

Un modelo para la evolución de las temperaturas de la Tierra

Agustín ALONSO RODRÍGUEZ

Real Centro Universitario
«Escorial-María Cristina»
San Lorenzo del Escorial

Resumen: Tomando como pretexto la celebración de la Cumbre sobre el Cambio Climático, en Copenhagen, se analiza la serie de temperaturas globales medias, en desviaciones a la media del periodo 1951-1980, elaborada por la NASA, desde 1900 hasta 2008. Un modelo de *espacio de estados* se estima para estos datos, y al final del periodo se hace un ejercicio de simulación para ver la posible evolución de las temperaturas hasta el año 2020.

Abstract: At the occasion of celebrating the Global Summit on Climate Change, in Copenhagen, the series of global temperatures in deviations to the mean of the period 1951-1980, prepared by NASA, is analyzed. A *state space model* is estimated for this series, and, at the end of the period, a simulation exercise is performed to *vision* the evolution of global temperatures to year 2020.

Palabras clave: Anomalías de las temperaturas, dTs, NASA, GISS, modelos de *espacio de estados*, filtro de Kalman, filtrado, suavizado y predicción recursiva, simulación.

Keywords: Mean anomalies of global temperatures, dTs, NASA, GISS, state space models, Kalman filter, filtering, smoothing, forecasting, recursion, simulation.

Sumario:

- I. Introducción.
- II. La observación meteorológica.
- III. El método de análisis estadístico a utilizar.

- IV. Los modelos de espacio de estados.**
- V. Modelos lineales dinámicos.**
- VI. Dos tipos de modelos lineales dinámicos.**
- VII. El modelo para la serie de temperaturas en desviaciones: dTs.**
- VIII. Predicción hasta el año 2020.**
- IX. Conclusión.**
- X. Bibliografía.**

I. INTRODUCCIÓN

La reciente cumbre sobre el Cambio Climático celebrada en Copenhagen, del 7 al 18 de diciembre de 2009, y su aparente fracaso, pone de relieve la complejidad del tema y la diversidad de opiniones al respecto.

En este trabajo se desea utilizar la estadística para analizar la serie de mediciones llevadas a cabo por la NASA, desde 1900 hasta el 2008, presentando al final, mediante simulación, la evolución futura de las temperaturas hasta el año 2020.

II. LA OBSERVACIÓN METEOROLÓGICA

En el *Goddard Institute for Space Studies* (GISS) de la NASA norteamericana se recogen mediciones de temperaturas desde distintas partes del Globo, y a distintos niveles, como mediciones en superficie, aéreas, etc.

El método de análisis actual fue establecido por James Hansen a finales de 1970, diseñando un método para medir los cambios de las temperaturas. Los primeros resultados publicados por Hansen et al (1981) permitían afirmar que un pequeño enfriamiento global, en el hemisferio norte, tuvo lugar tras 1940, si bien con una visión más amplia, y tomando como referencia las fechas entre 1880 y la década de 1970, el calentamiento neto global podría establecerse en torno a los 0.4 grados centígrados.

En la actualidad varias son las fuentes de procedencia de los datos. Se utilizan las mediciones de temperatura del aire de la *Global Historical Climatology Network*, GHCN, (Peterson y Vose, 1997); los datos del *United States Historical Climatology Network*, USHCN, y para la Antártida, los datos provienen del *Scientific Committee on Antarctic Research*, SCAR. El método de análisis utilizado es el establecido por Hansen et al. (1999), con las modificaciones descritas por Hansen et al (2001).

Los datos procedentes de GHNC/USHCN/SCAR son combinados en dos etapas. En la primera, múltiples mediciones del mismo tema, se combinan en un solo dato, y en la segunda etapa, los datos procedentes de estaciones meteorológicas de ciudades y de zonas rurales vecinas se ajustan de forma que su tendencia a largo plazo coincida con la media de las estaciones rurales vecinas, eliminando las mediciones de estaciones urbanas sin estaciones rurales vecinas.

Un índice global de temperaturas, Hansen et al. (1996) se obtiene combinando las mediciones de estaciones meteorológicas que recogen la temperatura de la superficie del mar, tomadas en barcos, y, recientemente, tomadas por satélite. Hay que reseñar también los datos procedentes del grupo británico *British HadISST* y de la *National Oceanic and Atmospheric Administration*, NOAA.

Aunque existen mediciones anteriores al año 1880, sin embargo, es esta fecha la establecida como el punto de arranque para las series oficiales de las mediciones.

Hay que señalar que los datos de las series no utilizan los valores numéricos de las temperaturas, sino lo que se denominan las *anomalías*, *dTs*, que son las desviaciones respecto a la media del periodo 1951-1980. La razón para considerar las *anomalías*, en lugar de los valores numéricos de las temperaturas, se debe al hecho de que las cifras absolutas de las temperaturas varían notablemente de unas regiones geográficas a otras separadas por cortas distancias, mientras que las *anomalías* son capaces de representar regiones geográficas más extensas. Para más detalles, véase, *The Elusive Absolute Surface Air Temperature (SAT)*; www.data.giss.nasa.gov/gistemp/abs_temp.html

Como resultado de estos esfuerzos, existen tablas de *anomalías* medias globales y por regiones. Aquí se utilizará la serie global de medias anuales *dTs*, procedente de las estaciones meteorológicas, desde 1900 a 2008. El lector interesado puede consultar estas tablas de datos en <http://data.giss.nasa.gov/gistemp>.

La representación gráfica de nuestra serie viene dada en la **figura 1**.

A simple vista, a partir de 1935 se nota una tendencia creciente, que vuelve a aparecer a partir de los 80.

El tema central en torno al denominado cambio climático radica en saber si estas pautas de crecimiento, es decir, de calentamiento, son un fenómeno natural o consecuencia de la actividad humana.

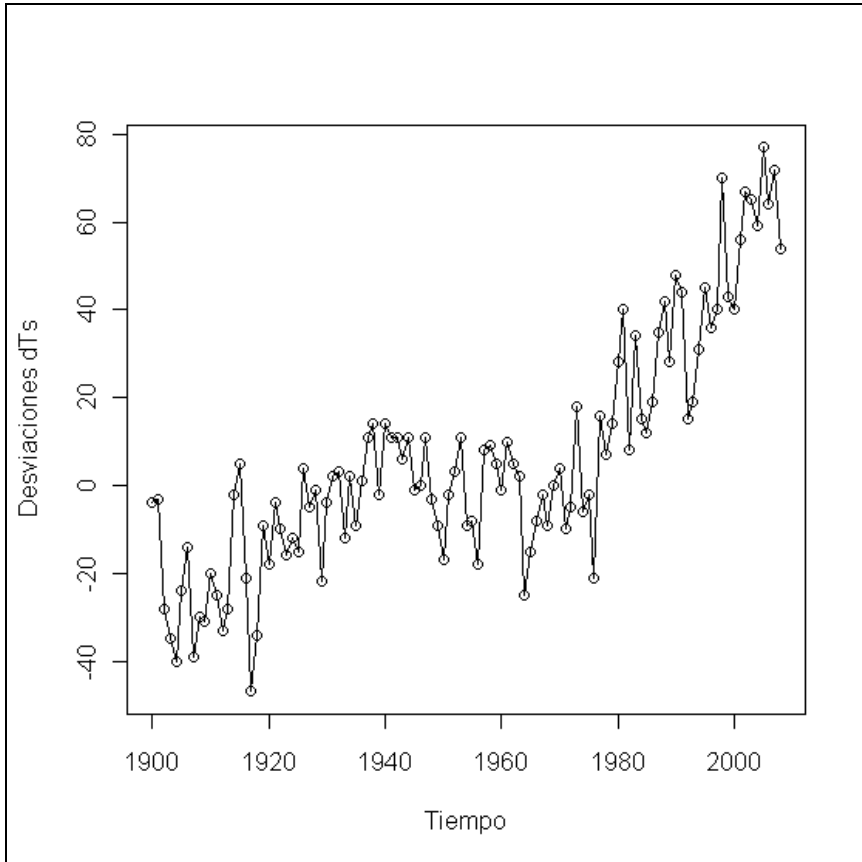


Figura 1

III. EL MÉTODO DE ANÁLISIS ESTADÍSTICO A UTILIZAR

Vamos a analizar la serie representada en la figura 1 utilizando la metodología de los *espacios de estados*, un enfoque de creciente interés en el estudio de las series temporales. Como referencias básicas cabe mencionar Harvey (1989), West y Harrison (1997), Durbin y Koopman (2001) y el más reciente y de carácter introductorio Commandeur y Koopman (2007).

Así mismo hay que mencionar a Akaike (1974), Aoki (1987), Brockwell y Davis (1991, 1997), Hamilton (1994), Ljung y Glad (1994), Venables y Ripley (2002), Peña (2005), Shumway y Stoffer (2006), Lütkepohl (2007), Hyndman et al. (2008), y el más reciente y adaptado al programa estadístico *R*, Petris et al. (2009). La lista empieza a ser interminable.

Los modelos de *espacio de estados* consideran una serie temporal como el output o salida de un sistema dinámico, alterado por perturbaciones aleatorias. Permiten considerar una serie temporal como combinación de varios componentes: tendencia, estacionalidad y elementos autorregresivos. Cabe también mencionar que en la base de estos modelos se encuentra una potente estructura probabilística, que permite un amplio abanico de aplicaciones. Los cálculos necesarios se implementan mediante algoritmos de naturaleza recursiva.

Los problemas inherentes a la estimación y predicción se resuelven calculando de forma recursiva las distribuciones de probabilidad condicionadas de las magnitudes de interés, a partir de la información muestral disponible. En este sentido, el enfoque bayesiano aparece como el marco de referencia para estos modelos.

Los modelos de *espacio de estados* permiten modelizar series univariantes y multivariantes, estacionarias y no estacionarias, con o sin presencia de estacionalidad, de cambios estructurales y de tendencias irregulares.

Para resaltar la diferencia con el enfoque más utilizado de los modelos ARMA, cabe decir que mientras estos modelos requieren series de carácter estacionario, bien en sí mismas, o mediante alguna transformación, los modelos de *espacio de estados* pueden ser utilizados con series no estacionarias sin requerir su transformación en estacionarias.

Los modelos de *espacio de estados* se originaron en el ámbito de la ingeniería, en la década de 1960, si bien los estudios relativos a la predicción de valores futuros, son muy anteriores. Cabe mencionar los trabajos de Wold (1938), Kolmogorov (1941), Wiener (1949). Sin embargo, fueron los trabajos de Kalman (1960) los que transformaron los anteriores enfoques.

Kalman utilizando las representaciones Bode-Shannon de los procesos estocásticos y el método del estado transición de los sistemas dinámicos, diseñó el denominado *filtro de Kalman* (Kalman, 1960, Kalman y Bucy, 1963), aplicable a procesos estocásticos, tanto estacionarios como no estacionarios.

En el ámbito de la ingeniería, el *filtro de Kalman* ha permitido la solución de un amplio abanico de problemas, desde el tratamiento de las órbitas espaciales hasta los problemas oceanográficos, desde la agricultura a los aspectos económicos, a la temática del reconocimiento de voz, etc.

En el ámbito de las estadística, la adaptación de estos métodos ha sido más lenta, si bien merecen destacarse la idea de las variables latentes, la estimación

recursiva, con nombres como el de Thile (1880) y Panckett (1950). Cf. Lauritzen (1981). Una posible razón para este retraso puede deberse al hecho de que las publicaciones sobre el *filtro de Kalman* han estado reducidas a las publicaciones de carácter técnico relativas a la ingeniería, con su específica terminología y problemática.

Las primeras aplicaciones de los modelos de *espacio de estados* en estadística estuvieron orientadas al tratamiento de las series temporales, y son de la década de 1970. Merecen citarse, Akaike (1974), Harrison y Stevens (1976), Aoki (1987), Harvey (1989), West y Harrison (1997). En la actualidad es un campo en continuo desarrollo.

IV. LOS MODELOS DE ESPACIO DE ESTADOS

Consideremos la serie temporal Y_t . Establecer la distribución de probabilidad conjunta de (Y_1, Y_2, \dots, Y_t) no es una operación fácil. De hecho, los supuestos de independencia o intercambiabilidad de las observaciones de la serie temporal son difícilmente defendibles, pues en este caso el papel del tiempo sería irrelevante.

Se dice que Y_t es una cadena de Markov si para todo $t > 1$,

$$\pi(y_t | y_{t:t-1}) = \pi(y_t | y_{t-1})$$

Lo que significa que la información en Y_t , dadas las observaciones hasta el momento $t-1$ es exactamente la misma que la aportada por y_{t-1} en solitario. O, dicho con otras palabras, Y_t e $Y_{1:t-2}$ son independientes dada y_{t-1} .

Tenemos así que para una cadena de Markov de dimensión finita, se puede escribir la distribución conjunta de probabilidad como

$$\pi(y_{1:t}) = \pi(y_1) \times \prod_{j=2}^t \pi(y_j | y_{j-1})$$

Sin embargo, suponer una estructura de cadena de Markov no es siempre apropiado.

Los modelos de *espacio de estados* se basan en el concepto de dependencia de Markov, para, sobre esta base, elaborar modelos más complejos.

En un modelo de *espacio de estados* se supone que existe una cadena de Markov no observable θ_t , denominada el proceso o serie de estados, y una serie temporal Y_t que de forma imprecisa contiene a θ_t .

En el ámbito de la ingeniería, θ_t representa el estado de un sistema físico observable, que genera el resultado u output Y_t . En econometría θ_t jugaría el papel de una variable latente u oculta, que sin embargo tiene una interpretación útil.

En cualquier caso se puede pensar que θ_t es una serie temporal auxiliar, que facilita o permite el establecimiento de la distribución de probabilidad conjunta de las variables observadas en Y_t .

Más formalmente, un modelo de *espacio de estados* consiste en la serie temporal $\theta_t : t = 0, 1, \dots$, de dimensión p , en R^p junto con otra serie temporal en R^m , $Y_t : t = 1, 2, \dots$, que satisfacen los supuestos siguientes:

S. 1: θ_t es una cadena de Markov, y

S. 2: los valores de Y_t , condicionados por θ_t , son independientes, siendo Y_t solo dependiente de θ_t .

Como consecuencia de ambos supuestos, un modelo de *espacio de estados* queda completamente especificado por la distribución inicial $\pi(\theta_0)$ y las funciones de densidad condicionadas $\pi(\theta_t | \theta_{t-1})$ y $\pi(y_t | \theta_t)$ para $t \geq 1$. De hecho, para $\forall t > 0$

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

expresión que permite obtener las distribuciones marginales de interés.

El flujo informativo subyacente en los modelos de *espacio de estados* puede ser representado mediante el siguiente *diagrama de flujos*, **figura 2**

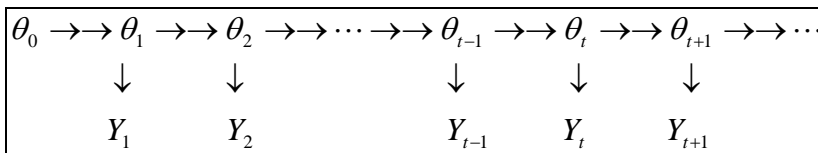


Figura 2

Es decir, si tenemos dos conjuntos de variables aleatorias, A y B, se puede demostrar que son independientes dado un tercer conjunto de variables aleatorias C, si y sólo si C separa A y B, es decir, si cada senda o camino que conecte A con B pasa a través de C. Cf. Cowell et al. (1999).

En la figura 2, Y_t y $(\theta_{0:t-1}, Y_{1:t-1})$ son independientes dado θ_t . Basta con comprobar en la figura 2 que para conectar Y_t con cualquier Y_s , ($s < t$) hay que hacerlo a través de θ_t , luego $\{\theta_t\}$ separa $\{\theta_{0:t-1}, Y_{1:t-1}\}$ e $\{Y_t\}$. En consecuencia

$$\pi(y_t | \theta_{0:t-1}, y_{1:t-1}) = \pi(y_t | \theta_t)$$

De manera semejante se puede demostrar que θ_t y $(\theta_{0:t-2}, Y_{1:t-1})$ son independientes dado θ_{t-1} , lo que permite escribir

$$\pi(\theta_t | \theta_{0:t-1}, y_{1:t-1}) = \pi(\theta_t | \theta_{t-1})$$

Cuando los estados son variables aleatorias de carácter discreto, los modelos de *espacio de estados* se conocen también como modelos ocultos de Markov (hidden Markov models).

V. MODELOS LINEALES DINÁMICOS

La primera clase importante de modelos de *espacio de estados* son los modelos lineales gaussianos, también llamados modelos lineales dinámicos (MLD).

Un MLD viene especificado por una distribución de probabilidad *a priori* normal para el vector de estados de dimensión p , en $t = 0$

$$\theta_0 \sim N_p(m_0, C_0) \quad (1.a)$$

junto al par de ecuaciones siguientes para $t \geq 1$

$$Y_t = F_t \theta_t + v_t; v_t \sim N_m(0, V_t) \quad (1.b)$$

$$\theta_t = G_t \theta_{t-1} + w_t; w_t \sim N_p(0, W_t) \quad (1.c)$$

Siendo G_t y F_t matrices de valores conocidos, de orden $p \times p$ y $m \times p$, respectivamente, y v_t y w_t secuencias independientes de vectores gaussianos, de elementos independientes, con media cero y matrices de varianzas-covarianzas V_t y W_t .

La ecuación (1.b) se conoce como la *ecuación de observaciones*, y la ecuación (1.c) es la *ecuación de estados* o *ecuación del sistema*. Se supone que θ_0 es independiente de v_t y de w_t .

Se puede demostrar que un MLD satisface los supuestos S.1 y S.2, con

$$Y_t | \theta_t \sim N(F_t \theta_t, V_t), \text{ y}$$

$$\theta_t | \theta_{t-1} \sim N(G_t \theta_{t-1}, W_t)$$

VI. DOS TIPOS DE MODELOS LINEALES DINÁMICOS

1.- El modelo paseo aleatorio con ruido blanco.

El modelo MLD más sencillo para una serie temporal univariante $Y_t, t = 1, 2, \dots$ recibe el nombre de *paseo aleatorio con ruido blanco* (random walk plus noise), que viene definido mediante las ecuaciones

$$Y_t = \mu_t + v_t; v_t \sim N(0, V)$$

$$\mu_t = \mu_{t-1} + w_t; w_t \sim N(0, W)$$

siendo las secuencias de variables aleatorias v_t y w_t independientes en sí y entre sí.

Al comparar con el modelo en (1.b) y (1.c), vemos que

$$m = p = 1; \theta_t = \mu_t; F_t = G_t = 1.$$

Desde el punto de vista de las aplicaciones, este modelo es apropiado para series temporales que no muestren una clara tendencia o variaciones estacionales.

Las observaciones en la serie $Y_t, t = 1, 2, \dots$, son modelizadas como observaciones aleatorias en torno al nivel marcado por μ_t , que a su vez se comporta en el tiempo como un *paseo aleatorio* o *random walk*. Es oportuno destacar que por ser μ_t un *random walk*, la serie no es estacionaria, y, por tanto, un MLD puede modelizar esta serie o comportamiento, lo que no puede hacerse con los modelos ARMA, que requieren series estacionarias o transformadas hasta conseguir que lo sean.

Dado el papel decisivo del nivel μ_t , este modelo se conoce también como el *modelo de nivel local* (local level model).

2.- El modelo lineal creciente.

Este segundo modelo añade al anterior una tendencia lineal. Tiene la misma *ecuación de observaciones* pero añade una pendiente variante en el tiempo a la dinámica de μ_t . Es el modelo *local lineal con tendencia* (linear growth model):

$$\begin{aligned} Y_t &= \mu_t + v_t; & v_t &\sim N(0, V) \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + w_{t,1}; & w_{t,1} &\sim N(0, \sigma_\mu^2) \\ \beta_t &= \beta_{t-1} + w_{t,2}; & w_{t,2} &\sim N(0, \sigma_\beta^2) \end{aligned}$$

El modelo tiene ahora una pendiente variante con el tiempo, que afecta a la dinámica de μ_t . Se supone que v_t , $w_{t,1}$ y $w_{t,2}$ son perturbaciones o errores incorrelacionados.

Si comparamos este modelo con la formulación en (1.b) y con (1.c), tenemos

$$\theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}; G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; W = \begin{pmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix} \text{ y } F = (1 \ 0)$$

las varianzas σ_μ^2 , σ_β^2 pueden ser iguales a cero. De ser $\sigma_\beta^2 = 0$ tendríamos una velocidad constante en la dinámica del sistema. De ser las matrices G_t , F_t y las matrices de varianzas-covarianzas V_t y W_t constantes, el modelo se dirían que es invariante en el tiempo. Por esta razón, los modelos ARMA gaussianos pueden deducirse como casos especiales de los MLD, siendo modelos equivalentes en el caso de no variabilidad en el tiempo.

Es posible añadir al anterior modelo componentes estacionales, que amplían la aplicabilidad del MLD. Sin embargo, como nuestra serie de temperaturas no presenta este tipo de elementos, detenemos aquí la exposición teórica, para pasar a la modelización de nuestra serie.

VII. EL MODELO PARA LA SERIE DE TEMPERATURAS EN DESVIACIONES: DTS

Con ayuda del paquete *dIm*, escrito en R por Giovanni Petris, instalable desde CRAN (the Comprehensive R Archive Network), en Viena, procedemos al análisis de nuestra serie.

Siguiendo las indicaciones que acompañan a las tablas de temperaturas en términos de anomalías, dTs , dividimos por cien los valores de la serie de medias anuales globales, de enero a diciembre, para convertir las observaciones en grados centígrados, grados Celsius. En la **figura 3**, viene representada la nueva serie, desde 1900 a 2008.

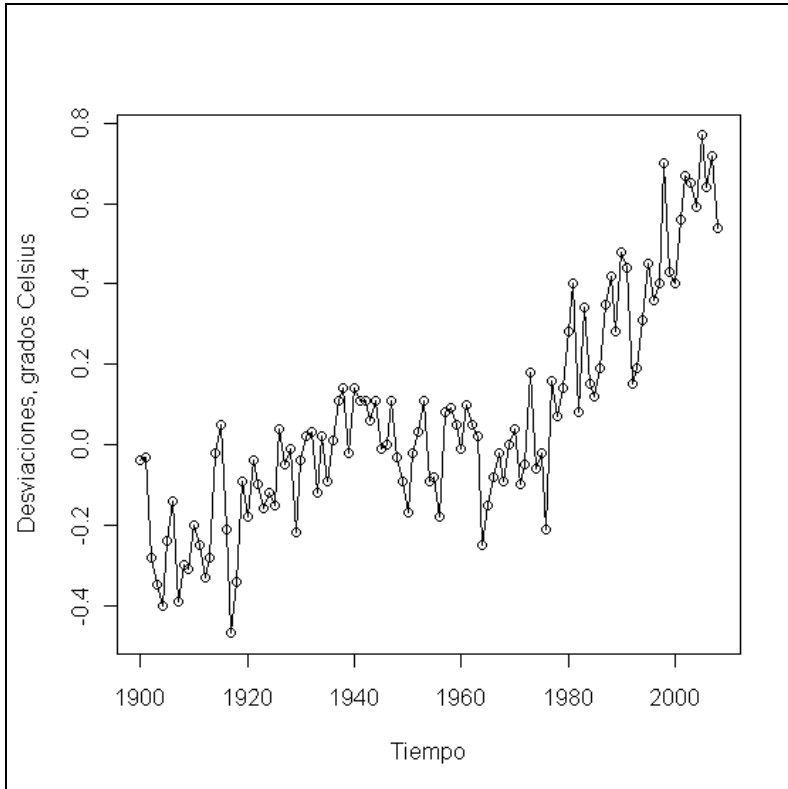


Figura 3

Visualmente, la serie muestra una ligera tendencia creciente, por lo que un modelo de *espacio de estados* del tipo de los modelos lineales dinámicos parece apropiado. En concreto, el modelo lineal con tendencia es el elegido.

$$Y_t = \mu_t + v_t; \quad v_t \sim N(0, V)$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{t,1}; \quad w_{t,1} \sim N(0, \sigma_\mu^2)$$

$$\beta_t = \beta_{t-1} + w_{t,2}; \quad w_{t,2} \sim N(0, \sigma_\beta^2)$$

Con ayuda de la función *dlnMLE*, las estimaciones máximo-verosímiles de las varianzas de las perturbaciones son:

$$V = \hat{\sigma}_v^2 = 0.01146983$$

$$W = \begin{pmatrix} 0.001078437 & 0 \\ 0 & 0.000002273492 \end{pmatrix}$$

sobre estas estimaciones se construye el modelo.

En los modelos de *espacio de estados* la principal labor consiste en inferir la evolución de los *no-observables estados*, y también predecir las futuras observaciones de la serie temporal, sobre la base de las observaciones disponibles de la serie.

Para estimar el *vector de estados* hay que computar las densidades condicionadas $\pi(\theta_s | y_{1:t})$. Y cabe distinguir tres situaciones:

- a) si $s = t$, se habla de filtrado (*filtering*) de la serie;
- b) si $s < t$, se habla de suavizado (*smoothing*) de la serie, y
- c) si $s > t$, se habla de predicción.

En el caso de *filtrado*, se supone que las observaciones llegan secuencialmente en el tiempo. Es el caso más frecuente, y, en concreto, el nuestro.

En estas circunstancias, se estima el *vector de estados* sobre la base de las observaciones hasta el momento t , hasta el momento presente, el “ahora”, y se actualiza la estimación y se predicen nuevos estados al llegar nuevas observaciones. Por tanto, en *filtering* se computa la densidad $\pi(\theta_t | y_{1:t})$. El *filtro de Kalman* permite las actualizaciones, es decir, el pasar de $\pi(\theta_t | y_{1:t})$ a $\pi(\theta_{t+1} | y_{1:t+1})$.

En el caso de *smoothing*, también denominado *análisis retrospectivo*, se trata de estimar la secuencia de estados en $t = 1, 2, \dots, t$, dadas las observaciones y_1, y_2, \dots, y_t . Ocurre a veces que se tienen observaciones de una serie temporal en un determinado periodo de tiempo, y se desea tener una visión retrospectiva del comportamiento del sistema subyacente en la serie temporal. Acudiendo al ámbito de la economía, el investigador puede tener los datos de una serie, por ejemplo, referentes al consumo privado en un cierto número de años, y desea tener una visión retrospectiva del comportamiento del sistema socio-económico subyacente.

La situación se resuelve computando la distribución condicionada de $\theta_{1:t}$ dados los valores de $y_{1:t}$. Al igual que en la situación anterior, se aplica un algoritmo recursivo como es el *filtro de Kalman*.

Por lo que respecta al tercer caso, *forecasting*, quizás el principal objetivo del análisis, se trata de predecir el valor futuro Y_{t+1} sobre la base de la información presente en $y_{1:t}$. Para ello, primero se estima el valor de θ_{t+1} de los estados y luego se calcula la predicción de Y_{t+1} . La densidad de la predicción un-periodo-adelante es $\pi(\theta_{t+1} | y_{1:t})$, a partir de la densidad de θ_t , y desde ella se obtiene la densidad de predicción un-periodo-adelante $\pi(y_{t+1} | y_{1:t})$.

Se puede hacer también predicción k -periodos hacia adelante, tras estimar el *vector de estados* θ_{t+k} , $k \geq 1$, y a partir de él, predecir Y_{t+k} . La predicción del *vector de estados* se obtiene computando la densidad de predicción $\pi(\theta_{t+k} | y_{1:t})$, y a partir de ésta, computando $\pi(y_{t+k} | y_{1:t})$. Por supuesto que cuanto mayor sea el valor de k , más imprecisa o insegura se vuelve la predicción. Cabe, sin embargo, cuantificar la incertidumbre a partir de $\pi(Y_{t+1} | y_{1:t})$ de forma recursiva. En concreto, la media condicionada $E(Y_{t+1} | y_{1:t})$ suministra la predicción puntual un-periodo-adelante del valor Y_{t+1} , que minimiza el error cuadrático medio. En cuanto función de k , $E(Y_{t+k} | y_{1:t})$ recibe el nombre de *función de predicción*.

Tras este *excursus*, volvamos a nuestra serie de temperaturas.

La función `dlmFilter` del paquete `dlm` aplica el *filtro de Kalman*, obteniendo el *vector de estados*, junto con sus matrices de varianzas y covarianzas. A la vez calcula las predicciones un-periodo-adelante del *vector de estados*, permitiendo obtener los residuos del modelo, como un primer paso para evaluar la idoneidad estadística del modelo estimado.

A partir de las estimaciones de las varianzas iniciales, se construye el modelo, y se aplica el *filtro de Kalman* mediante las instrucciones, en nuestro caso,

```
modFilt = dlmFilter(tempC, modelo)
```

siendo `tempC` la serie de anomalías de temperaturas en grados Celsius, y `modelo` el modelo establecido según la sintaxis del paquete `dlm`. Los resultados quedan archivados en `modFilt`.

El resumen de la aplicación de `dlmFilter` lo da la instrucción

```
summary(modFilt)
```

que genera el siguiente *output*:

	Length	Class	Mode
y	109	ts	numeric
mod	10	d1m	list
m	220	mts	numeric
U.C	110	-none-	list
D.C	220	-none-	numeric
a	218	mts	numeric
U.R	109	-none-	list
D.R	218	-none-	numeric
f	109	ts	numeric

nuestro vector de estados es m, en este caso una matriz de dos columnas. En la primera columna está el filtrado o *vector de estados* del nivel de este modelo, y en la segunda, el filtrado de la pendiente del modelo. Encabezando ambas columnas, están las estimaciones del nivel y de la pendiente. En nuestro caso, con los respectivos siguientes valores:

0.66242538 y 0.01507321

La representación de la serie de anomalías de las temperaturas, junto con el *vector de estados*, el filtrado del nivel, viene recogida en la **figura 4**

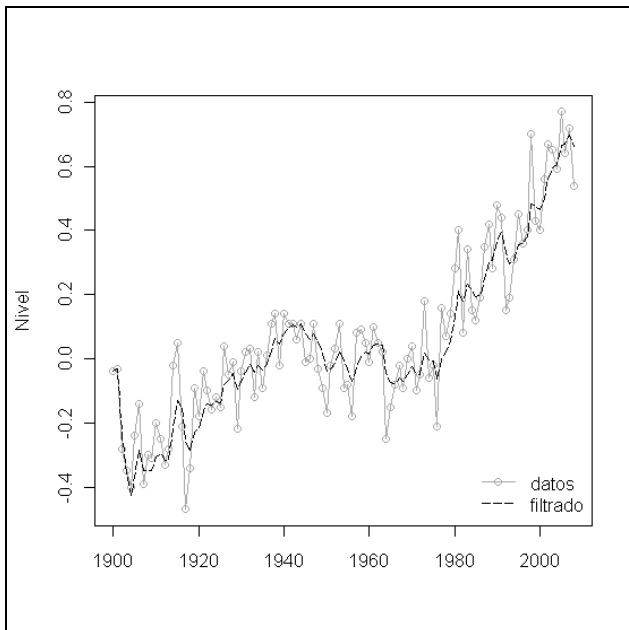


Figura 4

La tendencia está compuesta por el nivel y su pendiente. En la siguiente figura 5, aparece el filtrado de la pendiente. Dado que la cambiante pendiente representa la tasa de cambio de la serie, es susceptible de la siguiente interpretación.

Cuando es positiva, la tendencia de la serie es creciente, y cuando es negativa, la tendencia decrece. Así mismo, cuando la pendiente es positiva y creciente, el crecimiento de la tendencia de la serie es pronunciado, mientras que el crecimiento de la tendencia de la serie es menos pronunciado cuando la pendiente es positiva pero decreciente. Y viceversa, cuando la pendiente es negativa y decreciente, la tendencia decrece de forma pronunciada, mientras que el decrecimiento se debilita cuando la pendiente es negativa, pero creciente.

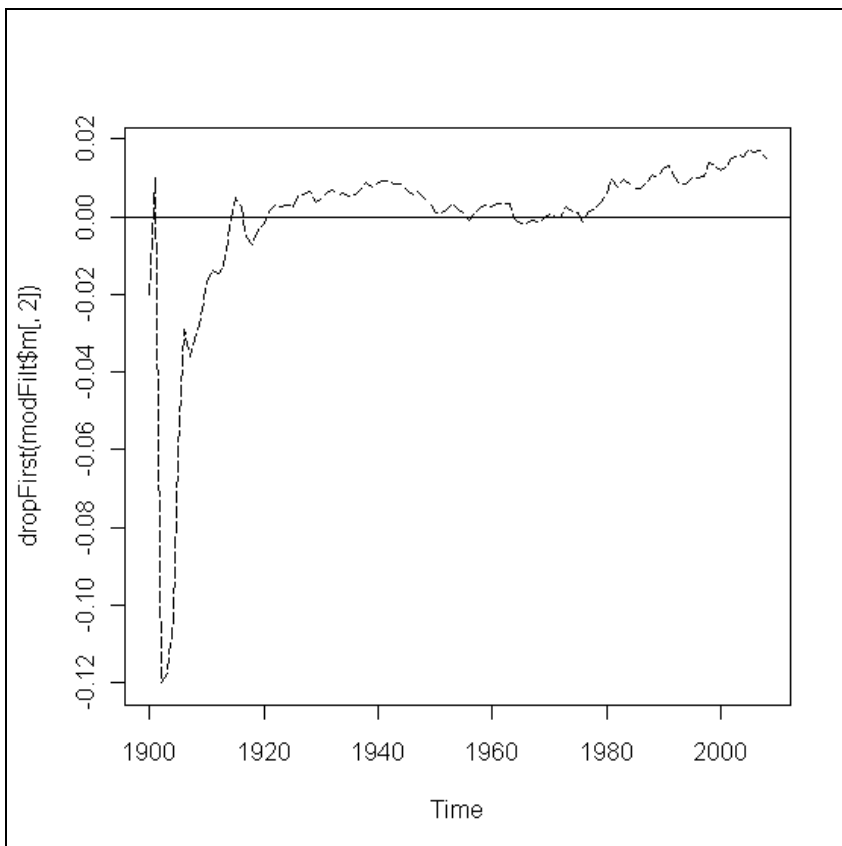


Figura 5

Los anteriores comentarios sobre esta **figura 5**, reflejan la evolución de la serie de anomalías de las temperaturas de la figura anterior.

El paso siguiente consiste en la validación del modelo. La función *dlmFilter* permite también la obtención de los residuos, en nuestro caso, mediante la instrucción

```
residuals(modFilt, sd=FALSE)
```

y los residuos pueden ser representados y analizados, permitiendo la consiguiente validación del modelo.

Para el modelo estimado, los residuos resultan ser aceptablemente normales, con media cero e independientes entre sí. Además, el *periodograma acumulado* no detecta la presencia de componentes periódicos. En definitiva, podemos considerar estadísticamente válido el modelo estimado.

Habiendo obtenido un modelo estadísticamente válido, podemos pasar a la fase de su utilización para predicción.

A partir del 2008, se desea hacer la predicción de la serie de anomalías de las temperaturas hasta el 2020. Hay que comenzar diciendo que los datos de la serie entran en la distribución de probabilidad predictiva a través de la media de la distribución del filtrado, en el momento de llegada de la última observación. La función *dlmForecast* computa las medias y varianzas de la distribución predictiva de las observaciones y estados. La función permite también generar una muestra aleatoria de futuros estados y observaciones.

Las instrucciones que siguen permiten obtener las predicciones de la serie de temperaturas anómalas *dTs* hasta el 2020, junto con ocho sendas simuladas de futuros valores, generadas de acuerdo con la distribución de predicción.

```
set.seed(123)
modFore = dlmForecast(modFilt, nAhead=12, sampleNew=8)
names(modFore)
attach(modFore)

plot(window(tempC, start=c(1980,1)), type="o",
      xlim=c(1980,2020), ylim=c(0.00,1.00), xlab="",
      ylab="Temperaturas dTs en Grados Celsius")
invisible(lapply(newObs, function(x)
  lines(x, col="darkgrey", type="o", pch=4)))
lines(f, type="o", lwd=2, pch=16)
abline(v=mean(c(time(f)[1], time(tempC)[length(tempC)])),
      lty="dashed")
```

El resultado viene recogido en la **figura 6**.

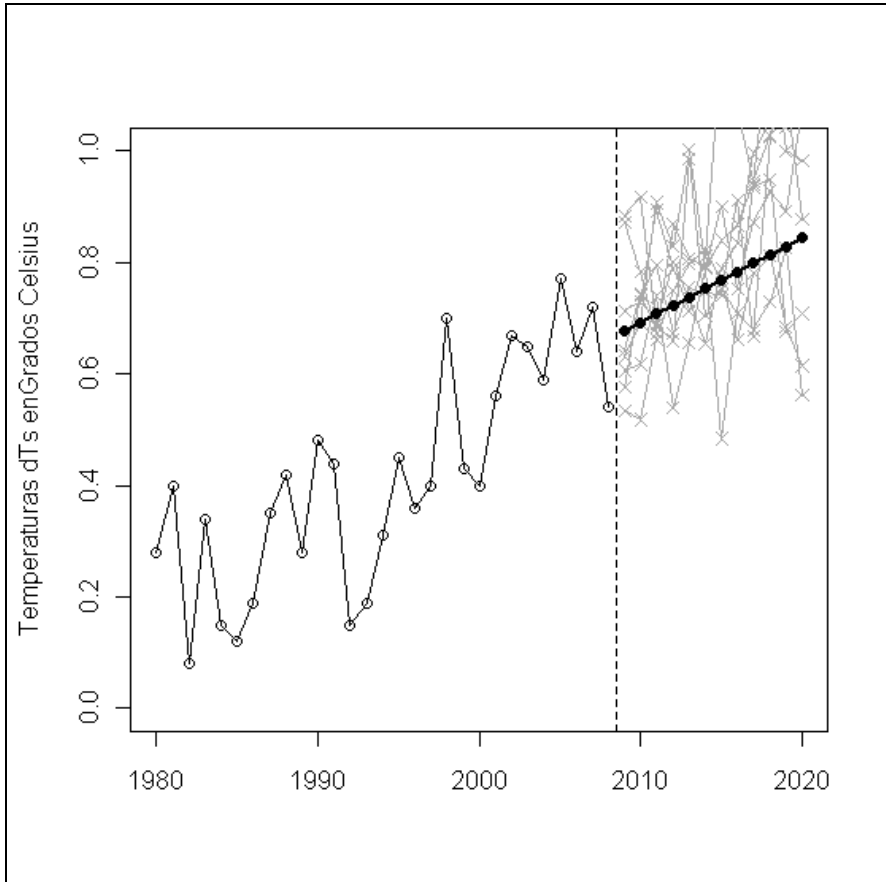


Figura 6

Resaltar que, como indica el eje de ordenadas, el aumento que presenta la serie de predicciones dTs viene expresado en décimas de grados centígrados.

Para mayor clarificación de la figura 6, se reproducen los valores numéricos de las predicciones, con las bandas de confianza del 95%. En la columna central están las predicciones dTs .

	lim.inf	predic.	lim.sup.
2009	0.4678173	0.6774986	0.9273493
2010	0.4732926	0.6925718	0.9538591
2011	0.4784618	0.7076450	0.9807337
2012	0.4833371	0.7227182	1.0079584
2013	0.4879300	0.7377914	1.0355197
2014	0.4922509	0.7528647	1.0634051

	lim.inf	predic.	lim.sup.
2015	0.4963095	0.7679379	1.0916030
2016	0.5001147	0.7830111	1.1201029
2017	0.5036746	0.7980843	1.1488951
2018	0.5069969	0.8131575	1.1779704
2019	0.5100885	0.8282307	1.2073205
2020	0.5129560	0.8433039	1.2369378

Valores que vienen representados en la **figura 7**

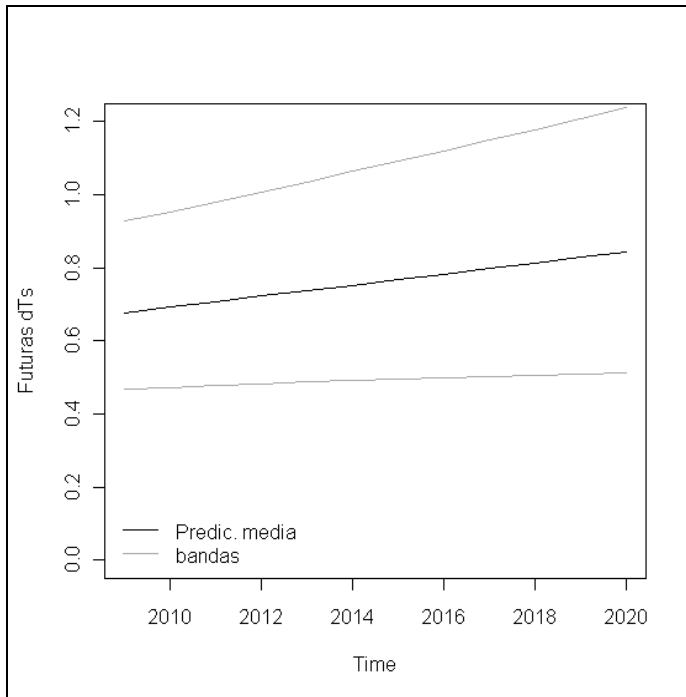


Figura 7

Si examinamos más en detalle las predicciones, los siguientes comentarios parecen en orden.

El rango de variación de las predicciones va desde 0.6774986 hasta 0.8433039, grados dTs. La diferencia entre estos extremos es: 0.1658054, o sea, una diferencia de 0.17 grados dTs.

El valor medio de las predicciones es 0.7604, y su error estándar: 0.01568870. El valor de la mediana es 0.7604. Esta coincidencia pone de relieve el carácter gaussiano o normal de la distribución de las predicciones.

En la **figura 8**, se presentan el *boxplot* y la función de densidad aproximada de la distribución de las predicciones.

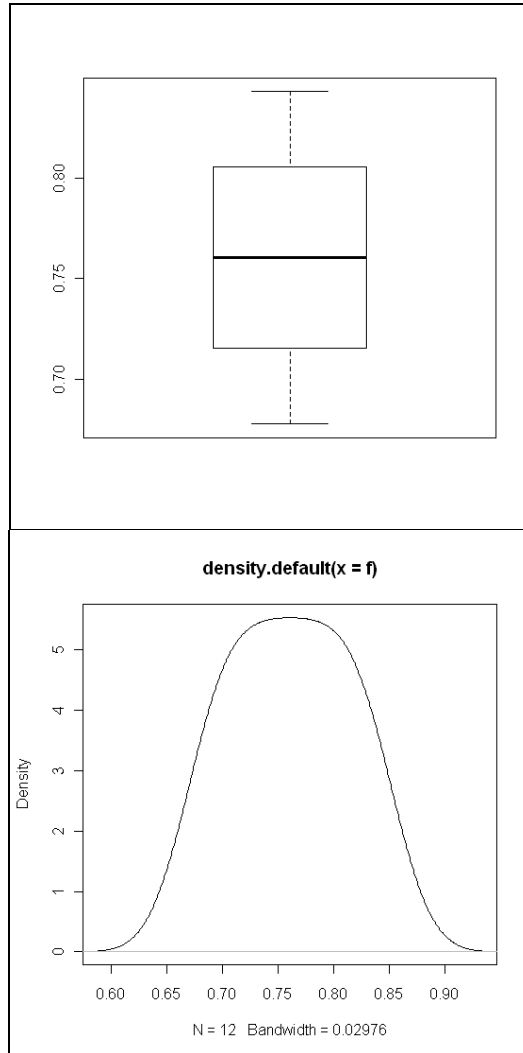


Figura 8

Por último, el contraste de normalidad de Shapiro-Wilk, da como resultado

Shapiro-Wilk normality test

```
data: f  
W = 0.9669, p-value = 0.8757
```

IX. CONCLUSIÓN

A manera de conclusión y dados los resultados, podemos decir que la elección de un modelo de *espacio de estados*, el modelo lineal dinámico y con tendencia aleatoria, parece apropiado para explicar estadísticamente la evolución de las temperaturas dTs.

Así mismo, la predicción muestra una tendencia creciente en esas temperaturas hasta el año 2020. Un crecimiento moderado, cuantificable, aproximadamente, en 0.17 grados dTs, y siempre que permanezcan las actuales circunstancias climáticas.

X. BIBLIOGRAFÍA

- AKAIKE, H., “Markovian representation of stochastic processes and its applications to the analysis of autoregressive moving average processes”, en *Annals of the Institute of Statistical Mathematics*, 24 (1974) 363-387.
- AOKI, M., *State Space Modeling of Time Series*, Springer, Berlin 1987.
- BROCKWELL, P. J. Y DAVIS, R. A., *Time Series: Theory and Methods*, 2ª ed., Springer, New York 1991.
- BROCKWELL, P. J. Y DAVIS, R. A., *Introduction to Time Series and Forecasting*, Springer, New York 1996.
- COMMANDEUR, J. F. Y KOOPMAN, S., *An Introduction to State Space Time Series Analysis*, Oxford University Press, Oxford 2007.
- COWELL, R., DAWID, P.; LAURITZEN, S., y SPIEGELHALTER, D., *Probabilistic networks and expert systems*, Springer, New York 1999.
- DURBIN, J. Y KOOPMAN, S., *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford 2001.
- HAMILTON, J. D., *Time Series Analysis*, Princeton University Press, Princeton 1994.
- HANSEN, J.; JOHNSON, D.; LACIS, A.; LEBEDEFF, S.; LEE, P.; RIND, D., y RUSSELL, G., “Climate impact of increasing atmospheric carbon dioxide”, en *Science*, 213 (1981) 957-966.

- HANSEN, J.; RUEDY, R.; SATO, M., AND REYNOLDS, R., “Global surface air temperature in 1995: Return to pre-Pinatubo level”, en *Geophys. Res. Lett.*, 23 (1996) 1665-1668.
- HANSEN, J.; RUEDY, R.; GLASCOE, J., y SATO, M., “GISS analysis surface temperature change”, en *J. of Geophys. Res.*, 104 (1999) 30997-31022.
- HANSEN, J.E.; RUEDY, R.; SATO, M.; IMHOFF, M.; LAWRENCE, W.; EASTERLING, D.; PETERSON, T., y KARL, T., “A closer look at United States and global surface temperature change”, en *J. Geophys. Res.*, 106 (2001) 23947-23963.
- HARRISON, P., y STEVENS, C., “Bayesian Forecasting (with discussion)”, en *Journal of the Royal Statistical Society*, series B, 38 (1976) 205-247.
- HARVEY, A., *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge 1989.
- HYNDMAN, R. J.; KOEHLER, A. B.; ORD, J. K., y SNYDER, R. D., *Forecasting with Exponential Smoothing, the State Space Approach*, Springer, Berlin-Heidelberg 2008.
- KALMAN, R., “A new approach to linear filtering and prediction problems”, en *Journal of Basic Engineering, Transactions of the ASME*, Series D, 82 (1960) 35-45.
- KALMAN, R., y BUCY, R., “New results in linear filtering and prediction theory”, en *Journal of Basic Engineering, Transactions of the ASME*, Series D, 83 (1963) 95-108.
- KOLMOGOROV, A., “Interpolation and extrapolation of stationary time series”, en *Bulletin de l'Académie des Sciences de l'URSS*, Serie Mathematique, 5 (1941) 3-14.
- LAURITZEN, S., “Time Series Analysis in 1880: a discussion of contributions made by T. N. Thiele”, en *International Statistical Review*, 49 (1981) 319-331.
- LJUNG, L., y GLAD, T., *Modeling of Dynamic Systems*, Prentice Hall, Englewood Cliffs 1994.

- LÜTKEPOHL, H., *New Introduction to Multiple Time Series Analysis*, Springer, Berlin- Heidelberg 2007.
- PEÑA, D., *Análisis de series temporales*, Alianza Editorial, Madrid 2005.
- PETERSON, T. C., y VOSE, R. S., “An overview of the Global Historical Climatology Network temperature database”, en *Bulletin of the American Meteorological Society*, 78 (1997) 2837-2849.
- PETRIS, G.; PETRONE, S., y CAMPAGNOLI, P., *Dynamic Linear Models with R*, Springer, New York 2009.
- R, <http://www.r-project.org/>
- SHUMWAY, R. H., y STOFFER, D. S., *Time Series Analysis and Its Applications, With R Examples*, 2a. ed., Springer, New York 2006.
- VENABLES, W. N., y RIPLEY, B. D., *Modern Applied Statistics with S*, 4ª. Ed., Springer, New York 2002.
- WEST, M., y HARRISON, J., *Bayesian Forecasting and Dynamic Models*, 2ª. ed., Springer, New York 1997.
- WIENER, N., *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, New York 1949.
- WOLD, H., *A Study in the Analysis of Stationary Time Series*, Almqvist-Wiksell, Uppsala 1938.

