



UNIVERSIDAD DE LA REPÚBLICA

Facultad de Ciencias Económicas y Administración

Instituto de Estadística

## **Título**

Manuel Hernández Banadik; Ignacio Álvarez-Castro

Abril, 2019

## **Serie Documentos de Trabajo**

DT (numero) - ISSN : 1688-6453

Forma de citación sugerida para este documento:

**queda de postre ...**

## Titulo ...

Manuel Hernández Banadik <sup>1</sup>; Ignacio Álvarez-Castro <sup>2</sup>

*Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, 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.

## 1. Introducción

El estudio de las temperaturas extremas ... bla bla

En la figura 1 se grafica la serie de temperaturas (máxima y mínima) para el año 2012, y de fondo aparecen gráficos de caja con la distribución de la temperatura para cada día. En los días señalados en azul (del 5 al 9 de junio de 2012), se nota que la temperatura observada se acerca a la cola inferior de la distribución para ese día. Eso nos indica que ese día fue bastante más frío de lo esperado, más aún, es muy raro observar la temperatura que se observó en dicho período. Podemos considerar que durante esos días estuvimos ante la presencia de una ola de frío. ... bla bla ... Veremos una definición más precisa de ola de frío ...

---

<sup>1</sup>*email:* [mhernandez@iesta.edu.uy](mailto:mhernandez@iesta.edu.uy), ORCID: 0000-0002-2505-4238

<sup>2</sup>*email:* [nachalca@iesta.edu.uy](mailto:nachalca@iesta.edu.uy), ORCID: 0000-0003-3123-2165

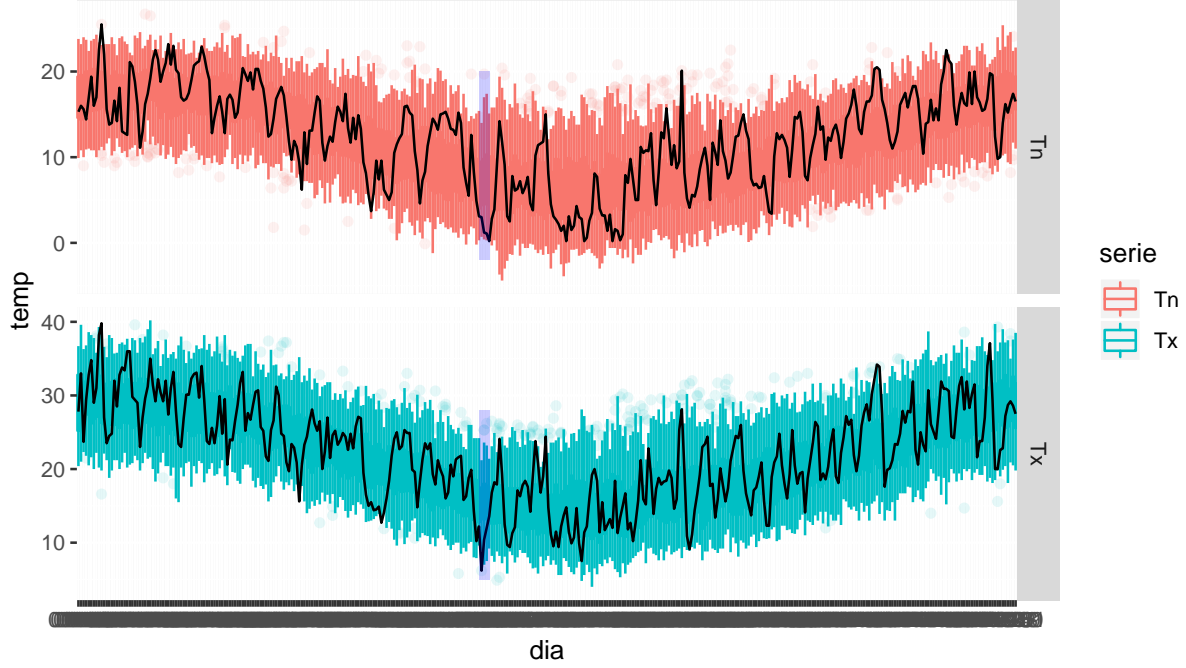


Figura 1: Gráfico

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 durante la cual, la temperatura observada debe mantenerse 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. Siendo  $\{Y_t^n\}$  y  $\{Y_t^x\}$  la serie diaria de temperaturas mínima y máxima respectivamente, para cada  $t \in \{1, \dots, 365\}$  definimos el percentil 10 de mínima como  $p_{10_t}^n := \inf\{y : \mathbb{P}(Y_t^n \leq y) \geq 0.1\}$ . Análogamente definimos el percentil 10 de máxima para el día  $t$  como  $p_{10_t}^x := \inf\{y : \mathbb{P}(Y_t^x \leq y) \geq 0.1\}$ .

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:  $\begin{cases} y_{t_i}^n < p_{10_i}^n \\ y_{t_i}^x < p_{10_i}^x \end{cases}$  para  $i = 1, \dots, k$

En la sección 2 introducimos la familia de Modelos Lineales Dinámicos, que fueron usados para modelar los datos, así como la estimación de los percentiles. En la sección 2.1 mostramos la implementación en **R** de la estimación de un modelo simulado. En la sección 3 aplicamos la metodología descripta a los datos de temperaturas de Uruguay y presentamos los resultados.

## 2. Modelos lineales dinámicos

Un modelo lineal dinámico (DLM) considera una serie de tiempo ( $Y_t$ ), como una observación ruidosa de un proceso dinámico no observable ( $\theta_t$ ). El modelo está definido a través de una ecuación de evolución y una ecuación de observación. La ecuación de evolución responde a la dinámica del proceso estocástico: 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  $t$  dado el estado en el que se encuentra a tiempo  $t$ . En ambas ecuaciones, se puede introducir una componente aleatorio. Un diagrama de esto se puede ver en la figura 2.

Tenemos entonces la siguiente expresión para el modelo:

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

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

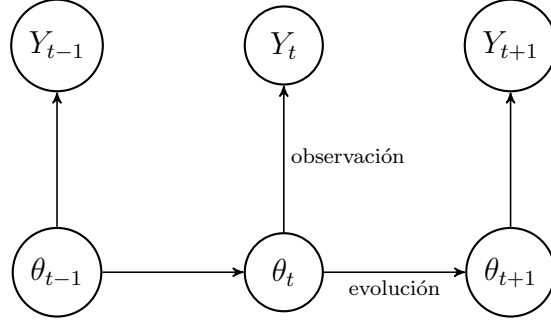


Figura 2: Diagrama de evolución y observación del proceso

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.

La figura 3 muestra un ejemplo simulado de un proceso univariado concebido de esta forma, bajo el modelo descrito en (2). Así mismo, se muestra la estimación del proceso  $\theta_t$  obtenida a través de la implementación del filtro de Kalman en **R**.

$$\begin{aligned} Y_t &= \theta_t + v_t & v_t &\sim \mathcal{N}(0, 0.1^2) \\ \theta_t &= \theta_{t-1} + w_t & w_t &\sim \mathcal{N}(0, 0.3^2) \end{aligned} \quad (2)$$

En este caso  $G_t$  y  $F_t$  se definieron constantes con valor 1. El proceso  $\{\theta_t\}$  describe un paseo al azar, mientras que podemos considerar a  $\{y_t\}$  como una observación ruidosa de  $\{\theta_t\}$

## 2.1. Implementación computacional en R

El paquete **d1m** 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 los parámetros desconocidos del modelo con la función **d1mMLE**,

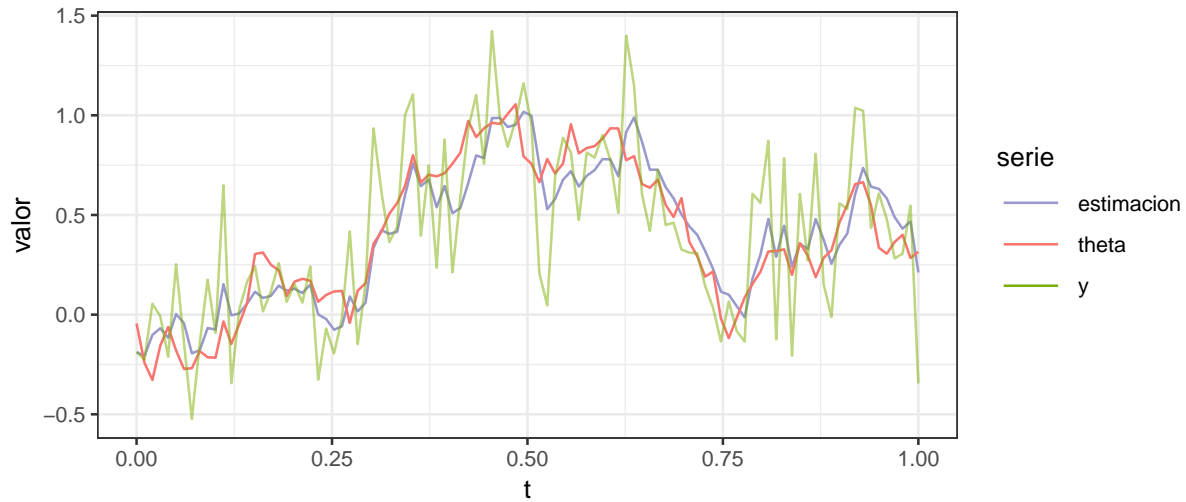


Figura 3: Ejemplo de un proceso de espacio de estado y su estimación.

que aproximará una estimación óptima, con un algoritmo iterativo. Para ello, primero debemos crear una función auxiliar (`parMLE`), cuyo argumento sea un vector de valores iniciales propuesto para los parámetros:

`parMLE <- function(par) dlmModPoly(1, dV = par[1], dW = par[2])`. En nuestro caso desconocemos  $V_t$  y  $W_t$ , la varianza del ruido de evolución y la varianza del ruido de observación.

En este caso, si el objeto `y` contiene la serie de datos, la función `dlmMLE` devolverá el modelo, con los parámetros estimados:

```
mod<- modPoly <- dlmMLE(y,parm = c(1,1),build = parMLE)
```

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

```
mod_filter <- dlmFilter(y, parMLE(modPoly$par))
```

```
parMLE <- function(par) dlmModPoly(1, dV = par[1], dW = par[2])
modPoly <- dlmMLE(y,parm = c(1,1),build = parMLE)
mod_filter <- dlmFilter(y, parMLE(modPoly$par))
mod_filter$m
```

### 3. Modelo estacional para el modelado de 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 de período 365s 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  $\alpha$  en la base canónica nos habla de su naturaleza temporal -*qué temperatura esperamos cada día del año*-. Por distintos motivos será de mayor interés considerar la representación de  $\alpha$  en una base 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:

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

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.

#### 3.1. 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-Octubre 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 la distribución de  $\theta_t|\{y_1, \dots, y_t\}$ . Se puede probar que, bajo el modelo descrito en la expresión (1), se cumple que  $\theta_t|\{y_1, \dots, y_t\} \sim \mathcal{N}(m_t, C_t)$ , y  $y_t|\theta_t \sim \mathcal{N}(Fm_t, FC_tF^T + v_t)$ . 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  $\hat{p}_{10_t}^n := Fm_t + \sqrt{FC_tF^T + v_t} \times \Phi(0.10)$ , donde  $\Phi(x) = \mathbb{P}(X \leq x)$  con  $X \sim N(0, 1)$ . Análogamente se hace lo mismo para el percentil 10 de temperaturas máximas.

## 4. Implementación en R - paquete ‘dlm’

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

```
parMLE <- function(par)  dlmModTrig(s = 365,q=2, dV = par, dW = 0)
mod<- dlmMLE(y,param = 16,build= parMLE)
Tn.filter <- dlmFilter(y,parMLE(mod$par))
```

El objeto `Tn.filter$m` tendrá la estimación.

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,param = 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 4 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.

## 5. Resultados

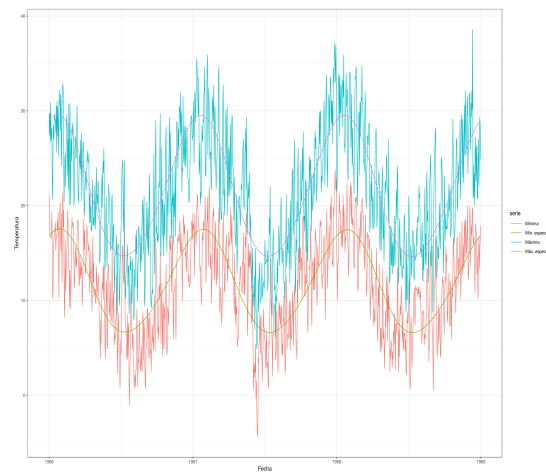


Figura 4: Gráfico de mínimas y máximas del 1966 al 1969, y la estimación del valor esperado del proceso, mediante el filtro de Kalman

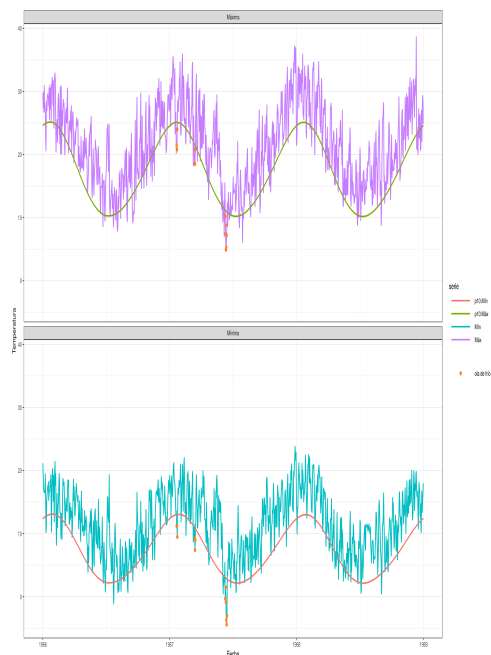


Figura 5: Mínimas y máximas del 1966 al 1969, percentiles 10, y olas de frío detectadas

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

Tabla 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,

Palabras clave: CÓDIGOS JEL: Clasificación MSC2010:

titulo in english...

authors ...

## ABSTRACT

oh boy ....

**Key words:**

**JEL CODES:**

**Mathematics Subject Classification MSC2010:**

- 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>

# Instituto de Estadística

---

## Documentos de Trabajo



**Eduardo Acevedo 1139. CP 11200 Montevideo, Uruguay**  
**Teléfonos y fax: (598) 2410 2564 - 2418 7381**  
**Correo: [ddt@iesta.edu.uy](mailto:ddt@iesta.edu.uy)**  
**[www.iesta.edu.uy](http://www.iesta.edu.uy)**  
**Área Publicaciones**

**Diciembre, 2017**  
**DT (17/3)**