significativas.

Como se mencionó previamente, el ajuste de este modelo se realiza con el 70 % de la base de datos, la validación del modelo se realiza de manera posterior, con el 30 % final de la base.

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	370.62
P-value(Q):	0.00
Heteroskedasticity (H):	0.94
P-value(H) (two-sided):	0.09
Jarque-Bera (JB):	7740.54
P-value(JB):	0.00
Skew:	0.24
Kurtosis:	7.91

Por otro lado, el valor del AIC de este modelo durante la validación es 19150.63.

La figura VIII.a.1 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.60. El 17.16 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 3.17 mayor 2°C y un 0.74 % mayor a 3 °C.

## VIII.a.2. Modelo 2

Este modelo parte de las mismas condiciones que el modelo anterior, pero no considera una componente diferencial.

Las sinusoides de estacionalidades anuales también resultaron no significativas. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	0.01
S:	24 (horas) o 36 intervalos
D:	0
d:	0
$\beta_1$ :	-4.07
$\beta_2$ :	1.57
$\phi_1$ :	0.95
$\Phi_1$ :	0.99
$\theta_1$ :	0.15
$\theta_2$ :	0.06
$\Theta_1$ :	-0.88
$\sigma^2$ :	0.75

$$X_{regt} = \left[ \sin\left(\frac{\text{hora} * 2 * \pi}{24} + \frac{\text{Minuto} * 2 * \pi}{60}\right); \cos\left(\frac{\text{hora} * 2 * \pi}{24} + \frac{\text{Minuto} * 2 * \pi}{60}\right) \right] \quad (15)$$

Nuevamente, como se mencionó previamente, el ajuste de este modelo se realiza con el 70 % de la base de datos, la validación del modelo se realiza de manera posterior, con el 30 % final de la base.

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

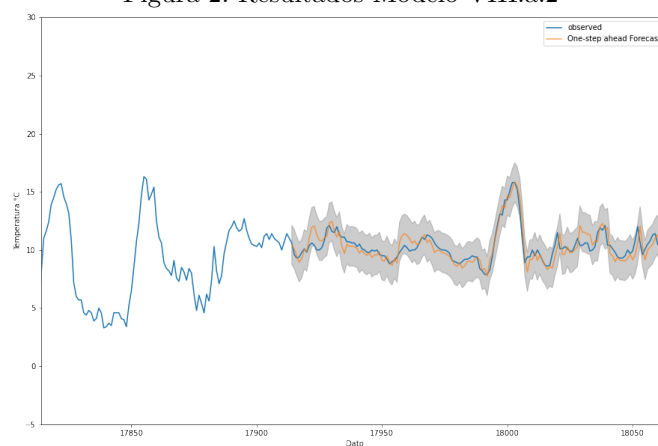
Test	Valor
Ljung-Box (Q):	334.82
P-value(Q):	0.00
Heteroskedasticity (H):	0.94
P-value(H) (two-sided):	0.13
Jarque-Bera (JB):	7507.07
P-value(JB):	0.00
Skew:	0.27
Kurtosis:	7.81

El valor del AIC de este modelo durante la validación es 19060.15.

La figura VIII.a.2 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.60. El 16.82 % de las predicciones tienen un valor absoluto del error mayor a 1 °C, un 3.02 mayor 2 °C y un 0.64 % mayor a 3 °C.

Figura 2: Resultados Modelo VIII.a.2



### VIII.a.3. Modelo 3

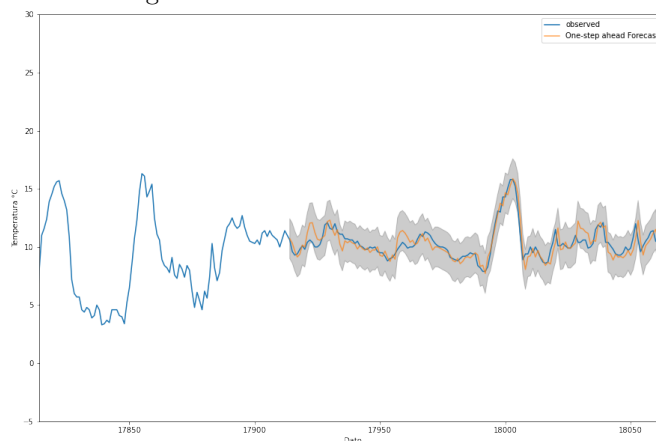
Este modelo se calcula retirando las componentes sinusoidales. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	$5.31 \cdot 10^{-6}$
S:	12 (horas) o 18 intervalos
$\beta$ :	0
d:	1
D:	0
$\phi_1$ :	0.35
$\Phi_1$ :	-0.01
$\Phi_2$ :	0.99
$\theta_1$ :	-0.22
$\Theta_1$ :	0
$\Theta_2$ :	-0.89
$\sigma^2$ :	0.77

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.



Figura 3: Resultados Modelo VIII.a.3



Test	Valor
Ljung-Box (Q):	1180.10
P-value(Q):	0.00
Heteroskedasticity (H):	1.06
P-value(H) (two-sided):	0.12
Jarque-Bera (JB):	9406.32
P-value(JB):	0.00
Skew:	0.17
Kurtosis:	8.43

El valor del AIC de este modelo durante la validación es 19623.15.

La figura VIII.a.3 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.61. El 18.25 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 3.62 mayor 2°C y un 0.85 % mayor a 3 °C.

#### VIII.a.4. Modelo 4

Este modelo se obtiene a partir de un modelo de componentes AR para todos los intervalos posibles hasta alcanzar un día, más una componente MA que es estacional de duración de un día. A este modelo se retiran componentes no significativos y de alto orden, para posteriormente seguir un procedimiento equivalente a los anteriores para llegar al modelo elegido.

A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	$9.39 \cdot 10^{-6}$
S:	24 (horas) o 36 intervalos
$\beta$ :	0
d:	1
D:	1
$\phi_1$ :	0.12
$\phi_2$ :	0.02
$\phi_3$ :	0.01
$\Phi_1$ :	0.05
$\Theta_1$ :	-0.91
$\sigma^2$ :	0.77

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	1229.68
P-value(Q):	0.00
Heteroskedasticity (H):	1.07
P-value(H) (two-sided):	0.07
Jarque-Bera (JB):	8892.17
P-value(JB):	0.00
Skew:	0.15
Kurtosis:	8.29

El valor del AIC de este modelo durante la validación es 19626.25.

La figura VIII.a.4 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.62. El 18.50 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 3.66 mayor 2°C y un 0.83 % mayor a 3 °C.

## VIII.b. Pichilemu

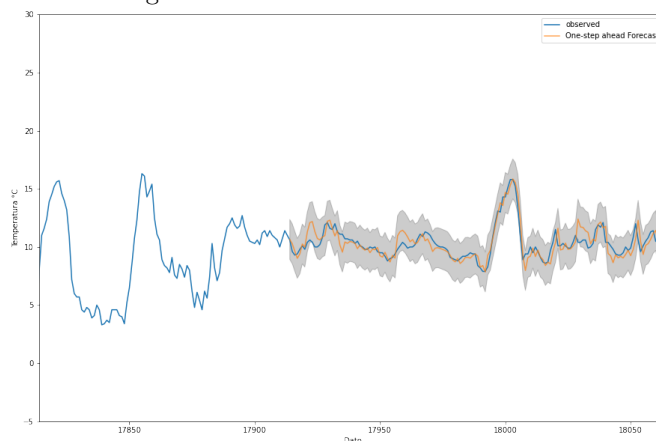
Los desarrollos para la obtención de los modelos y sus resultados se pueden observar en este [link](#).

### VIII.b.1. Modelo 1

En este caso, se parte desde un modelo de temperatura en que los únicos predictores son sinusoides con estacionalidad diaria y anual basada en los meses.

El Modelo 1 considera una componente diferencial con un retraso de un intervalo, este es un factor diferenciador respecto del comienzo del Modelo 2. A continuación, se parametriza las ecuaciones 2 a 9

Figura 4: Resultados Modelo VIII.a.4



con que se describe el modelo resultante.

Parámetro	Valor
Intercepto:	$1.25 \cdot 10^{-7}$
S:	24 (horas) o 36 intervalos
D:	0
d:	1
$\beta_1$ :	-1.64
$\beta_2$ :	0.71
$\phi_1$ :	0.89
$\Phi_1$ :	1.00
$\theta_1$ :	-0.95
$\Theta_1$ :	-0.95
$\sigma^2$ :	0.31

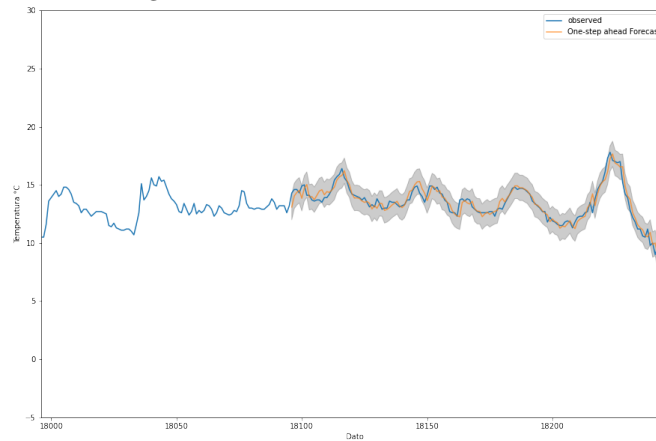
$$X_{regt} = \left[ \sin\left(\frac{\text{hora} * 2 * \pi}{24} + \frac{\frac{\text{Minuto} * 2 * \pi}{60}}{24}\right); \cos\left(\frac{\text{hora} * 2 * \pi}{24} + \frac{\frac{\text{Minuto} * 2 * \pi}{60}}{24}\right) \right] \quad (16)$$

Las sinusoides de estacionalidades anuales resultaron no significativas.

Como se mencionó previamente, el ajuste de este modelo se realiza con el 70 % de la base de datos, la validación del modelo se realiza de manera posterior, con el 30 % final de la base.

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Figura 5: Resultados Modelo VIII.b.1



Test	Valor
Ljung-Box (Q):	254.90
P-value(Q):	0.00
Heteroskedasticity (H):	0.81
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	10631.87
P-value(JB):	0.00
Skew:	0.17
Kurtosis:	8.73

Por otro lado, el valor del AIC de este modelo durante la validación es 12899.93.

La figura VIII.b.1 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.39. El 7.18 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 0.83 mayor 2°C y un 0.14 % mayor a 3 °C.

### VIII.b.2. Modelo 2

Este modelo parte de las mismas condiciones que el modelo anterior, pero no considera una componente diferencial.

De las sinusoides de estacionalidades anuales solo una resulta no significativa. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	1.09
S:	12 (horas) o 18 intervalos
D:	0
d:	0
$\beta_1$ :	-1.54
$\beta_2$ :	0.4195
$\beta_3$ :	1.8568
$\phi_1$ :	0.92
$\theta_1$ :	0.04
$\theta_2$ :	0.08
$\Theta_1$ :	0.02
$\Theta_2$ :	0.13
$\sigma^2$ :	0.33

$$X_{regt} = \left[ \sin\left(\frac{hora * 2 * \pi}{24} + \frac{Minuto * 2 * \pi}{60}\right); \cos\left(\frac{hora * 2 * \pi}{24} + \frac{Minuto * 2 * \pi}{60}\right); \cos\left(\frac{mes * 2 * \pi}{12}\right) \right] \quad (17)$$

Nuevamente, como se mencionó previamente, el ajuste de este modelo se realiza con el 70 % de la base de datos, la validación del modelo se realiza de manera posterior, con el 30 % final de la base.

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

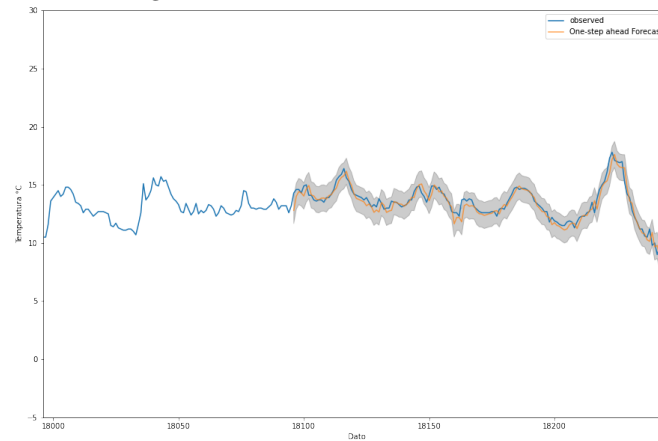
Test	Valor
Ljung-Box (Q):	473.90
P-value(Q):	0.00
Heteroskedasticity (H):	0.83
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	9954.07
P-value(JB):	0.00
Skew:	0.24
Kurtosis:	8.53

El valor del AIC de este modelo durante la validación es 13215.13.

La figura VIII.b.2 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.39. El 7.47 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 0.93 % mayor 2°C y un 0.13 % mayor a 3 °C.

Figura 6: Resultados Modelo VIII.b.2



### VIII.b.3. Modelo 3

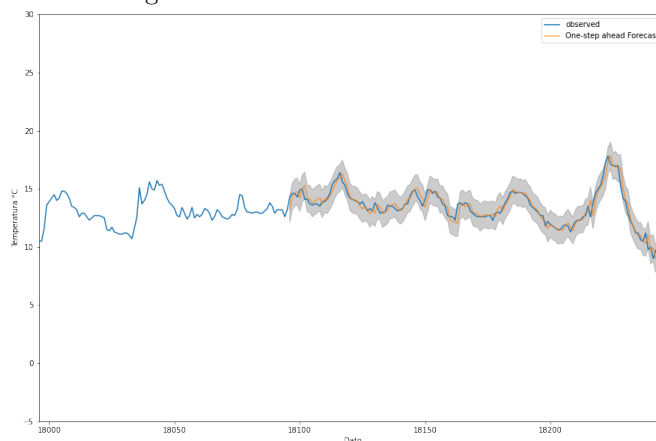
Este modelo se calcula retirando las componentes sinusoidales. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	$-3.62 \cdot 10^{-5}$
S:	12 (horas) o 18 intervalos
$\beta$ :	0
d:	1
D:	0
$\phi_1$ :	0.51
$\Phi_1$ :	-1.00
$\theta_1$ :	-0.46
$\Theta_1$ :	0.96
$\sigma^2$ :	0.35

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	926.86
P-value(Q):	0.00
Heteroskedasticity (H):	0.91
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	10190.79
P-value(JB):	0.00
Skew:	0.23
Kurtosis:	8.60

Figura 7: Resultados Modelo VIII.b.3



El valor del AIC de este modelo durante la validación es 13697.43.

La figura VIII.b.3 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.40. El 8.46 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 1.02 % mayor a 2°C y un 0.18 % mayor a 3 °C.

#### VIII.b.4. Modelo 4

Este modelo se obtiene a partir de un modelo de componentes AR para todos los intervalos posibles hasta alcanzar un día, más una componente MA que es estacional de duración de un día. A este modelo se retiran componentes no significativos y de alto orden, para posteriormente seguir un procedimiento equivalente a los anteriores para llegar al modelo elegido.

A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	0.51
S:	12 (horas) o 18 intervalos
$\beta$ :	0
d:	0
D:	0
$\phi_1$ :	1.04
$\phi_2$ :	0.04
$\phi_3$ :	-0.02
$\phi_4$ :	-0.08
$\phi_5$ :	0.02
$\phi_6$ :	-0.05
$\Phi_1$ :	-0.03
$\Phi_2$ :	0.21
$\sigma^2$ :	0.37

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	677.36
P-value(Q):	0.00
Heteroskedasticity (H):	0.81
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	11429.58
P-value(JB):	0.00
Skew:	0.26
Kurtosis:	8.94

El valor del AIC de este modelo durante la validación es 13879.25.

La figura VIII.b.4 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

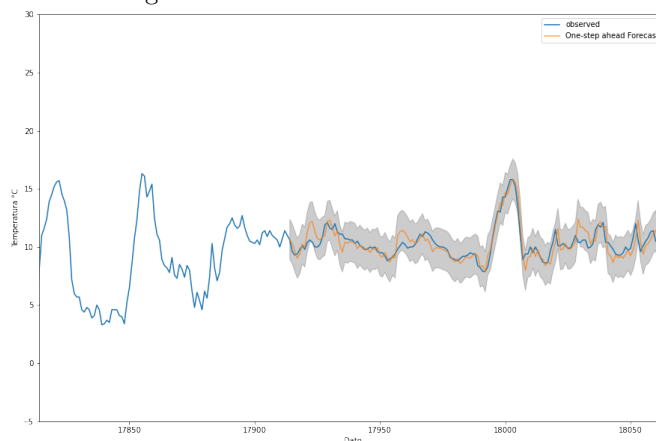
El valor promedio del valor absoluto del error en este modelo es 0.41. El 8.74 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 1.11 mayor 2°C y un 0.17 % mayor a 3 °C.

### VIII.c. Manquehue

Los desarrollos para la obtención de los modelos y sus resultados se pueden observar en este [link](#).



Figura 8: Resultados Modelo VIII.b.4



### VIII.c.1. Modelo 1

En este caso, se parte desde un modelo de temperatura en que los únicos predictores son sinusoides con estacionalidad diaria y anual basada en los meses.

El Modelo 1 considera una componente diferencial con un retraso de un intervalo, este es un factor diferenciador respecto del comienzo del Modelo 2. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parámetro	Valor
Intercepto:	$-3.68 \cdot 10^{-6}$
S:	24 (horas) o 36 intervalos
D:	0
d:	1
$\beta_1$ :	-3.31
$\beta_2$ :	1.72
$\phi_1$ :	0.61
$\Phi_1$ :	0.99
$\theta_1$ :	-0.30
$\Theta_1$ :	-0.92
$\sigma^2$ :	0.45

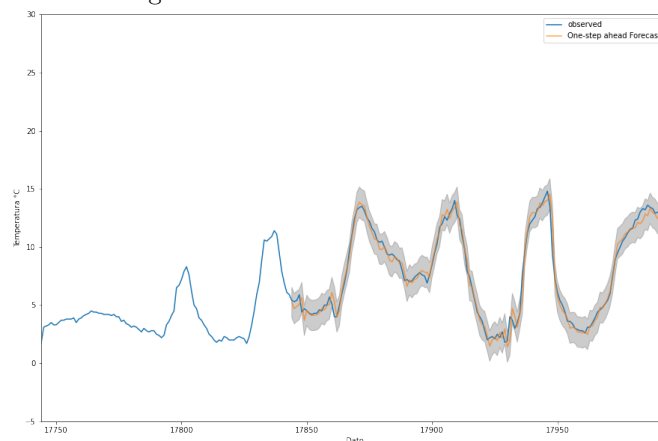
$$X_{regt} = \left[ \sin\left(\frac{\text{hora} * 2 * \pi}{24} + \frac{\frac{\text{Minuto} * 2 * \pi}{60}}{24}\right); \cos\left(\frac{\text{hora} * 2 * \pi}{24} + \frac{\frac{\text{Minuto} * 2 * \pi}{60}}{24}\right) \right] \quad (18)$$

Las sinusoides de estacionalidades anuales resultaron no significativas.

Como se mencionó previamente, el ajuste de este modelo se realiza con el 70 % de la base de datos, la validación del modelo se realiza de manera posterior, con el 30 % final de la base.

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Figura 9: Resultados Modelo VIII.c.1



Test	Valor
Ljung-Box (Q):	272.98
P-value(Q):	0.00
Heteroskedasticity (H):	1.17
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	10630.81
P-value(JB):	0.00
Skew:	0.06
Kurtosis:	8.77

Por otro lado, el valor del AIC de este modelo durante la validación es 15901.31.

La figura VIII.c.1 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.48. El 10.70 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 1.73 mayor 2°C y un 0.33 % mayor a 3 °C.

### VIII.c.2. Modelo 2

Este modelo parte de las mismas condiciones que el modelo anterior, pero no considera una componente diferencial.

Las sinusoides de estacionalidades anuales también resultaron no significativas. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	0.21
S:	24 (horas) o 36 intervalos
D:	0
d:	0
$\beta_1$ :	-3.34
$\beta_2$ :	1.42
$\phi_1$ :	1.66
$\phi_2$ :	-0.67
$\theta_1$ :	-0.30
$\Theta_1$ :	0.16
$\sigma^2$ :	0.48

$$X_{regt} = \left[ \sin\left(\frac{hora * 2 * \pi}{24} + \frac{Minuto * 2 * \pi}{60}\right); \cos\left(\frac{hora * 2 * \pi}{24} + \frac{Minuto * 2 * \pi}{60}\right) \right] \quad (19)$$

Nuevamente, como se mencionó previamente, el ajuste de este modelo se realiza con el 70 % de la base de datos, la validación del modelo se realiza de manera posterior, con el 30 % final de la base.

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	398.74
P-value(Q):	0.00
Heteroskedasticity (H):	1.11
P-value(H) (two-sided):	0.01
Jarque-Bera (JB):	8926.66
P-value(JB):	0.00
Skew:	0.07
Kurtosis:	8.29

El valor del AIC de este modelo durante la validación es 16208.82.

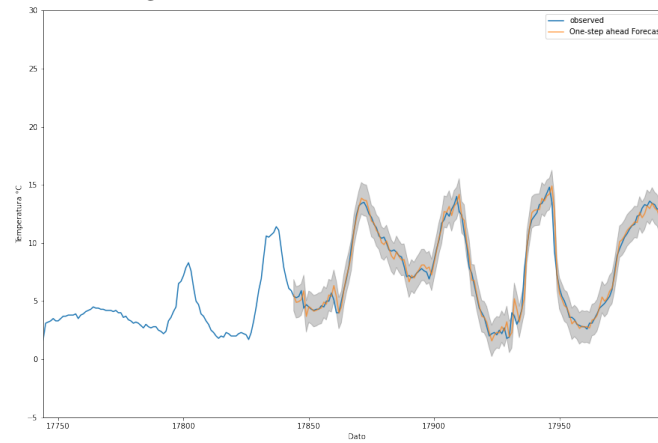
La figura VIII.c.2 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.50. El 11.40 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 1.78 mayor 2°C y un 0.37 % mayor a 3 °C.

### VIII.c.3. Modelo 3

Este modelo se calcula retirando las componentes sinusoidales. A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Figura 10: Resultados Modelo VIII.c.2



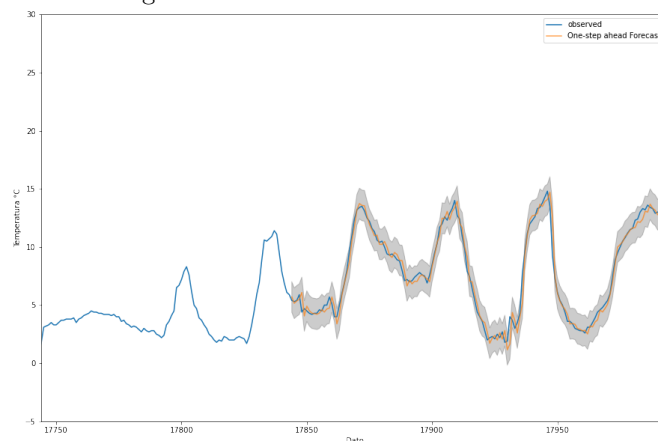
Parametro	Valor
Intercepto:	$-5.89 \cdot 10^{-6}$
S:	12 (horas) o 18 intervalos
$\beta$ :	0
d:	1
D:	0
$\phi_1$ :	0.63
$\Phi_1$ :	-0.00
$\Phi_2$ :	0.99
$\theta_1$ :	-0.32
$\Theta_1$ :	0
$\Theta_2$ :	-0.92
$\sigma^2$ :	0.46

A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	544.05
P-value(Q):	0.00
Heteroskedasticity (H):	1.27
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	15266.42
P-value(JB):	0.00
Skew:	-0.13
Kurtosis:	9.93

El valor del AIC de este modelo durante la validación es 16087.34.

Figura 11: Resultados Modelo VIII.c.3



La figura VIII.c.3 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.49. El 11.40 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 1.78 mayor 2°C y un 0.34 % mayor a 3 °C.

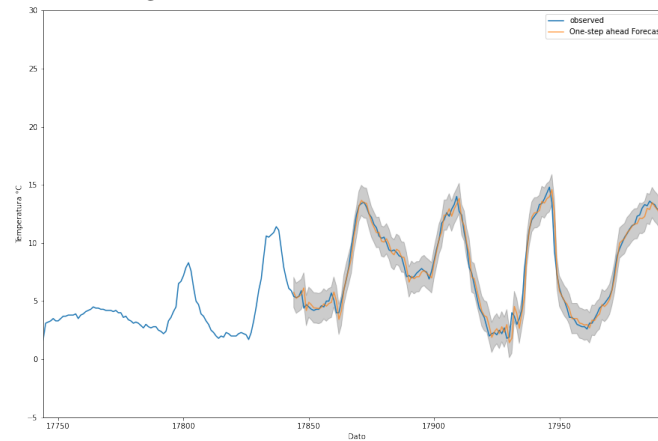
#### VIII.c.4. Modelo 4

Este modelo se obtiene a partir de un modelo de componentes AR para todos los intervalos posibles hasta alcanzar un día, más una componente MA que es estacional de duración de un día. A este modelo se retiran componentes no significativos y de alto orden, para posteriormente seguir un procedimiento equivalente a los anteriores para llegar al modelo elegido.

A continuación, se parametriza las ecuaciones 2 a 9 con que se describe el modelo resultante.

Parametro	Valor
Intercepto:	0.00
S:	12 (horas) o 18 intervalos
$\beta$ :	0
d:	0
D:	0
$\phi_1$ :	1.28
$\phi_2$ :	-0.15
$\phi_3$ :	-0.16
$\Phi_1$ :	-0.00
$\Phi_2$ :	0.99
$\Theta_1$ :	0
$\Theta_2$ :	-0.91
$\sigma^2$ :	0.44

Figura 12: Resultados Modelo VIII.c.4



A continuación, se exponen los resultados de las pruebas realizados para la validación del modelo.

Test	Valor
Ljung-Box (Q):	389.39
P-value(Q):	0.00
Heteroskedasticity (H):	1.25
P-value(H) (two-sided):	0.00
Jarque-Bera (JB):	14435.84
P-value(JB):	0.00
Skew:	-0.13
Kurtosis:	9.74

El valor del AIC de este modelo durante la validación es 15807.17.

La figura VIII.c.4 muestra el término del periodo de ajuste y el inicio del periodo de validación para las predicciones de temperatura.

El valor promedio del valor absoluto del error en este modelo es 0.48. El 11.30 % de las predicciones tienen un valor absoluto del error mayor a 1°C, un 1.67 mayor 2°C y un 0.26 % mayor a 3 °C.

## **IX. Validación**

A continuación, se analiza los resultados para la validación de los modelos.

### **IX.a. Chillán**

#### **IX.a.1. Estructura de Correlación**

Para todos los modelos utilizados en la estación de Chillán el valor de la prueba de Box-Ljung tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de independencia de los residuos con una significancia del 95 %.

#### **IX.a.2. Normalidad**

Para todos los modelos utilizados en la estación de Chillán el valor de la prueba de Box-Ljung tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de normalidad de los residuos con una significancia del 95 %.

#### **IX.a.3. Homocedasticidad**

Los valores p de la prueba de homocedasticidad para los distintos modelos son 0.09, 0.13, 0.12 y 0.07. Ninguno de ellos permite rechazar la hipótesis nula de homocedasticidad con una significancia mayor al 95 %.

#### **IX.a.4. AIC**

El AIC de los modelos durante la validación es de 19150.63, 19060.15, 19623.15 y 19626.25. Esto indica que claramente el Modelo 2 es el que tiene el menor valor de AIC entre los cuatro modelos estudiados.

#### **IX.a.5. Valor Absoluto del Error**

En todos los casos vistos hasta ahora, el promedio del valor absoluto del error es menor a 1 °C, menos del 20 % de las predicciones tiene un error mayor a 1°C, menos de un 5 % de las predicciones tiene un error mayor a 2°C y menos de un 1 % de las predicciones tiene un error mayor a 3°C. Para terminar con esta estación, el modelo 2 también es el que tiene menor promedio del valor absoluto del error y menor porcentajes de error mayores a 1, 2 y 3 °C.

### **IX.b. Pichilemu**

#### **IX.b.1. Estructura de Correlación**

Para todos los modelos utilizados en la estación de Pichilemu el valor de la prueba de Box-Ljung tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de independencia de los residuos con una significancia del 95 %.

### **IX.b.2. Normalidad**

Para todos los modelos utilizados en la estación de Pichilemu el valor de la prueba de Box-Ljung tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de normalidad de los residuos con una significancia del 95 %.

### **IX.b.3. Homocedasticidad**

Para todos los modelos utilizados en la estación de Pichilemu el valor de la prueba de Homocedasticidad tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de normalidad de los residuos con una significancia del 95 %.

### **IX.b.4. AIC**

El AIC de los modelos durante la validación es de 12899.93, 13215.13, 13697.43 y 13879.25. Esto indica que claramente el Modelo 1 es el que tiene el menor valor de AIC entre los cuatro modelos estudiados.

### **IX.b.5. Valor Absoluto del Error**

En todos los casos vistos hasta ahora, el promedio del valor absoluto del error es menor a 0.5 °C, menos del 10 % de las predicciones tiene un error mayor a 1°C, menos de un 1.2 % de las predicciones tiene un error mayor a 2°C y menos de un 0.2 % de las predicciones tiene un error mayor a 3°C. Para terminar con esta estación, el modelo 1 tiene menor promedio del valor absoluto del error y menor porcentajes de error mayores a 1, 2 °C. Pero el modelo 2 tiene el menor porcentaje de errores mayores a 3°C.

## **IX.c. Manquehue**

### **IX.c.1. Estructura de Correlación**

Para todos los modelos utilizados en la estación de Manquehue el valor de la prueba de Box-Ljung tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de independencia de los residuos con una significancia del 95 %.

### **IX.c.2. Normalidad**

Para todos los modelos utilizados en la estación de Manquehue el valor de la prueba de Box-Ljung tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de normalidad de los residuos con una significancia del 95 %.



### **IX.c.3. Homocedasticidad**

Para todos los modelos utilizados en la estación de Manquehue el valor de la prueba de Homocedasticidad tiene un valor p de aproximadamente 0, por lo que no se puede aceptar la hipótesis de normalidad de los residuos con una significancia del 95 %.

### **IX.c.4. AIC**

El AIC de los modelos durante la validación es de 15901.31, 16208.82, 16087.34 y 15807.17. Esto indica que claramente el Modelo 1 es el que tiene el menor valor de AIC entre los cuatro modelos estudiados.

### **IX.c.5. Valor Absoluto del Error**

En todos los casos vistos hasta ahora, el promedio del valor absoluto del error es menor a 0.5 °C, menos del 12 % de las predicciones tiene un error mayor a 1°C, menos de un 2 % de las predicciones tiene un error mayor a 2°C y menos de un 0.5 % de las predicciones tiene un error mayor a 3°C. Para terminar con esta estación, el modelo 1 tiene menor promedio del valor absoluto del error y menor porcentajes de error mayores a 1°C. Pero el modelo 4 tiene el menor porcentaje de errores mayores a 2 y 3 °C.

## X. Conclusión

Este proyecto ha permitido observar que es posible obtener bastante buenas aproximaciones de la temperatura mediante modelos SARIMAX considerando estacionalidades y valores históricos de la temperatura.

Respecto de los modelos, como se puede observar en los resultados, la magnitud de los errores es limitada por lo que los modelos son de utilidad, sin embargo, estos errores presentan fenomenologías que no han sido eliminadas durante la ejecución del proyecto, las que han impedido validarlos. Es posible que, al aumentar el tamaño de las bases de datos, considerando un mayor número de años, se obtenga modelos que permitan obtener residuos con mejores comportamientos.

Se ha cumplido el objetivo general del proyecto, porque los desarrollos permiten observar que es posible anticipar cambios de temperatura con una antelación de 40 minutos. Los modelos cometen un error absoluto que es mayor a  $1^{\circ}\text{C}$  entre el 10 % y el 20 % de los casos, mayor a  $2^{\circ}\text{C}$  entre el 1.2 % y el 5 %, y mayor a  $3^{\circ}\text{C}$  entre el 0.2 % y el 1 % de los casos.

Los objetivos específicos se cumplen parcialmente, ha sido posible construir modelos predictivos de la temperatura dentro de los próximos 30 a 60 minutos, se puede determinar que la temperatura y características estacionales de esta permiten hacer buenas predicciones y se determinó distintos modelos para las distintas estaciones.

Los diversos modelos construidos permiten observar distintos niveles de precisión de las predicciones, así como distintos estadísticos, comportamientos y formas de modelar algunos fenómenos. Es destacable que la fenomenología estacional asociada a la hora del día se representa de distintas maneras en distintos modelos. Algunos modelos favorecen la incorporación de esta estacionalidad mediante variables exógenas de comportamiento sinusoide. También es destacable que para un caso la estacionalidad asociada al mes es representada mediante una variable exógena sinusoide, esta fenomenología a diferencia de la horaria, no es representada de manera explícita por otros medios, aunque no sea significativa.

Queda propuesto para desarrollos próximos la construcción de modelos que permitan su validación completa, determinar más variables que puedan ser útiles para la predicción de la temperatura y analizar específicamente el poder predictivo de los modelos bajo ciertas temperaturas.

## XI. Anexos

### XI.a. Anexo I: Descarga

```
#!/bin/bash

for Estacion in 340093 390028 390043 340045 340115 340066 340031 350028
360047 360011 360045 360019 360046 370036 370067 380063 380013 380029
380018 380032 380033 390029 390006 390015
do
    for ano in 2018 2019
    do
        for mes in 01 02 03 04 05 06 07 08 09 10 11 12
        do
            for dato in Temperatura RadiacionGlobal
            PresionHumedad AguaCaida Viento
            do
                echo "https://climatologia.meteochile.gob
.....cl/application/productos/
.....getDatosHistoricosEma/
.....${Estacion}/${Estacion}_${ano}${mes}_${
.....{dato}.csv.zip"
                wget "https://climatologia.meteochile.gob.
.....cl/application/productos/
.....getDatosHistoricosEma/${Estacion}/${
.....{Estacion}_${ano}${mes}_${dato}.csv.zip"
            done
        done
    done
done
```

## XI.b. Anexo II: Descompresión

```
#!/bin/bash

for Estacion in 390043 340045 340093 340115 340066 340031 350028 360047
360011 360045 360019 360046 370036 370067 380063 380013 380029 390028
380018 380032 380033 390029 390006 390015
do
    for ano in 2018 2019
    do
        for mes in 01 02 03 04 05 06 07 08 09 10 11 12
        do
            for dato in Temperatura RadiacionGlobal
            PresionHumedad AguaCaida Viento
            do
                echo "${Estacion}_${ano}${mes}_${dato}."
                .....csv.zip"
                unzip "${Estacion}_${ano}${mes}_${dato}."
                .....csv.zip"
            done
        done
    done
done
done
done
```

## Antecedentes Bibliográficos

- [*Snyder, Melo-Abreu (2010)*] Richard L. Snyder, J. Paulo de Melo-Abreu, Traducción: Josep M. Villar-Mir; "*Protección contra las heladas: fundamentos, práctica y economía*"; Organización de las Naciones Unidas para la Agricultura y la Alimentación, Roma.
- [*Aguado-Rodríguez et al. (2016)*] G. Javier Aguado-Rodríguez, Abel Quevedo-Nolasco, Martiniano Castro-Popoca, Ramón Arteaga-Ramírez, M. Alberto Vázquez-Peña, B. Patricia Zamora-Morales; "*Predicción De Variables Meteorológicas Por Medio De Modelos Arima*".
- [*Irigoyen (2011)*] *Reconocimiento De Patrones Meteorológicos Mediante Técnicas Neuronales Y Estadísticas*"; Titulación de Miguel Irigoyen Mancho.
- [1] Selva Prabhakaran; [*ARIMA Model - Complete Guide to Time Series Forecasting in Python*]; [Link](#)
- [2] [*Guía de usuario de librería statsmodels*]; [Link](#); [Link](#)
- [3] [*GNU Bash*]; [Link](#)
- [4] [*Test for heteroskedasticity of standardized residuals*]; [Link](#); [Link](#)
- [5] [*State space models*]; [Link](#)
- [6] [*Regression Diagnostics and Specification Tests*]; [Link](#)
- [7] [*A Guide to Time Series Forecasting with ARIMA in Python 3*]; [Link](#)
- [8] [*Source code for statsmodels.tsa.statespace.mlemodel*]; [Link](#)
- [9] [Dirección Meteorológica de Chile]; [Link](#)