

Detección de olas de frío a través de modelos DLM

Estimación suavizada de un percentil

Manuel Hernández Banadik^{1,2} & Ignacio Álvarez-Castro¹

¹Instituto de Estadística. FCEA

²Departamento de Estadística, Biometría y Cómputo. FAGRO

mhernandez@fagro.edu.uy

nachalca@iesta.edu.uy

Universidad de la República

Resumen

Contamos con datos de mediciones diarias de temperaturas mínimas y máximas, de una estación meteorológica de Uruguay, desde Enero-1950 a Octubre-2014.

Se define una ola de frío, como un período de tiempo en el cual, la temperatura observada es inferior a un umbral. El objetivo es determinar dicho umbral a través de la estimación del percentil 10 de las temperaturas. Utilizaremos los modelos lineales dinámicos para modelar la serie, y para la estimación de los percentiles.

Olas de frío

Existen diversas formas de caracterizar una racha de frío, que responden a distintas aplicaciones de su estudio. Las diversas definiciones acuerdan en la necesidad de establecer un umbral de bajas temperaturas (puede ser absoluto o relativo) y en delimitar una ventana de tiempo mínimo. Durante ese tiempo, la temperatura observada debe mantenerse siempre por debajo del umbral definido.

Para el presente trabajo, hemos definido una ola de frío como un período de tiempo mayor o igual a 3 días, en los cuales las temperaturas mínimas y máximas son inferiores a los respectivos percentiles 10 esperados para tales días.

Esta definición requiere de definir para cada día del año un percentil 10 de temperatura mínima y máxima. Para cada $t \in \{1, \dots, 365\}$ definimos el percentil 10 de mínima como $p_{10_t}^n := \inf\{y : \mathbb{P}(Y_t^n \leq y) \geq 0,1\}$ siendo Y_t^n la temperatura mínima observada en el día t . Análogamente definimos el percentil 10 de máxima para el día t como $p_{10_t}^x := \inf\{y : \mathbb{P}(Y_t^x \leq y) \geq 0,1\}$ donde Y_t^x es la temperatura máxima observada en el día t .

Podemos decir entonces que una sucesión de días t_1, \dots, t_k constituyen una ola de frío de largo k si, siendo $k \geq 3$, se cumple simultáneamente que:
$$\begin{cases} y_{t_i}^n < p_{10_{t_i}}^n \\ y_{t_i}^x < p_{10_{t_i}}^x \end{cases} \quad \text{para } i = 1, \dots, k$$

Dynamic Linear Models

Un modelo lineal dinámico (DLM) considera una serie de tiempo, como una observación ruidosa de un proceso dinámico. La figura 1 muestra un ejemplo de un proceso concebido de esta forma. El modelo está definido a través de una ecuación de evolución y una ecuación de observación. La ecuación de evolución define cómo se vincula el estado del proceso en tiempo t con su estado en tiempo $t+1$; la ecuación de observación, explicita cómo es observado el proceso en tiempo t en función del estado en el que se encuentra a t . En una expresión y/o en la otra, se introduce un componente aleatorio.

Tenemos entonces la siguiente expresión para el modelo:

$$\begin{aligned} Y_t &= F_t \theta_t + v_t & v_t &\sim \mathcal{N}(0, V_t) \\ \theta_t &= G_t \theta_{t-1} + w_t & w_t &\sim \mathcal{N}(0, W_t) \end{aligned} \quad (1)$$

donde las matrices F_t, G_t, V_t, W_t serán los parámetros que caractericen el modelo. \mathcal{N} es la distribución normal.

Si se tiene las observaciones y_1, \dots, y_T , puede ser de interés obtener una estimación de $\mathbb{E}(\theta_t | y_1, \dots, y_t)$, es decir, el valor esperado del proceso que subyace a las observaciones en tiempo t , dado que tenemos las observaciones pasadas. Esta cantidad es estimada mediante el Filtro de Kalman.

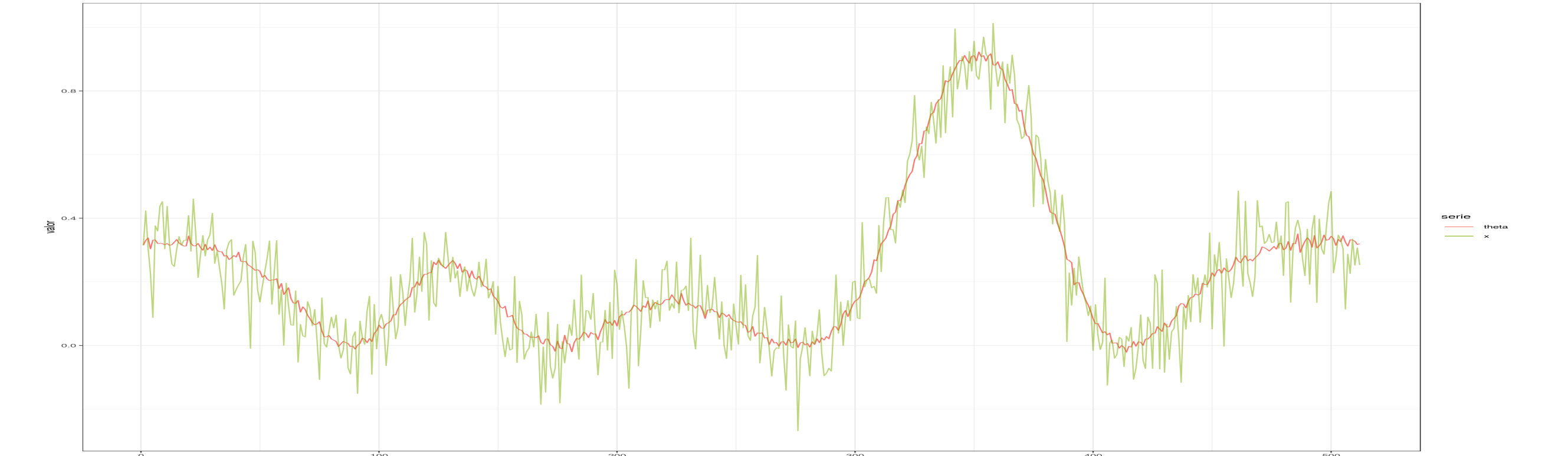


Figura 1: Ejemplo de un proceso de espacio de estado. En rojo, se representa el proceso *no-observable*, en verde se representa su observación.

DLM estacional para las temperaturas

Modelaremos la serie de mínimas y máximas de forma independiente, utilizaremos el mismo modelo para ambas.

Dada la fuerte característica estacional de los datos de temperaturas, es apropiado modelarla como una función periódica más un ruido aleatorio. En este caso, el proceso latente será modelado como una función periódica de período 365 y su dinámica no será aleatoria, esto se traduce en imponer matrices de varianza nulas para la ecuación de evolución. Tal función está caracterizada por un vector $\alpha \in \mathbb{R}^{365}$, la representación de α en la base canónica nos habla de su naturaleza temporal -*qué temperatura esperamos cada día del año*-. Será de mayor interés trabajar con su representación en series de Fourier. Esto nos permitirá trabajar en un espacio de dimensión significativamente menor, con muy poca pérdida de información, dado que unos pocos armónicos (en nuestro caso 2) tienen coeficientes de fourier 'significativos', definiendo así un criterio para descartar armónicos de mayor frecuencia, responsables de la rugosidad de la serie.

En este caso el modelo será:

$$\theta_t = \begin{pmatrix} S_1 := a_1 \cos\left(\frac{2\pi t}{365}\right) + b_1 \sin\left(\frac{2\pi t}{365}\right) \\ S_1^* := -a_1 \sin\left(\frac{2\pi t}{365}\right) + b_1 \cos\left(\frac{2\pi t}{365}\right) \\ S_2 := a_2 \cos\left(\frac{4\pi t}{365}\right) + b_2 \sin\left(\frac{4\pi t}{365}\right) \\ S_2^* := -a_2 \sin\left(\frac{4\pi t}{365}\right) + b_2 \cos\left(\frac{4\pi t}{365}\right) \end{pmatrix} \quad F_t = (1, 0, 1, 0) \quad \forall t$$

La matriz G_t (constante en t) será una matriz diagonal por bloques, cuyos dos bloques serán matrices de rotación de ángulos horarios $\frac{2\pi}{365}$ y $\frac{4\pi}{365}$. La matriz W_t se definió nula. La varianza observacional V_t será estimada por máxima verosimilitud.

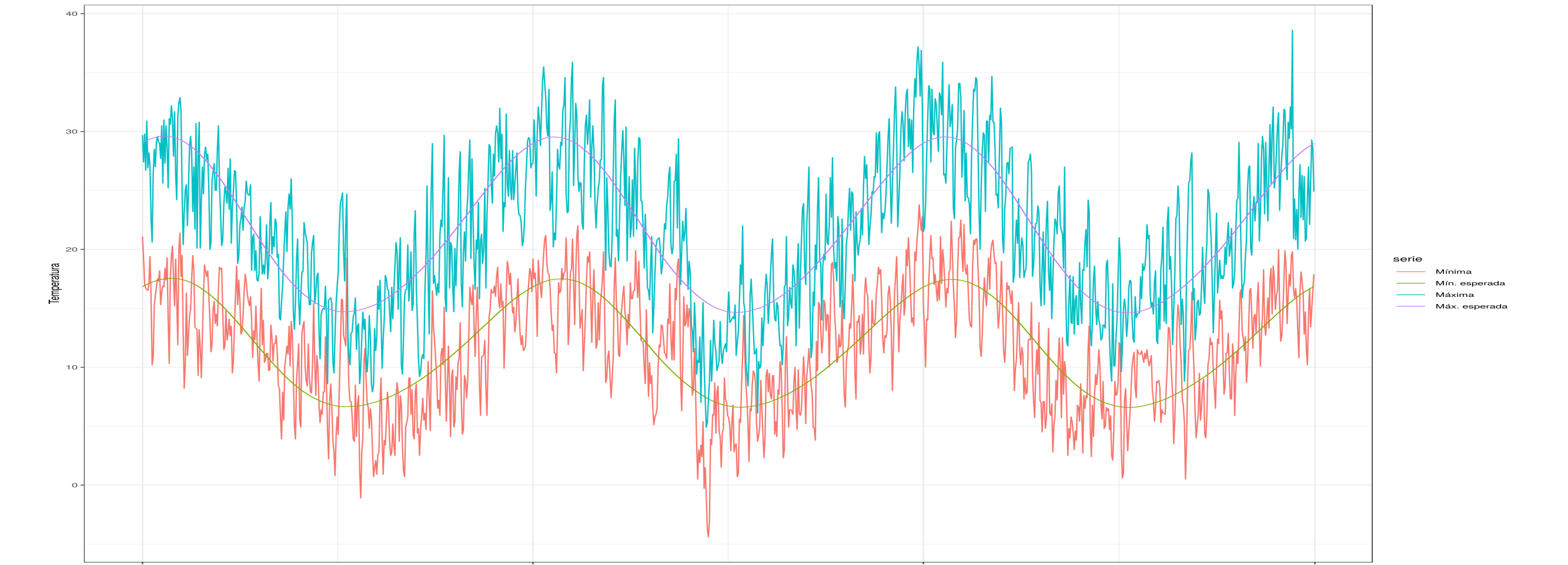


Figura 2: Gráfico de mínimas y máximas del 1966 al 1969, y la estimación del valor esperado del proceso, mediante el filtro de Kalman

Implementación en R - paquete 'dlm'

El paquete `dlm` permite trabajar con toda la familia de modelos lineales dinámicos.

Si el objeto `y` contiene una de las series de temperaturas, el siguiente programa nos dará la estimación del modelo.

Lo primero es estimar la varianza observacional con la función `dlmMLE`, que aproximará una estimación óptima, con un algoritmo iterativo. Para ello, primero debemos crear una función auxiliar (`parMLE`), cuyo argumento sea un valor inicial propuesto para el parámetro, que retorne un objeto de la clase `dlm`, en este caso de la familia de modelos trigonométricos, donde la frecuencia es 365. Le especificamos que mantenga 2 armónicos y descarte el resto:

```
parMLE <- function(par) dlmModTrig(s = 365,q=2, dV = par, dW = 0)
```

La función `dlmMLE` devolverá el modelo, con el/los parámetros estimados:

```
mod<- dlmMLE(y,parm = 16,build= parMLE)
```

Una vez estimado el modelo, podemos aplicar el filtro de Kalman (`dlmFilter`) para obtener estimaciones de la esperanza *a posteriori*:

```
Tn.filter <- dlmFilter(y,parMLE(mod$par))
```

En la figura 2 se muestra en un período de tres años, las series de mínima y máxima y la estimación de su valor esperado mediante el filtro de Kalman.

Estimación de percentiles

Como se explicó anteriormente, para caracterizar olas de frío debemos estimar los percentiles 10 de temperaturas mínimas y máximas para cada día del año. Tenemos las series temporales de los datos de mínima y máxima, a escala diaria, desde el 1-Enero de 1950 hasta el 10-October de 2014, por tanto para cada día del año (del 1-Enero al 31-Diciembre) tenemos más de sesenta observaciones de temperaturas. Esto da para cada día del año una distribución de temperaturas.

En este trabajo, para la estimación de tales percentiles, nos valdremos de que en el modelo descrito en la expresión (1) se cumple $\theta_t | \{y_1, \dots, y_t\} \sim \mathcal{N}(m_t, C_t)$, y $y_t | \theta_t \sim \mathcal{N}(F_t \theta_t, V_t)$. El filtro de Kalman nos devuelve las secuencias $(m_t)_{1 \leq t \leq T}$ y $(C_t)_{1 \leq t \leq T}$ para la serie de temperaturas mínimas y máximas. Por tanto el estimador para el percentil 10 de mínima para el día t será $\hat{p}_{10_t}^n := F_t m_t + \sqrt{F_t C_t F_t^T} + v_t \times z_{0,10}$. Análogamente se hizo lo mismo para el percentil 10 de las máximas.

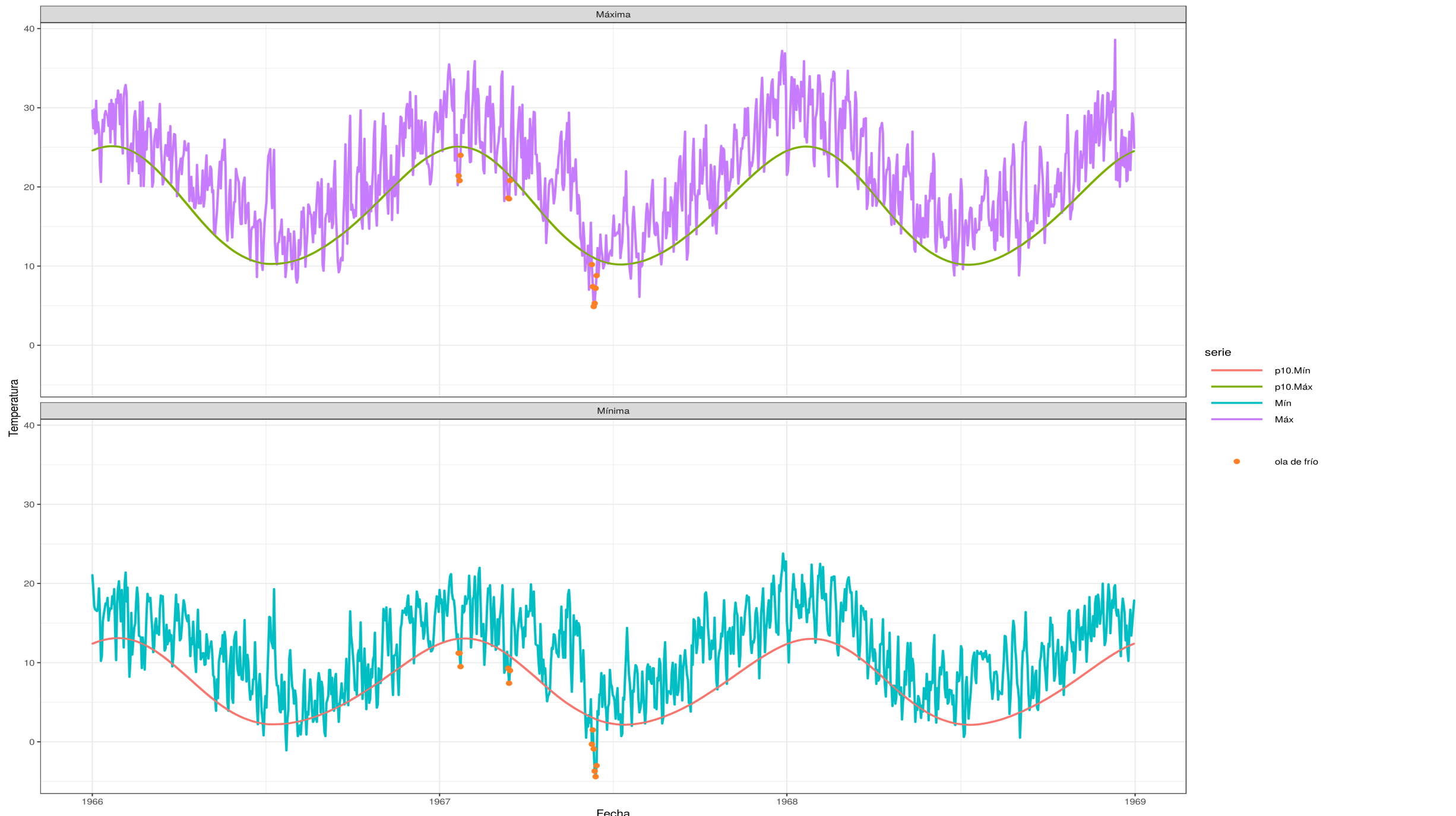


Figura 3: Mínimas y máximas del 1966 al 1969, percentiles 10, y olas de frío detectadas

Fecha	Mín	Máx	p10 mín	p10 máx
1967-06-09	5.40	3.14	15.50	11.25
1967-06-10	-0.30	3.08	10.20	11.19
1967-06-11	1.50	3.03	7.40	11.12
1967-06-12	-0.90	2.97	4.90	11.06
1967-06-13	-3.70	2.92	5.30	11.00
1967-06-14	-4.40	2.86	7.20	10.94
1967-06-15	-3.00	2.81	8.80	10.88
1967-06-16	3.90	2.77	12.30	10.83

Cuadro 1: Mínimas, máximas y percentiles 10 de mínimas y máximas, en ocasión de ola de frío detectada, desde el 10 al 16 de junio de 1967,

Referencias

- Petris, G., Petrone, S., & Campagnoli, P. (2009). Dynamic linear models. In Dynamic Linear Models with R (pp. 31-84). Springer, New York, NY.
- Niemi, J. (2012). STAT 615: Advanced Bayesian Methods [Beamer slides]. Retrieved from <http://www.jarad.me/courses/stat615/slides/DLMs/DLMs.pdf>