



UNIVERSIDAD DE LA REPÚBLICA

Facultad de Ciencias Económicas y Administración

Instituto de Estadística

Titulo

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 ¹; Ignacio Álvarez-Castro ²

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 eventos extremos ha tomado una gran relevancia en los últimos años debido principalmente al gran impacto que presentan en la sociedad y la economía de los países, así como en los ecosistemas. Dentro de la región del Sudeste de Sudamérica los principales eventos climáticos extremos que han sido analizados son los relacionados con la temperatura (olas de calor y de frío, heladas, días cálidos, etc) y a la precipitación (sequías, precipitaciones intensas, etc). Muchos de estos estudios son relativamente actuales, debido principalmente a que es necesario contar con series temporales lo suficientemente largas y fundamentalmente completas, a un paso temporal de por lo menos un día y de alta calidad.

Bajo el escenario de cambio climático, es necesario comprender cómo los eventos extremos cambian en frecuencia y/o intensidad, identificando cómo se han comportado en las

¹email: mhernandez@iesta.edu.uy, ORCID: 0000-0002-2505-4238

²email: nachalca@iesta.edu.uy, ORCID: 0000-0003-3123-2165

últimas 3 o 4 décadas. Identificar este tipo de eventos ayudará a comprender su dinámica para luego realizar previsiones a un plazo menor asociado a la variabilidad climática y hasta incluso estudiar el comportamiento a escala sinóptica o de días. Uruguay cuenta con registros suficientemente largos en ciertos puntos del país, pero se han detectado muchos períodos sin información, tanto para temperaturas extremas diarias como para la variable precipitación (Renom, 2009; Amiel, 2012)

En este trabajo se presenta la estimación y caracterización de olas de frío en una estación meteorológica basadas en modelos dinámicos de la serie de temperaturas diarias máximas y mínimas en dicha estación.

La siguiente sección presenta los datos utilizados y una definición de ola de frío. Luego, en la sección 3 introducimos la familia de Modelos Lineales Dinámicos, que fueron usados para modelar los datos, así como la estimación de los percentiles. En la sección 3.3 mostramos la implementación en **R** de la estimación de un modelo simulado. En la sección 4 aplicamos la metodología descripta a los datos de temperaturas de Uruguay y presentamos los resultados.

2. Los datos

Contamos con información de la temperatura máxima y mínima diaria en la estación meteorológica Estanzuela para el período XXX a ZZZ. Definimos $\{Y_{at}^n\}$ y $\{Y_{at}^x\}$ la serie diaria de temperaturas mínima y máxima respectivamente, para $t \in \{1, \dots, 365\}$ y $a \in \{1950, \dots, 2014\}$ (te parece bien esta notación?? si es asi hay que adaptar todo lo demás a esto)

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.

En 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. 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:

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

Vizualizar fenómenos de olas de extremos no es trivial, aplicando la definición anterior, debemos tener en cuenta varios factores. En primer lugar la condición recae simultáneamente sobre la temperatura mínima y máxima por lo que debemos prestar atención a ambas series. Al mismo tiempo debemos comparar la temperatura de cada día en particular con los percentiles para ese día, o más en general, con las temperaturas esperables. Finalmente, los días deben ser consecutivos para formar una ola de extremos.

La figura 1 muestra las series de temperaturas para el año 2012 ([la serie es del 2012 o 2011 ???](#)) su la distribución de frecuencias. El gráfico tiene 3 capas básicas, en primer lugar calculamos un histograma bidimensional en que los bins son hexágonos en el plano de $\{y_{ta}^n, t\}$ (panel superior), la intensidad del color representa la frecuencia. Sobre el histograma se dibuja las series de temperatura observada para cada día en 2012 y finalmente se divide en dos paneles con la serie de mínima (superior) y máxima (inferior) por separado.

[en la figura: caption con descripcion, nombres de los ejes y la etiqueta de color, nombres de los paneles](#)

El período de días del 5 al 9 de junio de 2012 (señalado en rojo en la Figura 1) la temperatura observada se acerca a la cola inferior de la distribución para ese día. Eso indica que ese período fue bastante más frío de lo esperado, podemos considerar que durante esos días estuvimos ante la presencia de una *ola de frío*.

[comentario de cierre??](#)

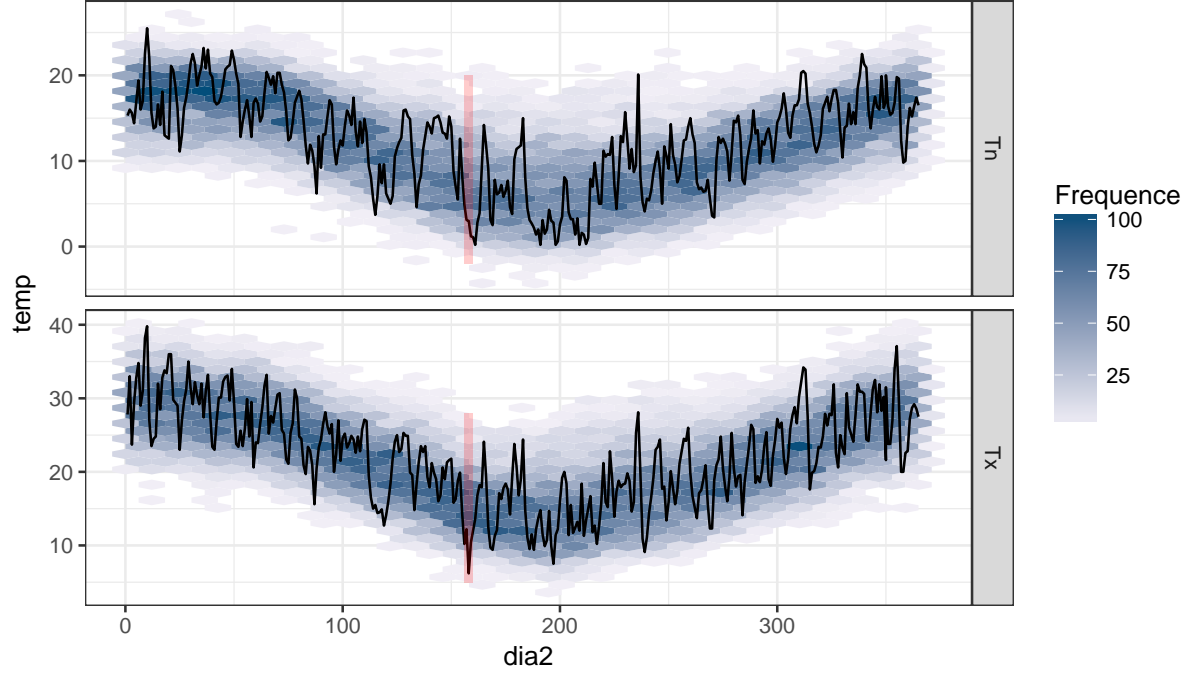


Figura 1: Gráfico

3. Modelos lineales dinámicos

Los modelos lineales dinámicos (DLM) son una familia de modelos que constituye una subclase de los llamados modelos de espacio de estado. En DLM se considera una serie de tiempo, Y_t , como una observación ruidosa de un proceso dinámico no observable θ_t . Estos modelos se definen a través de dos ecuaciones básicas: una *ecuación de evolución* y una *ecuación de observación*. La estructura general de un DLM se muestra en (1) y (2) donde F_t G_t son dos matrices conocidas de orden $p \times p$ y $m \times p$ respectivamente que caracterizan al modelo, las perturbaciones v_t y w_t son secuencias de vectores gaussianos independientes, mientras que V_t, W_t serán los parámetros que caractericen el modelo. Adicionalmente, se agrega una previa Normal para el momento inicial, es decir $\theta_0 \sim N(m_0, C_0)$.

$$Y_t = F_t \theta_t + v_t \quad v_t \sim \mathcal{N}(0, V_t) \quad (1)$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad w_t \sim \mathcal{N}(0, W_t) \quad (2)$$

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 $t + 1$. Por otro lado, 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.

3.1. Estructura de un DLM

La estructura de dependencia entre las observaciones es un aspecto clave para entender como funcionan los DLM. Al igual que la clase más general de modelos espacio-estado en los DLM se basa en dos supuestos fundamentales. En primer lugar se asume que el proceso inobservable o proceso de estado θ_t constituye una cadena de Markov. En tanto, las observaciones Y_t son condicionalmente independientes dado θ_t y se asume que Y_t depende únicamente de θ_t .

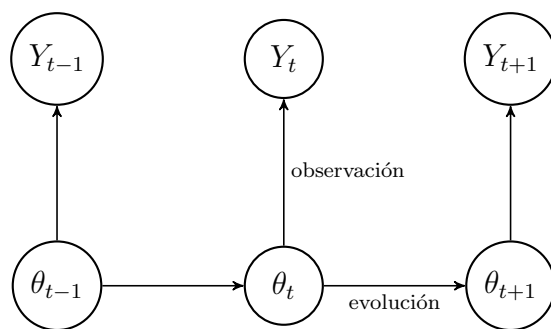


Figura 2: Diagrama de evolución y observación del proceso

La figura 2 representa un diagrama de la estructura de dependencia en DLM, donde los arcos indican dependencia estadística. Se puede ver como θ_t , toda la dependencia entre observaciones pasa a través del proceso latente. A partir de esta estructura, se puede considerar que las observaciones representan una medida ruidosa del proceso de estado.

La principal consecuencia de la estructura de dependencia es que la distribución conjunta de todas las variables de interés queda determinada por la distribución previa del estado

inicial y las densidades condicionales $p(\theta_t|\theta_{t-1})$, $p(y_t|\theta_t)$, de forma:

$$p(\theta_{0:t}, y_{1:t}) = p(\theta_0) \prod_{j=1}^t p(\theta_j|\theta_{j-1})p(y_j|\theta_j)$$

La familia de DLM es amplia y flexible, muchos modelos estadísticos para datos temporales pueden verse como casos particulares de esta familia o de la clase mas general de modelos espacio-estados. En particular los modelos ARMA con ruido gaussiano son equivalentes a DLM cuando los parametros F_t, G_t, V_t, W_t son constantes en el tiempo. Por ejemplo, el proceso AR(2): $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$, puede ser visto como un DLM si definimos $\theta_t = \begin{pmatrix} \theta_{1,t} \\ \theta_{2,t} \end{pmatrix}$ y luego,

$$y_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \theta_t = \theta_{1,t}$$

$$\theta_t = \begin{pmatrix} \theta_{1,t} \\ \theta_{2,t} \end{pmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{pmatrix} \theta_{1,t-1} \\ \theta_{2,t-1} \end{pmatrix} + \begin{bmatrix} e_t \\ 0 \end{bmatrix}$$

para ver la equivalencia entre ambas representaciones, debemos operar en la segunda ecuación, que representa la evolución del proceso latente y verificar que se verifica que $\theta_{1,t} = \phi_1 \theta_{1,t-1} + \phi_2 \theta_{1,t-2} + e_t$. Ver CITA.LIBRO (sección 3.2.5) por más detalles en la correspondencia entre modelos ARIMA(p, d, q) y DLM.

Aparte de DLM, la familia de modelos espacio-estados admite modelos no lineales y el uso de otras distribuciones no Gaussianas. Por ejemplo, modelos para las distribuciones de la familia exponencial pueden ser especificados para tener DLMG *generalizados*, por otro lado en casos uqe el proceso de estados tenga una distribucion discreta el modelo es referido como modelos ocultos de MARKov, tambien modelos de volatilidad estocastica pueden ser vistos como espacio-estados. El estudio de DLM entonces, es relevante pues permite introucarnos al trabajo con una familia de modelos que es muy amplica y puede aplicarse en una gran variedad de problemas concretos.

3.2. Estimacion del proceso de estados

Si se tienen las observaciones y_1, \dots, y_T , y se asume un modelo como el descrito por las ecuaciones (1) y (2), hay varias distribuciones posteriores que puede ser de interes. En relacion al proceso no observable se distinguen tres distribuciones relevantes condicional en los datos observados: filtrada, suavizada y de prediccion.

La **distribucion filtrada** (filtering distribution en ingles) consiste en la distribucion posterior del estado no observable θ_t dada la informacion observada hasta ese mismo momento, es decir $p(\theta_t|y_1, \dots, y_t)$, se tiene una distribucion para cada momento t . La posterior filtrada es de principal interes en aplicaciones secuenciales donde es necesario realizar una nueva estimacion para cada observacion nueva. En el caso de modelos DLM, donde operan las restricciones de normalidad y linealidad, es posible tener soluciones cerradas para $p(\theta_t|y_1, \dots, y_t)$ mediante el Filtro de Kalman (REF??) [es necesario incluir las formulas de recursion para estimar esto??](#)

Una segunda distribucion posterior que puede tener relevancia en aplicaciones es la **distribucion suavizada** (smoothing distribution en ingles) del proceso de estados. La distribucion suavizada consiste en $p(\theta_t|y_1, \dots, y_T)$, es similar a la posterior filtrada con la diferencia que el condicional se realiza sobre todos los datos observados disponibles. La distribucion suavizada es relevante para aplicaciones donde estamos interesados en un estudio retrospectivo y se desea analizar todo el proceso de estados latente. [a nosotros no deberia interesarnos mas smoothing que filtering ???](#)

[estimacion de los parametros de varianza? MLE.. otra forma?](#)

En caso que estemos interesados en predecir un valor futuro del estado se utilizan distribuciones de prediccion $p(\theta_{t+k}|y_1, \dots, y_t)$. [expandir y comentar sobre predictiva de observaciones](#)

[Incluir tratamiento de datos faltantes, caapz que para que se entienda son necesarias incluir formulas para calcular filtering y smoothing](#)

3.3. Un ejemplo simulado

En lo que sigue se ilustra la estimación de DLM mediante un ejemplo simulado de un proceso simple y su estimación mediante la biblioteca `d1m` (CITA??) que permite trabajar con toda la familia de modelos lineales dinámicos. La ecuación (3) muestra el modelo usado para simular datos, en este caso G_t y F_t se definieron constantes con valor 1. El proceso de estado $\{\theta_t\}$ describe un paseo al azar, mientras que podemos considerar a $\{y_t\}$ como una observación ruidosa de $\{\theta_t\}$. El objetivo será recumerar los valores de los parámetros de varianzas y el proceso de estado. *notar que no es estacionario...*

$$\begin{aligned} Y_t &= \theta_t + v_t & v_t &\sim \mathcal{N}(0, 0.1^2) \\ \theta_t &= \theta_{t-1} + w_t & w_t &\sim \mathcal{N}(0, 0.3^2) \end{aligned} \tag{3}$$

El siguiente código de **R** realiza la simulación de un conjunto de datos con el modelo descrito en (3) y su estimación con el Filtro del Kalman.

```
# Obtenemos estados y datos simulados
set.seed(1237); N <- 100

dd <- data_frame(t = seq(0,1, ,N), w = rnorm(N,0,.1), v = rnorm(N,0,.3) ) %>%
  mutate( theta = cumsum(w), y = ts(theta + v) )

# Función para estimar varianzas por Max.Ver
parMLE <- function(par) dlmModPoly(1, dV = par[1], dW = par[2])
modPoly <- dlmMLE(dd$y,parm = c(1,1),build = parMLE)

# Kalman filter y Kalman smoother
mod_filter <- dlmFilter(dd$y, parMLE(modPoly$par))
mod_smooth <- dlmSmooth(mod_filter)
```

En primer lugar, es estimar los parámetros desconocidos del modelo 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 vector de valores iniciales propuesto para los parámetros que se desea estimar, en nuestro caso desconocemos V_t y W_t , la varianza del ruido de evolución y la varianza del ruido de observación. El objeto `modPoly` contiene el modelo con los parámetros estimados.

Una vez estimado el modelo, podemos aplicar el filtro de Kalman mediante (`dlmFilter()`) para obtener estimaciones de la esperanza posterior,

```
## List of 9
## $ y : Time-Series [1:100] from 1 to 100: -0.18756 -0.226 0.05441 -0.00719 -0.21121
## $ mod:List of 10
## ..- attr(*, "class")= chr "dlm"
## $ m : Time-Series [1:101] from 0 to 100: 0 -0.1876 -0.2081 -0.102 -0.0685 ...
## $ U.C:List of 101
## $ D.C: num [1:101, 1] 3162.278 0.289 0.211 0.184 0.172 ...
## $ a : Time-Series [1:100] from 1 to 100: 0 -0.1876 -0.2081 -0.102 -0.0685 ...
## $ U.R:List of 100
## $ D.R: num [1:100, 1] 3162.278 0.309 0.238 0.214 0.204 ...
## $ f : Time-Series [1:100] from 1 to 100: 0 -0.1876 -0.2081 -0.102 -0.0685 ...
## - attr(*, "class")= chr "dlmFiltered"
```

las series de tiempo `m`, `a` y `f` contienen la estimación del estado, la predicción del estado a un paso y la predicción de las observaciones a un paso, respectivamente. Adicionalmente, la función `dlmSmooth()` aplica el suavizado de Kalman y nos devuelve la estimación suavizada del proceso de estados.

La figura 3 muestra los resultados de la estimación modelo, en cada panel se muestran los datos observados (serie gris), el verdadero valor del estado θ_t (serie negra) y una estimación del modelo (serie roja). Los tres tipos de estimación comentados previamente corresponden a cada panel de la figura, en el panel superior se muestra la estimación filtrada que utiliza datos hasta el momento en que se quiere estimar, es decir $m_t = E(\theta_t|y_1, \dots, y_t)$. El panel intermedio muestra la predicción a un paso del proceso de observación, $f_t = E(y_t|y_1, \dots, y_{t-1})$, y finalmente, en el panel inferior se muestra la estimación suavizada del estados, utilizando toda la información disponible para obtenerla, es decir $s_t = E(\theta_t|y_1, \dots, y_N)$

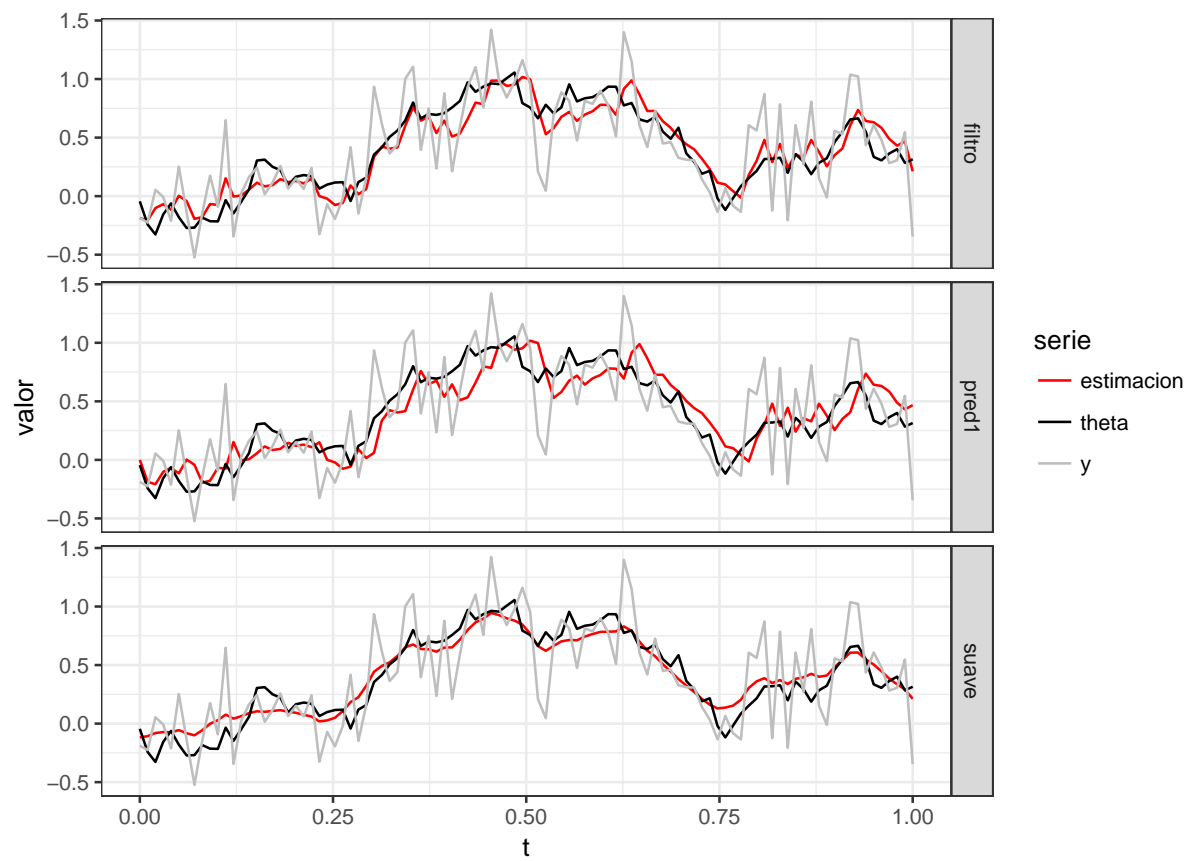


Figura 3: Ejemplo de un proceso de espacio de estado y su estimación.

4. 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 α 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 α 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.

4.1. 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 (??), 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.

5. 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, parm = 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,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 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.

6. Resultados

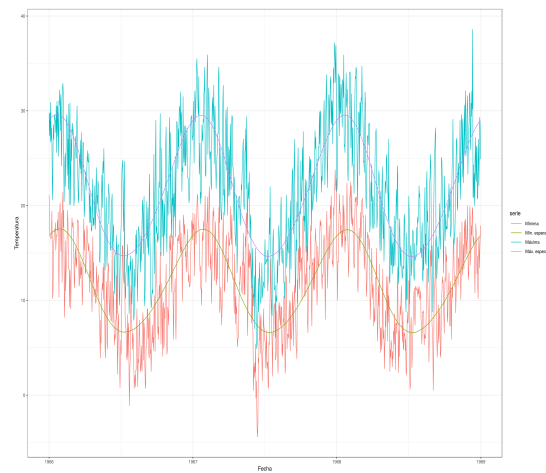


Figura 4: Gráfico de mínimas y máximas del 1966 al 1969, y la estimación del valor esperado del proceso, mediante el filtro de Kalman

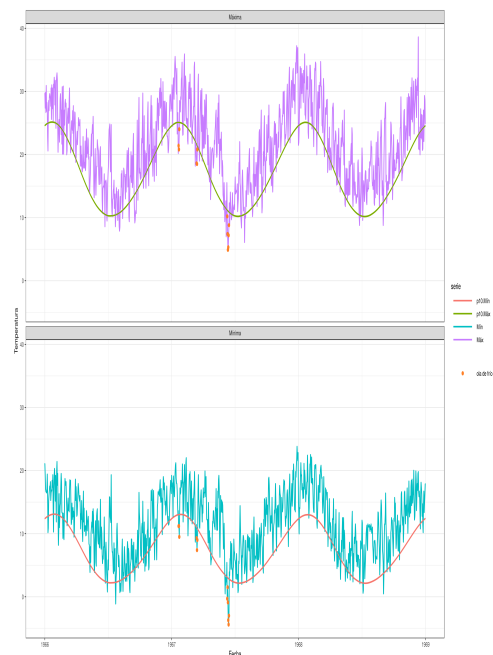


Figura 5: Mínimas y máximas del 1966 al 1969, percentiles 10, y olas de frío detectadas

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
www.iesta.edu.uy
Área Publicaciones

Diciembre, 2017
DT (17/3)