azti tecnalia

2016

23(1)

Un programa ejecutable para el análisis de series de datos mediante filtros Kolmogorov-Zurbenko

Manuel González
Almudena Fontán

Revista de Investigación Marina

M. González, A. Fontán, 2016. Un programa ejecutable para el análisis de series de datos mediante filtros Kolmogorov-Zurbenko. Revista de Investigación Marina, AZTI, 23(1): 1-8

La '*Revista de Investigación Marina'* de AZTI edita y publica investigaciones y datos originales resultado de la Unidad de Investigación Marina de AZTI. Las propuestas de publicación deben ser enviadas al siguiente correo electrónico aborja@azti.es. Un comité de selección revisará las propuestas y sugerirá los cambios pertinentes antes de su aceptación definitiva.

Edición: 1.ª Enero 2016

ISSN: 1988-818X
Unidad de Investigación Marina
Internet: www.azti.es
Edita: Unidad de Investigación Marina de AZTI
Herrera Kaia, Portualdea
20110 Pasaia
Foto portada: © Alex Iturrate

Un programa ejecutable para el análisis de series de datos mediante filtros Kolmogorov- Zurbenko

Manuel González[1], Almudena Fontán[1]

Resumen

Se facilitan dos subrutinas en Fortran para el cálculo de los filtros Kolmogorov-Zurbenko (KZ) y Kolmogorov-Zurbenko Adaptativo (KZA) y un programa ejecutable preparado para su uso. Además, se presentan algunas aplicaciones a series de datos con el fin de que el usuario se familiarice con el uso del programa y pueda comprobar los desarrollos que realice. Por una parte, se muestran casos en los que se han introducido cambios bruscos en los valores medios, así como variabilidad lineal a largo plazo y por otra, la aplicación de estos filtros a dos series de datos de información pública: la serie del Índice de Precios al Consumo (IPC) y la Encuesta de Población Activa (EPA) de España.

Palabras clave: Series de datos, filtro KZ, filtro KZA, Fortran

Abstract

Two subroutines in Fortran for calculating the Kolmogorov-Zurbenko (KZ) and Kolmogorov-Zurbenko Adaptive (KZA) filters are provided, together with an executable program ready for application. In addition, some applications to data sets are provided, which would allow users get familiar with the program and check if it is correctly run. On one hand, several cases are shown in which abrupt changes have been introduced in the mean values, plus long-term lineal variability. On the other hand, the filters are applied to two series of public information data in Spain: the Consumer Price Index (IPC) and the Labour Force Survey (EPA).

Keywords: time series, KZ filter, KZA filter, Fortran

Introducción

Los filtros Kolmogorov-Zurbenko (KZ) y Kolmogorov-Zurbenko Adaptativo (KZA) (Zurbenko *et al.,* 1996) son dos métodos sencillos para detectar períodos con valores medios constantes e identificar cambios (debidos a variabilidad natural, cambios en los sistemas de medida, en las condiciones ambientales, en los marcos regulatorios y legislativos, etc.) así como variabilidad a largo plazo (e.g. Civerolo *et al.*, 2001; Wise y Comrie, 2005; Chaves *et al.,* 2008, González *et al.,* 2013; Solaun *et al.,* 2013; Henneman *et al.,* 2015). El filtro KZA fue diseñado específicamente para detectar cambios bruscos en series con variabilidad periódica y ruido notable y cabe mencionar que existe a disposición un código en R (Yang y Zurbenko, 2010; https://cran.r-project.org/web/packages/kza/kza.pdf).

En un trabajo publicado en la Revista de Investigación Marina se presentó la aplicación de estos filtros a series océano-meteorológicas del golfo de Vizcaya (González *et al.,* 2011). Tras varios trabajos sobre el análisis de series de datos mediante estos métodos (e.g. ICES, 2011; Revilla *et al*., 2012) se han recibido algunas peticiones del código a las que se trata de dar respuesta en esta comunicación.

Métodos

KZ es básicamente una media móvil de ventana fija (el semiancho de la ventana de promediado se denota por la letra q) que se repite un cierto número de iteraciones (it), mientras que el KZA es una media móvil iterada pero, de ventana variable en la que también se ha de fijar un valor máximo del semiancho de la ventana de promediado (q) y un número de iteraciones (it).

Según Zurbenko *et al.* (1996) siendo una serie temporal de datos, la media móvil centrada (MMC) de semiancho q se define como:

$$MMC^q[x(t)] = \frac{1}{2q+1}\sum_{j=-q}^{q} x(t+j). \quad (1)$$

La primera iteración de semiancho q del filtro Kolmogorov-Zurbenko (KZ), denotada por $KZ_q^1[x(t)]$, es la media móvil centrada:

$$KZ_q^1[x(t)] = MMC^q[x(t)] = \frac{1}{2q+1}\sum_{j=-q}^{q} x(t+j) \quad (2)$$

[1] *AZTI, Marine Research Division, Herrera Kaia, Portualdea z/g, 20110 Pasaia, Gipuzkoa, Spain*
Email: mgonzalez@azti.es; afontan@azti.es

y, recursivamente, la iteración *k-ésima* de semiancho q del filtro, $KZ_q^k[x(t)]$, se define como la media móvil centrada de semiancho q de la iteración *k-1*:

$$KZ_q^k[x(t)] = \frac{1}{2q+1}\sum_{j=-q}^{q} KZ_q^{k-1}[x(t+j)];\ k \geq 2 \qquad (3)$$

Para calcular el filtro de Kolmogorov-Zurbenko Adaptativo, primero se aplica el filtro KZ de parámetros q y k a la serie de datos original:

$$z(t) = KZ_q^k[x(t)]. \qquad (4)$$

Para disminuir el suavizado en aquellos puntos donde es posible que haya un cambio brusco, el filtro adaptativo define un ancho de ventana a izquierda $q_H(t)$ y a derecha $q_T(t)$ de la siguiente forma:

$$q_H(t) = \begin{cases} q & \text{si} \quad D'(t) < 0 \\ q \cdot f(t) & \text{si} \quad D'(t) \geq 0 \end{cases} \qquad (5)$$

$$q_T(t) = \begin{cases} q & \text{si} \quad D'(t) > 0 \\ q \cdot f(t) & \text{si} \quad D'(t) \leq 0 \end{cases} \qquad (6)$$

donde:

$$D(t) = |z(t+q) - z(t-q)|, \qquad (7)$$

$$D'(t) = D(t+1) - D(t) \qquad (8)$$

$$f(t) = 1 - \frac{D(t)}{\max[D(t)]}. \qquad (9)$$

La primera iteración de semiancho q del filtro KZA, denotada por $KZA_q^1[x(t)]$ se define como:

$$KZA_q^1[x(t)] = \frac{1}{q_H + q_T + 1}\sum_{j=-q_H}^{q_T} z(t+j). \qquad (10)$$

Y, recursivamente, la iteración k de semiancho q del filtro KZA viene dada por:

$$KZA_q^k[x(t)] = \frac{1}{q_H + q_T + 1}\sum_{j=-q_H}^{q_T} KZA_q^{k-1}[x(t+j)]. \qquad (11)$$

El filtro $KZA_q^k(x)$ sobre una serie permite retirar de la misma, aproximadamente, los períodos inferiores a $q\sqrt{k}$ (Eskridge *et al.*, 1997) y permite establecer una primera aproximación a los valores del tamaño de la ventana y del número de iteraciones a usar.

Ejemplos con datos simulados

Como primer caso de ejemplo se ha generado una serie de 3000 datos. La señal tiene una parte aleatoria uniforme en el intervalo [-1,1]. A esta señal aleatoria entre el dato 1000 y el 2000 se le ha añadido un valor de 0,4 unidades. Se ha filtrado la señal total con un semiancho de ventana de 100 y 4 iteraciones. En la imagen superior de la Figura 1 puede verse la señal original y en la imagen inferior de la Figura 1 se muestra el salto introducido en la señal aleatoria (línea roja), el resultado del filtro KZ aplicado a la señal (línea negra) y el resultado del filtro KZA (línea gris). El tiempo de CPU requerido es de 0,03 s en un PC de 2,6 GHz y 4 Gb de RAM.

En la Figura 2 puede verse un segundo ejemplo, con 36.500 datos. En este caso la señal está formada por una onda sinusoidal de período 365, más una señal aleatoria de distribución normal [0,1] y un salto de 0,5 en el dato 10.000 con un incremento lineal de pendiente 1/16.000 hasta el salto 18.000, posteriormente hay un salto negativo de 0,75 y una tendencia lineal de disminución de 1/64.000 (imagen superior de la Figura 2). En la imagen inferior de la Figura 2 pueden verse los saltos bruscos y las líneas de tendencia de la serie (línea roja), el resultado del filtro de la señal con el KZ (línea negra) y el resultado del KZA (línea gris). Ambos filtros se han usado con q = 500 datos y 4 iteraciones, para filtrar la variabilidad inferior a 1.000 datos, valor superior a la periodicidad de 365 datos presente $(q\sqrt{k})$ señal. El tiempo de CPU requerido es de 0,08 s en un PC de 2,6 GHz y 4 Gb de RAM.

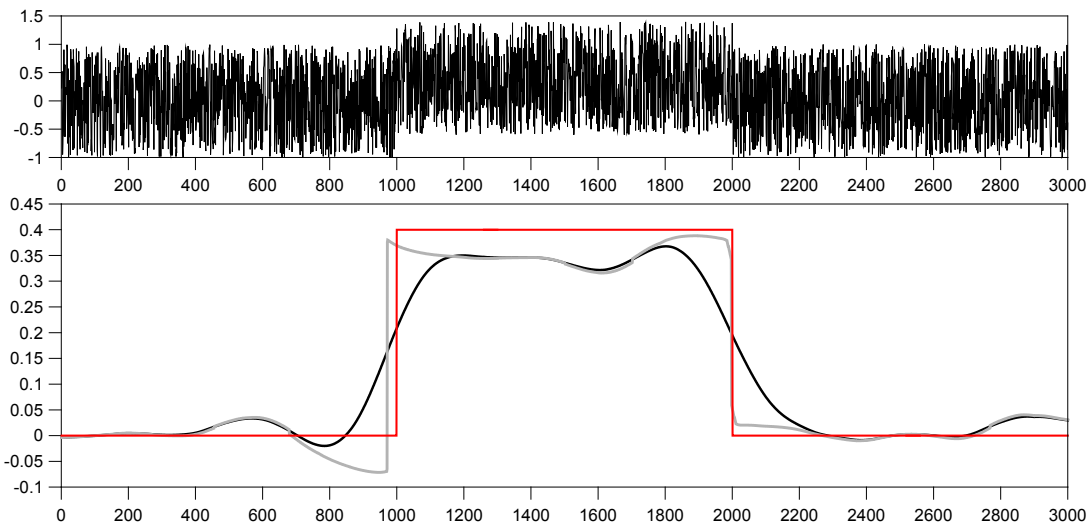

Figura 1. Imagen superior: señal aleatoria uniforme [-1,1] y un salto de 0,4 unidades entre el dato 1000 y el 2000. Imagen inferior: la línea roja es el salto introducido en la serie, la línea negra es el resultado del filtro KZ aplicado a la serie de datos con q=100 y 4 iteraciones y la línea gris es el KZA con q=100 y 4 iteraciones.

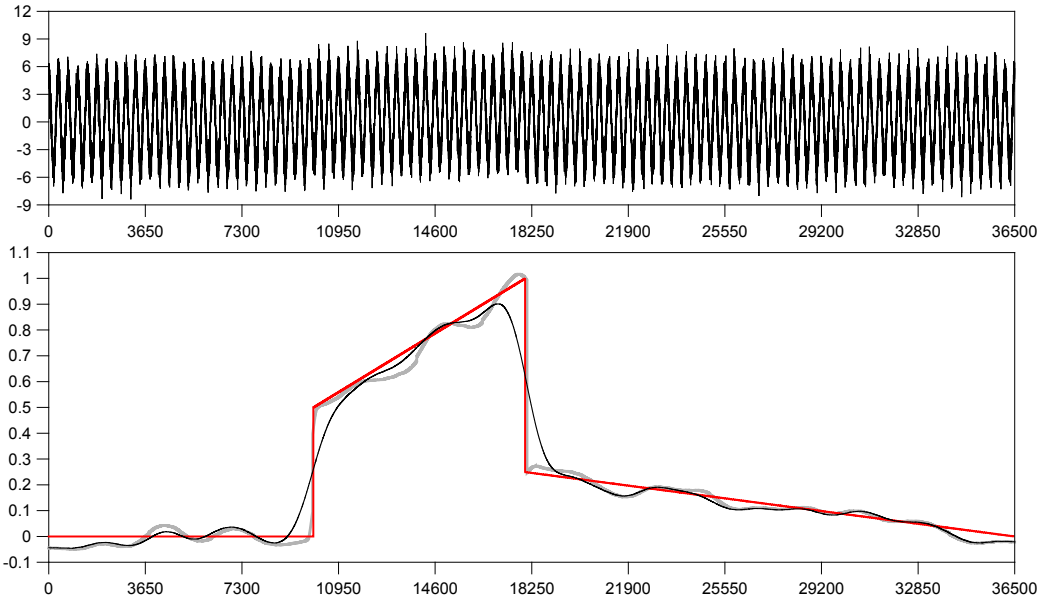

Figura 2. Imagen superior: onda sinusoidal de período 365 más una señal aleatoria normal [0,1] y un salto de 0,5 en el dato 10.000 con un incremento lineal de pendiente 1/16.000 hasta el salto 18.000, posteriormente hay un salto negativo de 0,75 y una tendencia línea de disminución de 1/64.000. Imagen inferior: la línea roja son los saltos y las líneas de tendencia introducidas en la serie, la línea negra es el resultado del filtro KZ aplicado a la serie de datos con q=500 y 4 iteraciones y la línea gris es el KZA con q=500 y *4* iteraciones.

Ejemplos con series de datos

Se presenta a continuación la aplicación de ambos filtros a la serie de datos mensual del Índice de Precios al Consumo (IPC) en España entre 1962 y 2015 (www.aragon.es/iaest). En este caso para obtener una aproximación a la variabilidad decadal de esta serie se han aplicado los filtros KZ y KZA con q=60 y 4 iteraciones. Los resultados pueden verse en la Figura 3 (la línea roja es la serie de datos, el resultado del filtro KZ es la línea negra y la línea gris es el resultado del KZA).

Dado que la metodología de cálculo del IPC se revisa y actualiza desde el año 2001 cada 5 años, hasta la fecha se ha modificado en 2006 (Índice de Precios al Consumo, Metodología, 2006) y 2011 (Índice de Precios al Consumo, Metodología, 2011) se ha analizado la serie de datos con los filtros KZ y KZA desde el año 2000. En la Figura 4 puede verse el IPC mensual desde el año 2000 hasta 2015 y los resultados de los filtros KZ y KZA con q= 30 meses (2,5 años) y 4 iteraciones para filtrar la variabilidad de período inferior a 5 años (valor superior a la variabilidad estacional de la serie).

En la Figura 5 pueden verse los datos trimestrales de población ocupada en España (Encuesta de Población Activa, EPA, www.aragon.es/iaest) desde el año 2001 hasta 2015 (línea roja) La línea negra y la línea gris son el resultado del filtro KZ y KZA con semiancho de ventana 4 datos (1 año) y 4 iteraciones, respectivamente (con estos valores se retira la variabilidad de período inferior a 2 años). Entre 2001 y 2015 el cambio metodológico más importante se produjo en 2005 (Encuesta de Población Activa. Metodología 2005, 2008).

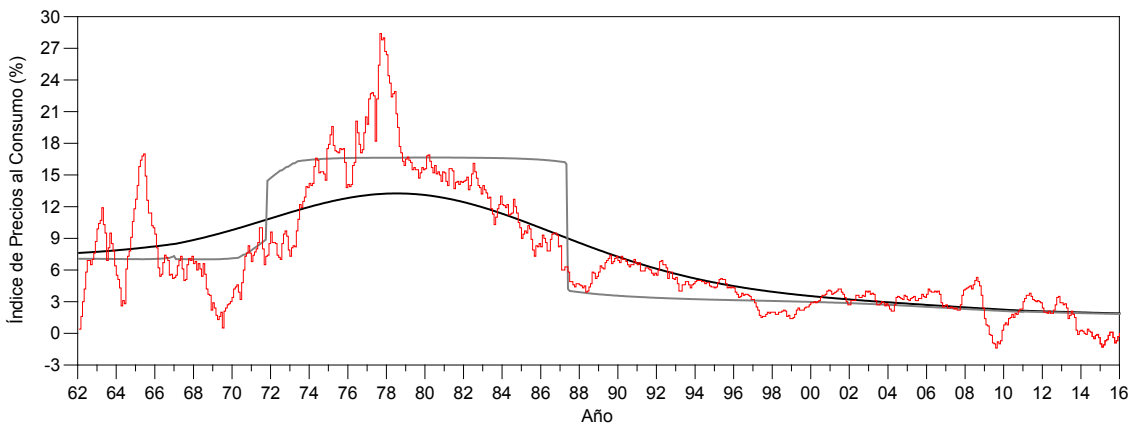

Figura 3. La línea roja son los datos mensuales del Índice de Precios al Consumo en España desde 1962 a 2015 (www.aragon.es/iaest). Las líneas de color negro y gris son el resultado de filtrar con KZ y KZA, con q=60 datos y 4 iteraciones, respectivamente.

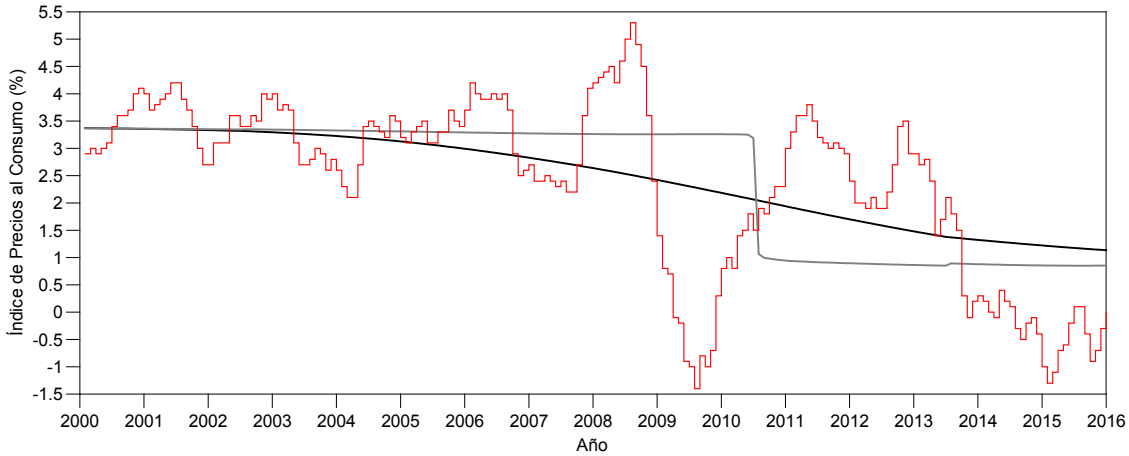

Figura 4. La línea roja son los datos mensuales del Índice de Precios al Consumo en España desde 2000 a 2015 (www.aragon.es/iaest). Las líneas de color negro y gris son el resultado de filtrar con KZ y KZA, con q=30 datos y 4 iteraciones, respectivamente.

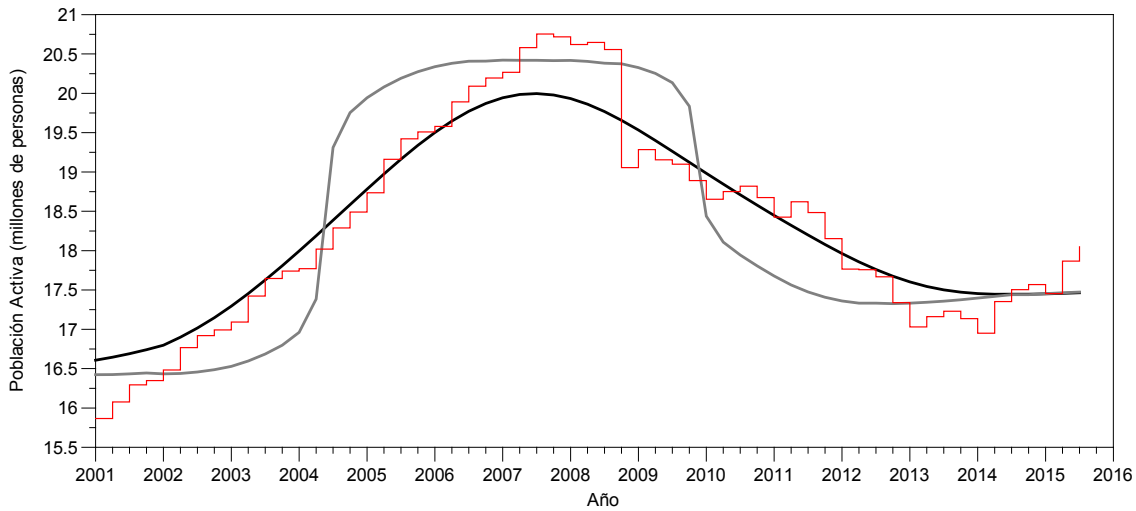

Figura 5. La línea roja son los datos trimestrales de la población ocupada en España (Encuesta de Población Activa) desde 2001 a 2015 (www.aragon.es/iaest). Las líneas de color negro y gris son el resultado de filtrar con KZ y KZA, con q=4 datos y 4 iteraciones, respectivamente.

Instrucciones de uso

Para la aplicación de los filtros KZ y KZA los datos deben de ser equiespaciados, ordenados de más antiguo a más recientes y sin huecos. El programa admite hasta 1 millón de datos. Para un ejemplo con 1 millón de datos, un tamaño de ventana de promediado de 1.000 datos y 4 iteraciones, el código desarrollado requiere 1,5 s de computación en un PC de 2,6 GHz y 4 Gb de RAM; en el caso de una ventana de 5.000 datos y 4 iteraciones se necesitan unos 7 s de computación. El programa ejecutable que se pone a disposición está compilado con la versión Compaq Visual Fortran Professional Edition 6.6.0 y funciona en Windows, no requiriendo tener instalado Fortran.

Cada dato debe de estar en una línea de un fichero en formato ASCII, con fecha y valor (separado por espacios o coma), se sugiere consultar los ejemplos que se ponen a disposición.

El nombre del fichero con los datos, los nombres de los ficheros de resultados de KZ y KZA y los valores de q, it se dan al programa a través del fichero "Datos.dat"

El fichero Datos.dat, el fichero donde se encuentran los datos a filtrar (Ejemplo_1.dat) y el código ejecutable que se facilita deben copiarse en una carpeta y realizarse la ejecución del código. Los resultados (KZ.dat, KZA.dat) son ficheros que el programa genera y en los que se encuentran los valores filtrados. Con el fin de facilitar el uso del código se facilitan los ficheros de datos y de resultados correspondientes a los ejemplos de la Figura 1 y de la Figura 2.

Ejemplo_1.dat	(Nombre del fichero donde están los datos)
KZ.dat	(Nombre del fichero de resultados del filtro KZ)
KZA.dat	(Nombre del fichero de resultados del filtro KZA)
100	(q, semiancho de la ventana de promediado)
4	(it, número de iteraciones)

Figura 6. Información del fichero Datos.dat

En el Anexo a este documento se facilita el código fuente para su adaptación a otros sistemas operativos, su inclusión en programas más complejos o su adaptación a series de datos con presencia de huecos.

Agradecimientos

Los autores desean agradecer las aportaciones de los revisores: Dra. Leire Ibaibarriaga y Dr. Guillem Chust, que han mejorado notablemente la calidad del manuscrito.

Esta es la contribución número 749 de AZTI (Unidad de Investigación Marina).

Bibliografía

Chaves, L. F., Kaneko, A., Taleo, G., Pascual, M., & Wilson, M. L., 2008. Malaria transmission pattern resilience to climatic variability is mediated by insecticide-treated nets. *Malar J*, 7(100), 10-1186.

Civerolo, K.L., Brankov, E., Rao, S.T., Zurbenko, I., 2001. Assessing the impact of the acid deposition control program. *Atmospheric Environment*, 35: 4135-4148.

Encuesta de Población Activa. Metodología 2005. Descripción de la encuesta, definiciones e instrucciones para la cumplimentación del cuestionario. Madrid, 2008

Eskridge, R.E., Ku, J.Y., Rao, S.T., Porter, P.S., Zurbenko, I.G., 1997. Separating different scales of motion in times series of meteorological variables. *Bulletin American Meteorological Society*, 78: 1473-1483.

González, M., Fontán, A., Esnaola, G., & Collins, M., 2013. Abrupt changes, multidecadal variability and long-term trends in sea surface temperature and sea level datasets within the southeastern Bay of Biscay. *Journal of Marine Systems*, 109: S144-S152.

González, M., Fontán, A., Esnaola, G., Valencia, V., 2011. Variaciones multidecadales notables en la temperatura atmosférica, temperatura superficial y nivel del mar en el sudeste del golfo de Vizcaya, detectadas mediante filtros "Kolmogorov-Zurbenko". *Revista de Investigación Marina*, 17(8): 1-15.

Henneman, L. R., Holmes, H. A., Mulholland, J. A., & Russell, A. G., 2015. Meteorological detrending of primary and secondary pollutant concentrations: Method application and evaluation using long-term (2000–2012) data in Atlanta. *Atmospheric Environment*, 119: 201-210.

ICES. 2011. Report of the Working Group on Phytoplankton and Microbial Ecology (WGPME), 21–24 March 2011, Galway, Ireland. ICES CM 2011/SSGEF:04. 32 pp.

Índice de Precios de Consumo. Base 2006. Metodología, Madrid. Subdirección General de Estadísticas de Precios y Presupuestos Familiares.

Índice de Precios de Consumo. Base 2011. Metodología. Subdirección General de Estadísticas Coyunturales y de Precios. Madrid, mayo 2012. INE. Instituto Nacional de Estadística.

Revilla M., Borja Á., Chust, G., Fontán, A., Franco, J., González, M., Novoa, S., Sagarminaga, Y., Valencia, V. 2012. Estudio de la clorofila, elemento clave para la Estrategia Marina Europea y la Directiva Marco del Agua. Informe elaborado por AZTI-Tecnalia para la Agencia Vasca del Agua. 102 pp.

Solaun, O., Rodríguez, J. G., Borja, A., González, M., & Saiz-Salinas, J. I., 2013. Biomonitoring of metals under the water framework directive: Detecting temporal trends and abrupt changes, in relation to the removal of pollution sources. *Marine Pollution Bulletin*, 67(1): 26-35.

Wise, E. K., & Comrie, A. C., 2005. Extending the Kolmogorov–Zurbenko filter: application to ozone, particulate matter, and meteorological trends. *Journal of the Air & Waste Management Association*, 55(8): 1208-1216.

Yang, W. and Zurbenko, I., 2010. Nonstationarity. Wiley Interdisciplinary Reviews: *Computational Statistics*, 2(1): 107–115.

Zurbenko, I.G., Porter, P.S., Rao, S.T., Ku, J.Y., Gui, R., Eskridge, R.E., 1996. Detecting discontinuities in time series of upper air data: development and demonstration of an adaptive filter technique. *Journal of Climate*, 9: 3548–3560.

ANEXO

```
    subroutine KZ(nmax,Ndata,xdata,iq,nveces,y)   ! y=KZ (xdata,iq,nvceces)
    implicit none
    integer*4 nmax,nveces,iveces,ndata,iq,jj,ii,kk
    real*8 xdata(nmax),x(nmax),y(nmax),b,c,d
    b=1.d00/(1.d0+iq)
    d=1.d00/(2.d0*iq)
    c=1.d00/(2.d0*iq+1)
                                                x=xdata
    do 10 iveces=1,nveces
y(1)=sum(x(1:iq+1))*b
    do 1 ii=2,iq+1
                        jj=iq+ii
                        y(ii)=(y(ii-1)*(jj-1)+x(jj))/(jj)
    1           end do
                do 2 ii=iq+2,ndata-iq
                        jj=ii-1
                        y(ii)=y(jj)+(x(ii+iq)-x(jj-iq))*c
    2           end do
                        y(ndata-iq+1)=sum(x(ndata-2*iq+1:ndata))*d
                        jj=ndata+iq+1
                do 4 ii=ndata-iq+2,ndata
                        kk=ii-1
                        y(ii)=(y(kk)*(jj-kk)-x(kk-iq))/(jj-ii)
    4           end do
                                                                x=y
    10      end do
                                    return
    end                 ¡ Subroutine KZ
```

```
      subroutine KZA(nmax,Np,iq,iqmin,nveces,X,yt,sigma)   !yt=KZA(x,iq,nveces)
      implicit none
      integer*4 nmax,np,iq,nveces,iveces,iqmin,ii,jj,j1,j2,j3,j4
      integer*4 qt(nmax),qh(nmax)
      real*8  X(nmax),yt(nmax),sigma(nmax),yy(2*iq+1),s,ymed,sigma_t,a
                                          call KZ(NMAX,Np,X,iq,nveces,yt)
      sigma(1:np)=0.d0
      qh(1:np)=iq
      qt(1:np)=iq
      sigma(iq+1:np-iq)=abs(yt(2*iq+1:np)-yt(1:np-2*iq))
      ymed=1.0d00/maxval(sigma(1:np))
      do 1 ii=1,np
            s=sigma(ii+1)-sigma(ii)
            jj=int(iq-sigma(ii)*ymed*iq)
                  if(s.ge.0.0d00) qh(ii)=max(iqmin,jj)
                  if(s.le.0.0d00) qt(ii)=max(iqmin,jj)
                  if((ii-qt(ii)).lt.1) qt(ii)=ii-1
                  if((ii+qh(ii)).gt.np) qh(ii)=np-ii
1     end do
            sigma(1:np)=X(1:np)
      do 10 iveces=1,nveces
            yt(1)=sum(sigma(1-qt(1):1+qh(1)))/(qh(1)+qt(1)+1)
            j1=1-qt(1)
            j2=1+qh(1)
            do 2 ii=2,np
                  j3=ii-qt(ii)
                  j4=ii+qh(ii)
                  a=1.0d00/(qh(ii)+qt(ii)+1)
                        if(j3.ge.j1.and.j4.ge.j2) then
                              s=sum(sigma(j2+1:j4))-sum(sigma(j1:j3-1))
                              jj=ii-1
                              yt(ii)=a*(yt(jj)*(qh(jj)+qt(jj)+1)+s)
                        else
                              yt(ii)=a*sum(sigma(j3:j4))
                        end if
                                                      j1=j3
                                                      j2=j4
2           end do
                                                      sigma(1:np)=yt(1:np)
10    end do
      sigma(1:np)=X(1:np)-yt(1:np)-(sum(X(1:np))-sum(yt(1:np)))/(np)
      a=sqrt(dot_product(sigma(1:np),sigma(1:np))/(np-1.d0))
      sigma_t=(a*np)/(2.d0*iq*dsqrt(1.d0*nveces))
      sigma_t=1.d0/sigma_t
      do 3 ii=1,np
       jj=1+qt(ii)+qh(ii)
       ymed=sum(yt(ii-qt(ii):ii+qh(ii)))/(jj)
       yy(1:jj)=yt(ii-qt(ii):ii+qh(ii))-ymed
       sigma(ii)=sigma_t*sqrt(dot_product(yy(1:jj),yy(1:jj))/(jj-1))
3     end do
                  return
      end                                 ¡Subroutine KZA
```

www.azti.es

azti tecnalia

Txatxarramendi ugartea z/g
48395 Sukarrieta (Bizkaia)
Tel.: +34 94 657 40 00
Fax: +34 94 657 25 55

Herrera Kaia, Portualdea z/g
20110 Pasaia (Gipuzkoa)

Parque Tecnológico de Bizkaia
Astondo bidea, Edificio 609.
48160 Derio (Bizkaia)