

Capítulo 5

Aplicación. Datos de meteorología e inmisión

En este capítulo, realizaremos un análisis exploratorio de la dirección de viento tomada en la estación meteorológica, A Mourela, durante el año 2010. Para ello realizaremos un análisis descriptivo de esta variable (sección 4.2.) así como realizaremos una estimación no paramétrica tipo núcleo de la densidad (sección 4.3.) y estimaremos el modelo de Möbius de series de tiempo (sección 4.5.).

Además, estudiaremos que tipo de relación hay entre el SO_2 y la dirección de viento en las estaciones de medida automáticas B1, B2, C9, F2 Y G2, construyendo un modelo de regresión circular-lineal (sección 4.4).

Cabe mencionar que tanto la estación meteorológica a Mourela como las estaciones de medida automáticas pertenecen al Sistema de Control Suplementario de la Contaminación Atmosférica de la U.P.T. de As Pontes propiedad de Endesa Generación S.A.

Por otra parte, notemos que para llevar a cabo este capítulo fue necesaria realizar una perturbación en los datos, la cual está descrita en la sección 4.1. Por tanto, en este capítulo no vamos a trabajar con los datos minutales recogidos en la estación meteorológica y en las estaciones de medida automáticas B1, B2, C9, F2, G2; sino que vamos a trabajar con los datos horarios perturbados.

5.1 Datos perturbados

Al analizar los datos obtenidos por los dispositivos de dirección se observó que había valores repetidos tanto en la dirección de viento como en las concentraciones de SO_2 . Esto pueden ser debido a que estos dispositivos no presentan una precisión tan alta como para evitarlo. La aparición de medidas repetidas va a ser un problema a la hora de implementar ciertas técnicas, como la validación-cruzada para escoger el parámetro de suavizado.

Con el fin de solucionar dicho problema, haremos una perturbación en ambas variables. La perturbación en el caso lineal, viene dada por [1], una muestra artificial de la concentración de SO_2 viene dada por:

$$\tilde{X}_i = X_i + b\epsilon_i$$

, dónde X_i denota los valores observados de SO_2 , $b = 1.03\hat{\sigma}n^{-1/3}$ y $\epsilon_i, i = 1, \dots, n$ son variables aleatorias independientes e idénticamente distribuidas siguiendo el núcleo de Epanechnikov en $(-\sqrt{5}, \sqrt{5})$. $\hat{\sigma}$ es un estimador robusto de la varianza, el cual puede ser obtenido usando el rango intercuartílico estandarizado. [1] muestra que esta elección de b para los datos perturbados permite una estimación consistente de la función de distribución, obteniendo un error cuadrático medio con la misma magnitud que el de la función empírica de distribución acumulada.

En el caso circular, la perturbación se realiza de forma similar al caso lineal. La muestra artificial de la dirección del viento es

$$\tilde{\theta}_i = \theta_i + d\epsilon_i$$

, dónde θ_i denota los valores observados de la dirección de viento y $\epsilon_i, i = 1, \dots, n$ son variables aleatorias independientes generadas de la distribución von Mises con $\mu = 0$ y $\kappa = 1$. En ([4]) se toma $d = n^{-1/5}$ basándose en los resultados de ([8]) para la estimación tipo núcleo de la distribución multivariante. Esta perturbación resuelve el problema de los valores repetidos y no afecta a la función de distribución subyacente.

5.2 Análisis de los datos de dirección de viento en A Mourela

En esta sección, vamos a analizar los datos de viento recogidos durante el año 2010 por la estación meteorológica A Mourela situada en As Pontes (Figura 5.1), que van a ser usados en todo el capítulo excepto en la sección 4.4. Cabe mencionar que esta estación meteorológica pertenece al Sistema de Control Suplementario de la Contaminación Atmosférica de la U.P.T. de As Pontes.

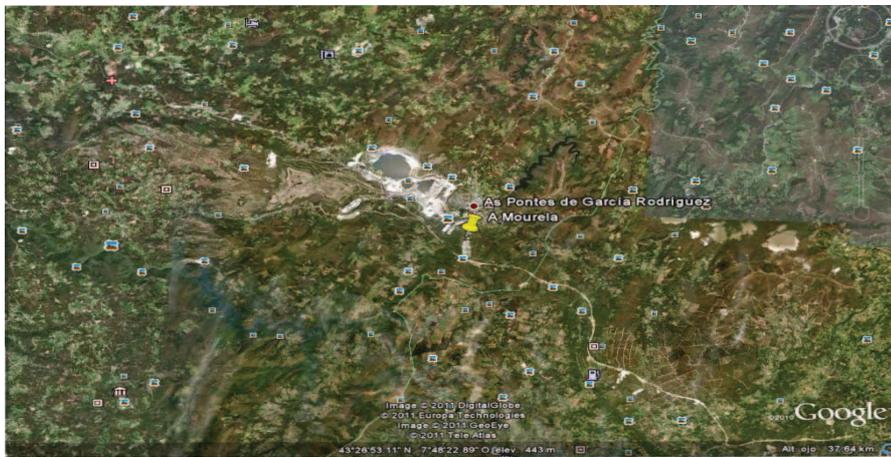


Figura 5.1: Estación Meteorológica A Mourela

Esta estación recoge datos minutales pero, en este caso, trabajaremos con las medias horarias con el objetivo de reducir la cantidad de datos. Aunque a la hora de presentar los resultados de los datos horarios de viento nos seguimos encontrando con una gran cantidad de datos, 8360. Luego, para que los gráficos sean más fáciles de visualizar hemos dividido el año 2010 en cuatro periodos de 2090 datos, (Tabla 5.1).

En lo que se refiere a la notación, cabe mencionar que la dirección de viento va a tomar valores entre $[0, 2\pi)$ y cada valor representa el ángulo que forma respecto al Este. Por tanto, vamos a usar la siguiente codificación $0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ representan las direcciones Este, Norte, Oeste y Sur, respectivamente.

	n	Media direccional	Media de longitud resultante
Primer periodo	2090	5.3634	0.2177
Segundo periodo	2090	0.6446	0.2850
Tercer periodo	2090	0.2134	0.1750
Cuarto periodo	2090	5.7374	0.2952

Tabla 5.1: Estudio de los datos de dirección de viento 2010 mediante una medida de localización (media direccional) y una medida de dispersión (media de longitud resultante)

En este análisis exploratorio de los datos, observamos que en el primer y el cuarto periodo del año 2010 la dirección media es la de Sureste y que la dirección media del segundo y tercer periodo del año 2010 es la de Noreste (Tabla 5.1). Además se observa que los datos están muy dispersos (Figuras 5.2, 5.4, 5.6, 5.8) pues la media de longitud resultante \bar{R} es próxima a 0 (Tabla 5.1). Lo que también se observa a la vista de (Figuras 5.3, 5.5, 5.7, 5.9). De hecho, podemos decir que la distribución de la dirección de viento en los 4 periodos en los que hemos dividido el año 2010 es prácticamente uniforme.

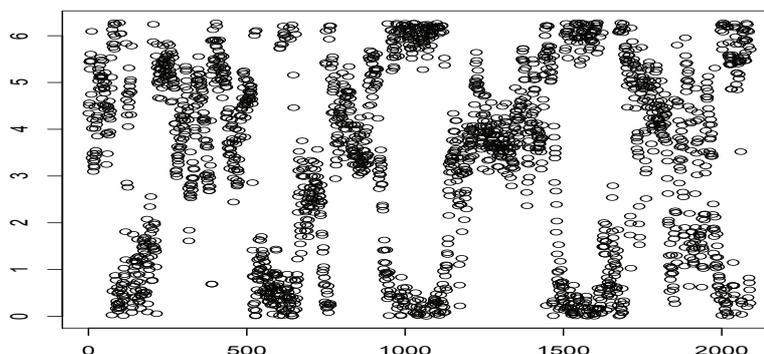


Figura 5.2: Gráfico de dispersión de los datos de dirección de viento del primer periodo de 2010

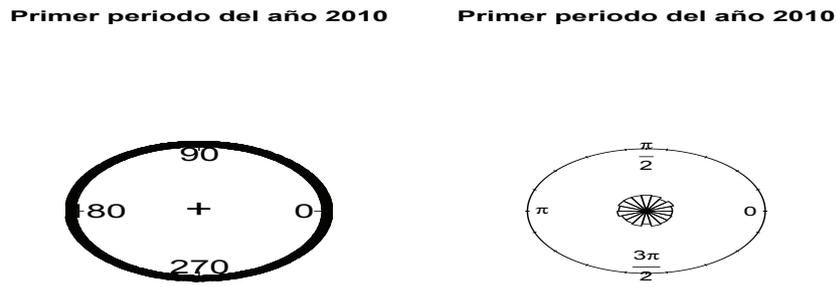


Figura 5.3: Gráfico de dispersión circular y rosa de los vientos de los datos de dirección de viento del primer periodo de 2010

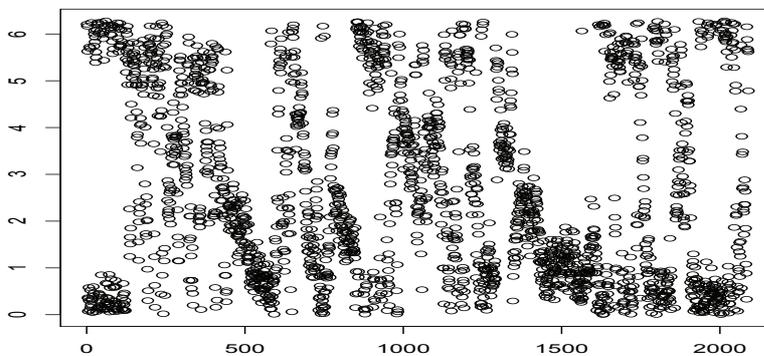


Figura 5.4: Gráfico de dispersión de los datos de dirección de viento del segundo periodo de 2010

Segundo periodo del año 2010 Segundo periodo del año 2010



Figura 5.5: Gráfico de dispersión circular y rosa de los vientos de los datos de dirección de viento del segundo periodo de 2010

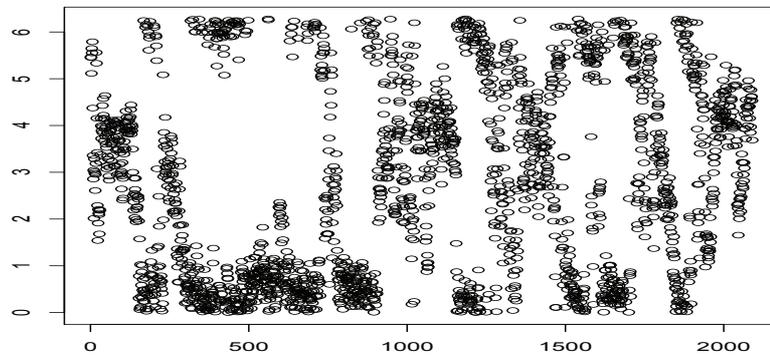


Figura 5.6: Gráfico de dispersión de los datos de dirección de viento del tercer periodo de 2010

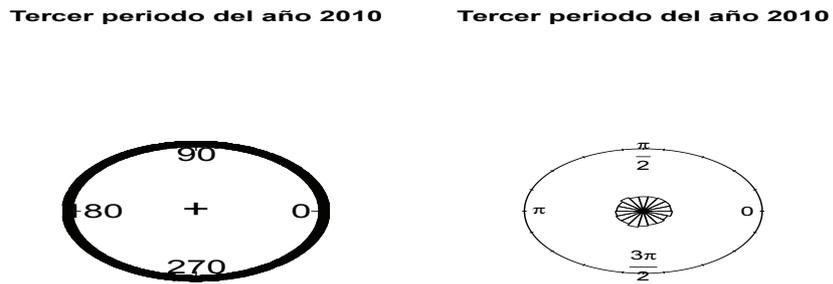


Figura 5.7: Gráfico de dispersión circular y rosa de los vientos de los datos de dirección de viento del tercer periodo de 2010

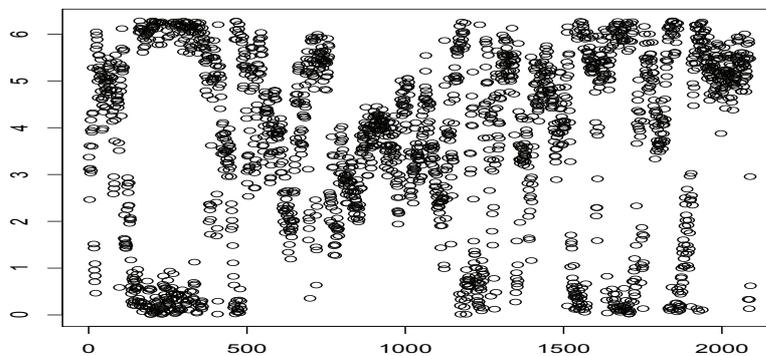


Figura 5.8: Gráfico de dispersión de los datos de dirección de viento del cuarto periodo de 2010

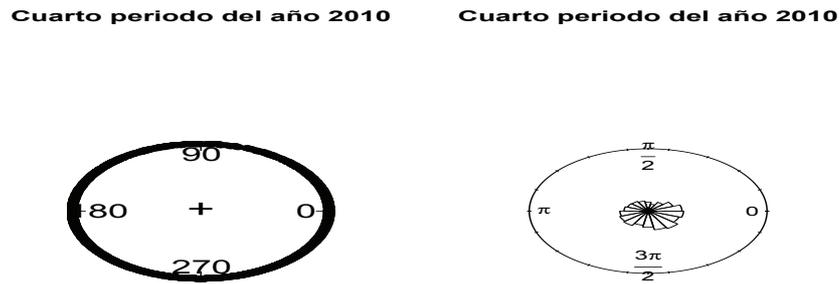


Figura 5.9: Gráfico de dispersión circular y rosa de los vientos de los datos de dirección de viento del cuarto periodo de 2010

5.3 Estimación tipo núcleo de la función de densidad circular

En esta sección obtendremos la estimación no paramétrica tipo núcleo de la densidad circular (2.1) para cada uno de los cuatro periodos en los que hemos dividido el año 2010. Para ello hemos implementado tres diferentes elecciones del parámetro de suavizado (2.4,2.5,2.3) para la correspondiente estimación de la densidad circular en cada uno de los periodos en los que hemos dividido el año 2010.

Los parámetros de suavizado obtenidos (Tabla 5.2) mediante validación cruzada (2.4,2.5) toman valores del mismo orden pero más altos que los obtenidos mediante la técnica plug-in (2.3) en todos los periodos. Por lo que las estimaciones no paramétricas tipo núcleo de la densidad obtenidas seleccionando el parámetro de suavizado mediante la técnica plug-in van a ser más suaves que con el dado por las técnicas de validación cruzada.