Análisis Estadístico de Datos Direccionales. Aplicaciones Medioambientales

María Leyenda Rodríguez

Proyecto Fin de Máster

Máster en Técnicas Estadísticas

Endesa Generación S.A. cuenta en el municipio de As Pontes de García Rodríguez, al noroeste de la provincia de A Coruña, con una importante Unidad de Producción Térmica (U.P.T. As Pontes)

- Central Térmica.
- Ciclo Combinado.

Valores límite de dióxido de azufre (SO_2).

Valor límite horario para la protección de la salud humana	**Valor límite diario para la protección de la salud humana**	**Nivel crítico para la protección de la vegetación**
350 $\mu g/m^3$ No debe superase más de 24 veces/año	125 $\mu g/m^3$ No debe superarse más de 3 veces/año	20 $\mu g/m^3$ (del 1 de Octubre al 31 de Marzo)

Valores límite de dióxido de nitrógeno (NO_2).

Valor límite horario para la protección de la salud humana	**Valor límite anual para la protección de la salud humana**	**Nivel crítico para la protección de la vegetación**
200 $\mu g/m^3$ No debe superase más de 18 veces/año	40 $\mu g/m^3$	30 $\mu g/m^3$ de NO_x expresado en NO_2

Las Centrales tienen implantado un Sistema de Control Suplementario de la Contaminación Atmosférica.

- Red de Vigilancia de la Calidad Atmosférica.
  - Es una red de 10 estaciones de medida automáticas, convenientemente situadas, que transmiten en continuo medidas de la evolución de la calidad del aire.
- Estación Meteorológica de A Mourela.
- Sistema de Control de Emisiones de la Central Térmica y del Ciclo Combinado.

Sistema de Predicción Estadística de Inmisión (SIPEI).

- Predicciones de los valores de dióxido de azufre y de óxidos de nitrógeno, con media hora de antelación.
  - *TIT Modelos De Predicción Medioambiental. María Piñeiro Lamas.*
- Predice cuál es el origen del episodio de alteración de calidad de aire
  - Este puede ser causado por la Central Térmica, el Ciclo Combinado u otros posibles focos como por ejemplo el tráfico o las actividades agrícolas de la zona.
  - Esta última metodología se incorporó recientemente.
  - *PFM Modelos de Clasificación Estadística Basados en Indicadores de SO_2 Y NO_x en la U.P.T. de As Pontes. Francisco Manuel Prieto Magdalena.*

En este trabajo nos hemos propuesto analizar la dirección de viento medida en la estación meteorólogica A Mourela a 80 metros, considerando la naturaleza angular de la dirección de viento.

- Análisis descriptivo direccional (ver Mardia y Jumpp, 2000).
- Estimación no paramétrica tipo núcleo de la densidad circular haciendo un especial hincapié en la selección del parámetro de suavizado (ver Taylor, 2008; Hall, Watson y Cabrera, 1987).
- Estimación del modelo de Möbius de series de tiempo (ver Hughes, 2007)

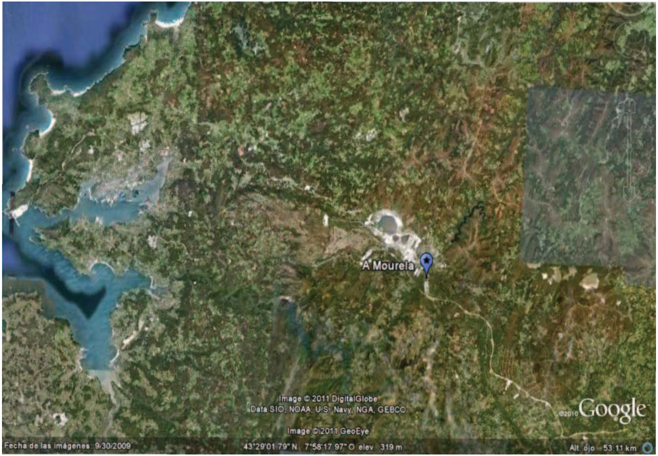
Por otra parte, nos interesa estudiar la relación entre las concentraciones de dióxido de azufre (SO_2) y la dirección de viento en cinco de las estaciones automáticas de medida.

- Estimación no parámetrica tipo núcleo de la función de regresión circular lineal (ver Marzio, Panzera y Taylor, 2009).
- Para realizar esta estimación necesitamos que tanto los datos de SO_2 y dirección de viento estén recogidos por la misma estación.

1. Análisis descriptivo de dirección de viento en A Mourela
2. Estimación de la función de densidad circular
    - Aplicación Medioambiental
3. Modelo de Möbius de series de tiempo
    - Aplicación Medioambiental
4. Estimación de la regresión circular-lineal
    - Aplicación Medioambiental
5. Conclusiones

1. Análisis descriptivo de dirección de viento en A Mourela
2. Estimación de la función de densidad circular
   - Aplicación Medioambiental
3. Modelo de Möbius de series de tiempo
   - Aplicación Medioambiental
4. Estimación de la regresión circular-lineal
   - Aplicación Medioambiental
5. Conclusiones

Comencemos con un análisis descriptivo de los datos de viento recogidos durante el año 2010 por la estación meteorológica A Mourela que situada en As Pontes.



- Esta estación recoge datos minutales.
- Trabajaremos con las medias horarias reduciendo así la gran cantidad de datos y la dependencia.
- Hemos dividido el año 2010 en cuatro periodos de 2090 datos. Con el objetivo de que los resultados sean más fáciles de visualizar.
- Sólo mostraremos los gráficos del tercer periodo.

Notación

Los valores 0, $\frac{\pi}{2}$, π, $\frac{3\pi}{2}$ representan las direcciones Este, Norte, Oeste y Sur, respectivamente.

Medidas de localización y dispersión

- Cada punto **x** en el círculo unidad puede ser representado por un ángulo θ.

$$\mathbf{x} = (\cos\theta, \operatorname{sen}\theta)$$

- La dirección media $\bar{\theta}$ de $\theta_1, \ldots, \theta_n$ es la dirección de $x_1 + \ldots + x_n$ que es el centro de masa $\bar{x}$ de $x_1, \ldots, x_n$.
- Si las coordenadas cartesianas de x_j son $(\cos\theta_j, \operatorname{sen}\theta_j)$, entonces $(\bar{C}, \bar{S})$ son las coordenadas cartesianas del centro de masas.

$$\bar{C} = \frac{1}{n}\sum_{j=1}^{n} \cos\theta_j; \;\; \bar{S} = \frac{1}{n}\sum_{j=1}^{n} \operatorname{sen}\theta_j$$

Medidas de localización y dispersión

Además $\bar{\theta}$ es la solución de las siguientes ecuaciones

$$\bar{C} = \bar{R}\cos\bar{\theta}$$
$$\bar{S} = \bar{R}\,\mathrm{sen}\,\bar{\theta}$$

(supuesto $\bar{R} > 0$), dónde la longitud media resultante $\bar{R}$ es dada por

$$\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$$

Es la medida de dispersión más importante en datos direccionales.

n=2090	Media direccional, $\bar{\theta}$	Medida de dispersión, $\bar{R}$
Primer periodo	5.3634	0.2177
Segundo periodo	0.6446	0.2850
Tercer periodo	0.2134	0.1750
Cuarto periodo	5.7374	0.2952

Tabla: Estudio de los datos de dirección de viento 2010 mediante una medida de localización (media direccional) y una medida de dispersión ($\bar{R}$)

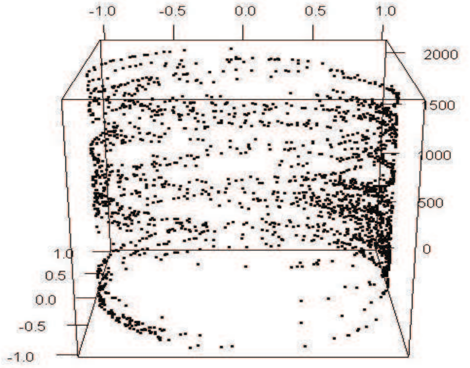


Figura: Gráfico de dispersión de las medias horarias de la dirección de viento correspondientes al tercer periodo de 2010.

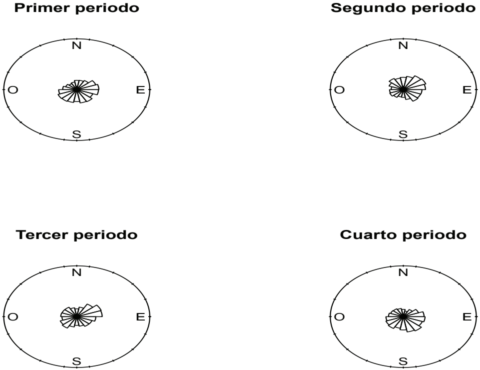


Figura: Diagrama de rosa de las medias horarias de la dirección de viento.

1. Análisis descriptivo de dirección de viento en A Mourela
2. Estimación de la función de densidad circular
   - Aplicación Medioambiental
3. Modelo de Möbius de series de tiempo
   - Aplicación Medioambiental
4. Estimación de la regresión circular-lineal
   - Aplicación Medioambiental
5. Conclusiones

En el caso lineal, si partimos de una muestra aleatoria simple $X_1, \ldots, X_n$, el estimador natural de la probabilidad en cada uno de los intervalos $[x_m, x_{m+1})$, con $x_m = x_0 + hm, \; m \in \mathbb{Z}$, es el histograma

$$\hat{f}_{n,H}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left(X_i \in [x_m, x_{m+1})\right)$$

donde $\mathbb{I}$ denota a la función indicadora.

Modificaremos esta estimación de manera que para cada punto x se construye un intervalo de la forma $(x-h, x+h)$; lo que da lugar a otra estimación no paramétrica de la densidad denominada histograma móvil

$$\hat{f}_{n,N}(x) = \frac{1}{nh}\sum_{i=1}^{n}\omega\left(\frac{x-X_i}{h}\right)$$

donde ω es la densidad de la distribución uniforme en (-1, 1).

- Si se reemplaza ω por una densidad K (denominada núcleo) con una única moda en cero y simétrica se obtiene el estimador tipo núcleo.

Estimador de Parzen-Rosemblat

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

- h es el parámetro de suavizado o ventana.
- Usualmente se supone que la función núcleo es una función de masa de probabilidad simétrica, por ejemplo, la densidad Gausiana.

Función núcleo circular: Densidad von Mises

$$g(\theta;\mu,\kappa)=\frac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta-\mu)}$$

- I_0 denota la función de Bessel modificada de primer tipo y orden 0

$$I_0(\kappa)=\frac{1}{2\pi}\int_0^{2\pi}e^{\kappa\cos\theta}d\theta.$$

- El parámetro μ es la media de las direcciones.
- κ es el parámetro de concentración.

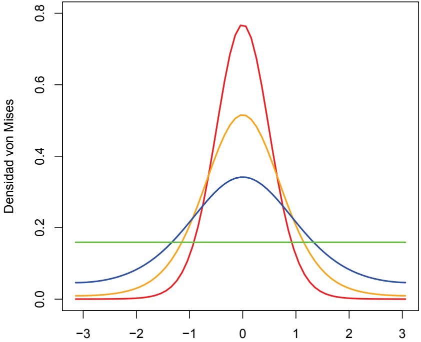


Figura: Representación de la función de densidad von Mises $vM(0,\kappa)$, κ=0 (verde), κ=1 (azul), κ=2 (naranja), κ=4 (rojo).

- Cuando usamos datos en el círculo, la distancia entre dos puntos en el círculo viene dada por

$$d_i = ||x - X_i||^2 = 2(1 - x^T X_i) = 2(1 - \cos(\theta - \Theta_i)),$$

donde $x^T = (\cos\theta, \operatorname{sen}\theta)$ y $X_i = (\cos\Theta_i, \operatorname{sen}\Theta_i)^T$.

- El estimador de Parzen-Rosemblat permite una representación de la estimación no paramétrica tipo núcleo de la densidad circular,

$$\hat{f}(\theta, \nu) = \frac{1}{n}\sum_{i=1}^{n} L_\nu\left(1 - x^T X_i\right) = \frac{1}{n}\sum_{i=1}^{n} L_\nu\left(1 - \cos(\theta - \Theta_i)\right)$$

donde L_ν es una función núcleo circular.

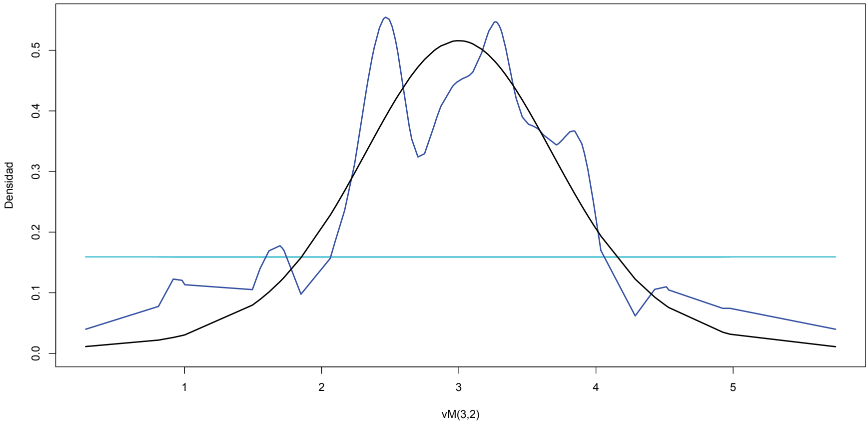
- Dada una muestra aleatoria de ángulos $\Theta_1, \ldots, \Theta_n \in [0, 2\pi]$,

Estimador no paramétrico tipo núcleo de la densidad circular

$$\hat{f}(\theta, \nu) = \frac{1}{2n\pi I_0(\nu)} \sum_{i=1}^{n} \exp(\nu \cos(\theta - \Theta_i))$$

con función núcleo la función de densidad von Mises.

- El parámetro de concentración ν ha asumido el papel de la inversa del parámetro de suavizado.


Densidad
0.0
0.1
0.2
0.3
0.4
0.5
1
2
3
4
5
vM(3,2)

Selección del parámetro de suavizado.Validación cruzada

Obtendremos las funciones de validación cruzada para minimizar la función pérdida dada por el error cuadrático medio, L_2, y la dada por Kullback-Leibler. Ambas dadas por Hall, Watson y Cabrera, (1987).

$$L_2(\nu) = \int_0^{2\pi} \mathbb{E}\left(\hat{f}(\theta,\nu) - f(\theta)\right)^2 d\theta.$$

$$L_{KL}(\nu) = \int_0^{2\pi} f(\theta)\mathbb{E}\left[\log\left\{f(\theta)/\hat{f}(\theta,\nu)\right\}\right] d\theta.$$

Selección del parámetro de suavizado. Validación cruzada; L_2

El parámetro de suavizado obtenido mediante el método de validación cruzada para minimizar la función de pérdida L_2 viene dado por

$$\widehat{v}_2 = \arg\min_{\nu \geq 0} \widehat{CV}_2(\nu)$$

donde, $$\widehat{CV}_2(\nu) = \int_0^{2\pi} \hat{f}^2(\theta,\nu)d\theta - 2n^{-1}\sum_{i=1}^{n}\hat{f}_i(\Theta_i,\nu),$$

siendo $\hat{f}_i(\cdot)$ la estimación no paramétrica tipo núcleo de la densidad construída dejando fuera el valor Θ_i de la muestra

$$\hat{f}_i(\theta,\nu) = \frac{1}{2n\pi I_0(\nu)}\sum_{j\neq i}\exp\left(\nu\cos(\theta-\Theta_j)\right).$$

Selección del parámetro de suavizado. Validación cruzada; Kullback-Leibler

El parámetro de suavizado obtenido mediante el método de validación cruzada para minimizar la función de Kullback-Leibler viene dado por

$$\widehat{v}_{KL} = \arg\min_{\nu \geq 0} -\widehat{CV}_{KL}(\nu),$$

siendo

$$\widehat{CV}_{KL}(\nu) = n^{-1} \sum_{i=1}^{n} \log \left\{ \hat{f}_i(\Theta_i, \nu) \right\}$$

Selección del parámetro de suavizado. Plug-in

- Taylor obtuvo la expresión del Error Cuadrático Medio Integrado Asintótico (AMISE)
- Bajo la hipótesis de que la distribución von Mises con parámetro de concentración κ es la que genera los datos.

$$AMISE(\nu) = 3\kappa^2 I_2(2\kappa) / \left\{32\pi\nu^2 I_0(\kappa)^2\right\} + \nu^{1/2} / \left(2n\pi^{1/2}\right)$$

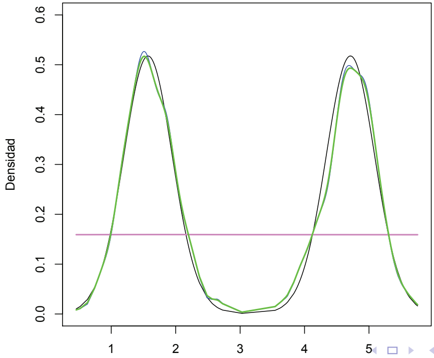
donde I_2 denota la función de Bessel de primer tipo de orden 2

- El parámetro de suavizado ν se puede obtener minimizando el AMISE y utilizando en la estimación de κ, mediante máxima verosimilitud (ver Taylor, 2008):

$$\hat{\nu} = [3n\hat{\kappa}^2 I_2(2\hat{\kappa}) \left\{4\pi^{1/2} I_0(\hat{\kappa})^2\right\}^{-1}]^{2/5}.$$

Ejemplo

Supongamos que la muestra de tamaño 1000 está generada por una mixtura de von Mises de parámetros $\mu_1 = \pi/2$, $\kappa_1 = 7$, $\mu_2 = 3\pi/2$ y $\kappa_2 = 7$ con $\rho = 0{,}5$.



- La estimación tipo núcleo seleccionando el parámetro de suavizado mediante validación cruzada puede ser inconsistente bajo una variedad de circunstancias.
  - para datos discretos con múltiples valores repetidos
  - validación cruzada tiende a sugerir parámetros de suavizado que suavizan muy poco, es decir, ν tiende a ser muy grande.
- Perturbación Caso Circular: Garcia-Portugués et al. (2011)

$$\tilde{\theta}_i = \theta_i + d\epsilon_i, \epsilon_i \sim vM(0,1)$$
$$d = n^{-1/5}$$

Datos originales	Media direccional, $\bar{\theta}$	Medida de dispersión, $\bar{R}$
Primer periodo	5.3727	0.2253
Segundo periodo	0.6239	0.2914
Tercer periodo	0.2315	0.1840
Cuarto periodo	5.1451	0.3038

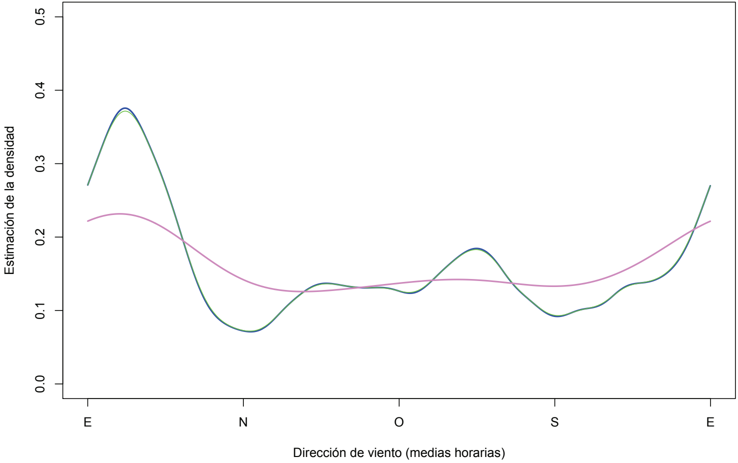
Datos perturbados	Media direccional, $\bar{\theta}$	Medida de dispersión, $\bar{R}$
Primer periodo	5.3634	0.2177
Segundo periodo	0.6446	0.2850
Tercer periodo	0.2134	0.1750
Cuarto periodo	5.7374	0.2952

	plug-in	L_2	Kullback-Leibler
Tercer periodo	2.35	235.67	204.92

Tabla: Estimaciones del parámetro de suavizado para la estimación no paramétrica tipo núcleo de la función de densidad circular (sin perturbar).

	plug-in	L_2	Kullback-Leibler
Tercer periodo	2.17	41.44	36.17

Tabla: Estimaciones del parámetro de suavizado para la estimación no paramétrica tipo núcleo de la función de densidad circular (perturbando).


Estimación de la densidad
0.0
0.1
0.2
0.3
0.4
0.5
E
N
O
S
E
Dirección de viento (medias horarias)

1. Análisis descriptivo de dirección de viento en A Mourela
2. Estimación de la función de densidad circular
   - Aplicación Medioambiental
3. Modelo de Möbius de series de tiempo
   - Aplicación Medioambiental
4. Estimación de la regresión circular-lineal
   - Aplicación Medioambiental
5. Conclusiones

Estudio de la dependencia en los datos de dirección de viento

Coeficiente de autocorrelación basado en la estimación de la correlación circular entre dos variables aleatorias Θ, Φ introducido por Fisher y Lee (1983).

$$\rho_T = \frac{\mathbf{E}\left[\mathrm{sen}(\Theta_1 - \Theta_2)\,\mathrm{sen}(\Phi_1 - \Phi_2)\right]}{\left\{\mathbf{E}\left[\mathrm{sen}^2(\Theta_1 - \Theta_2)\right]\mathbf{E}\left[\mathrm{sen}^2(\Phi_1 - \Phi_2)\right]\right\}^{1/2}}.$$

donde (Θ_1, Φ_1) y (Θ_2, Φ_2) son independientes e identicamente distribuídas como (Θ, Φ).

Estudio de la dependencia en los datos de dirección de viento

Calcularemos la autocorrelación circular k-lag $\hat{\rho}_T^k$ como

$$\hat{\rho}_T^k = \frac{\sum_{1\leq i<j\leq n-k} \operatorname{sen}(\theta_i - \theta_j) \operatorname{sen}(\theta_{i+k} - \theta_{j+k})}{\left[\sum_{1\leq i<j\leq n-k} \operatorname{sen}^2(\theta_i - \theta_j) \sum_{1\leq i<j\leq n-k} \operatorname{sen}^2(\theta_{i+k} - \theta_{j+k})\right]^{1/2}}.$$

- Por tanto, si representamos $\hat{\rho}_T^k$ frente a k obtendremos la autocorrelación circular para los distintos periodos.
- Estudiaremos el tercer periodo.

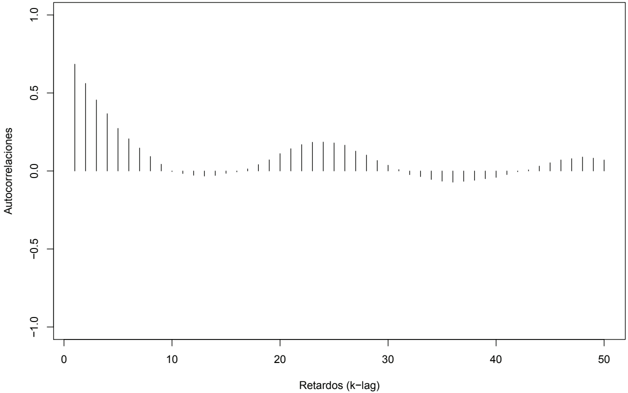


Figura: Autocorrelaciones circulares de las medias horarias de la dirección de viento correspondientes al tercer periodo de 2010.

- La componente determinística del modelo de regresión estudiado por Downs y Mardia (2002) une a la variable angular dependiente v con la variable angular independiente u mediante

$$\tan\frac{1}{2}(v-\beta)=\omega\tan\frac{1}{2}(u-\alpha)$$

- $\omega \in [-1,1]$ es el parámetro pendiente.
- $-\pi \leq \alpha, \beta < \pi$ son los parámetros angulares de localización.
- Definimos una relación uno a uno entre u y v,siempre que $\omega \neq 0$ de la siguiente forma

$$v=\beta+2\tan^{-1}\left\{\omega\tan\frac{1}{2}(u-\alpha)\right\}$$

- Reemplazamos el ángulo v por θ_t y el ángulo u por θ_{t-1}, $t = 2, \ldots, n$. Esta sustitución sugiere un único parámetro de localización, $\alpha = \beta$

$$v = \beta + 2\tan^{-1}\left\{\omega \tan \frac{1}{2}(u - \alpha)\right\}$$

$$\theta_t = \alpha + 2\tan^{-1}\left\{\omega \tan \frac{1}{2}(\theta_{t-1} - \alpha)\right\} = \mu_t$$

Modelo de Möbius de series de tiempo

$$\theta_t = \mu_t + \epsilon_t$$

donde $\epsilon_t \sim vM(0, \kappa)$

Estimación de los parámetros. Máxima verosimilitud

- Estimamos los parámetros α y ω mediante máxima verosimiltud; es decir, maximizando la función

$$l(\alpha, \omega) = \sum_{t=2}^{n} \cos \left[\theta_t - \alpha - 2 \tan^{-1} \left\{ \omega \tan \frac{1}{2} (\theta_{t-1} - \alpha) \right\} \right]$$

- La maximización la haremos usando la función *nlm* de R.
- Esta función utiliza un algoritmo tipo Newton para minimizar una función dada. Por tanto hemos minimizado la función $-l_C(\alpha, \omega)$.

Estimación de los parámetros. Máxima verosimilitud

Una vez que $l_C(\alpha, \omega)$ ha sido máximizada con respecto a α y ω, se obtiene una estimación de κ maximizando

$$l_C(\hat{\alpha}, \hat{\omega}, \kappa) = const. - (n-1)\log I_0(\kappa) + \kappa l\left(\hat{\alpha}, \hat{\omega}\right)$$

respecto de κ.

	$\widehat{\alpha}$	$\widehat{\omega}$	$\widehat{\kappa}$
Primer periodo	-0.7492	0.9634	4.12
Segundo periodo	0.5133	0.9296	4.09
Tercer periodo	-0.1431	0.9644	4.11
Cuarto periodo	-1.3190	0.9382	4.11

Tabla: Estimaciones de los parámetros que están involucrados en la obtención del modelo de Möbius de series de tiempo

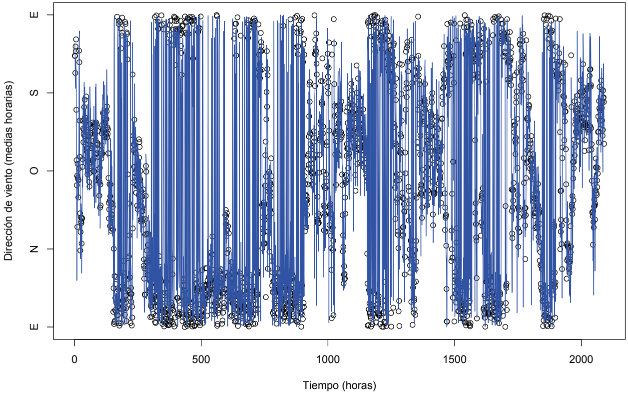
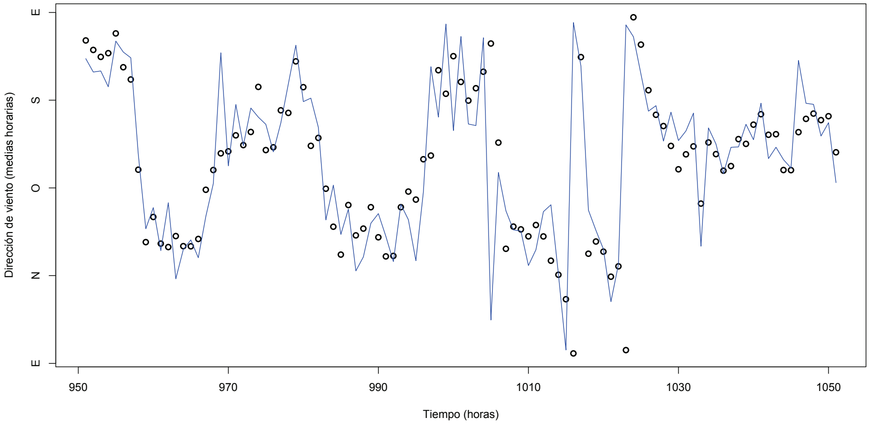


Figura: Modelo de Möbius de series de tiempo de las medias horarias de la dirección de viento correspondientes al tercer periodo de 2010.


Dirección de viento (medias horarias)
E
S
O
N
E
950
970
990
1010
1030
1050
Tiempo (horas)

- El modelo de series de tiempo obtenido en cada uno de los periodos parece que modela bien nuestros datos de dirección de viento.
- Calculamos el error cuadrático medio mediante la distancia longitud de arco, que viene dado por

$$\frac{1}{n}\sum_{i=1}^{n} d(\theta_i, \hat{\theta}_i)^2 = \frac{1}{n}\sum_{i=1}^{n} (\pi - |\pi - |\theta_i - \hat{\theta}_i||)^2,$$

  - $\theta_i, \ i = 1, \ldots, n$ corresponden a las medias horarias de la dirección de viento del periodo correspondiente.
  - $\hat{\theta}_i, \ i = 1, \ldots, n$ corresponden a las estimaciones dadas por modelo de Möbius de series de tiempo de las medias horarias de la dirección de viento del periodo correspondiente.
  - d es la distancia entre dos ángulos definida por la longitud de arco.

- Se obtiene que las desviaciones estandar de las predicciones con respecto a las observaciones está alrededor de $\sqrt{0,25}$ radianes, o lo que es lo mismo a **0.5** radianes.

Consideración

- Sea $\Theta_i \sim U_{[0,2\pi)}$, $i = 1, \ldots, n$ (observado)
- Sea $\hat{\theta}_i = \pi$, $i = 1, \ldots, n$ (esperado)

$$\frac{1}{n}\sum_{i=1}^{n} d(\theta_i, \hat{\theta}_i)^2 \overset{\sim}{\to} Var(\Theta) = \frac{(2\pi - 0)^2}{12} = \frac{4\pi^2}{12}$$

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n} d(\theta_i, \hat{\theta}_i)^2} \overset{\sim}{\to} \frac{\pi}{\sqrt{3}} = \mathbf{1.814}$$

1. Análisis descriptivo de dirección de viento en A Mourela
2. Estimación de la función de densidad circular
   - Aplicación Medioambiental
3. Modelo de Möbius de series de tiempo
   - Aplicación Medioambiental
4. Estimación de la regresión circular-lineal
   - Aplicación Medioambiental
5. Conclusiones

- Dadas n observaciones de una variable direccional explicativa $\Theta_1, \ldots, \Theta_n$ y una variable respuesta lineal $Y_1, \ldots, Y_n$.
- Suponemos que $Y_i = m(\Theta_i) + \epsilon_i$, donde ϵ_i son variables con media cero, independientes e identicamente distribuídas.

Estimación no paramétrica tipo núcleo de la regresión circular-lineal

$$\hat{m}(\theta; \nu) = \frac{\sum_{i=1}^{n} Y_i g(\theta - \Theta_i, 0, \nu)}{\sum_{i=1}^{n} g(\theta - \Theta_i, 0, \nu)}$$

donde g denota la función de densidad von Mises.

- Este estimador es análogo al estimador propuesto por Nadaraya-Watson en 1964 para el caso lineal.

Datos lineales. Núcleo Gausiano

- Silverman recomienda un parámetro de suavizado de $0{,}9\hat{\sigma}n^{1/5}$.
- Scott $h_s = 1{,}06\hat{\sigma}n^{1/5}$.
- $\hat{\sigma}$ es dado por el mínimo de la desviación típica muestral $x_1, \ldots, x_n$ y el rango intercuartílico dividido por 1.34.

Densidad von Mises & Densidad Gausiana

- Si κ es grande, sea $\theta \sim vM(\mu, \kappa)$ y sea $\xi = \kappa^{1/2}(\theta - \mu)$.

$$\theta \sim vM(\mu, \kappa) \Rightarrow \kappa^{1/2}(\theta - \mu) \sim N(0, 1), \ \kappa \to \infty.$$

Selección del parámetro de suavizado. Plug-in

$$\widehat{\nu} = \frac{1}{h_s^2}$$

Selección del parámetro de suavizado. Validación Cruzada

- Como en el caso lineal, escogemos ν de modo que minimize a la función

$$CV(\nu) = n^{-1} \sum_{j=1}^{n} [y_j - \hat{m}^{-j}(\theta_j;\nu)]^2$$

- donde son las estimaciones sin el dato j-ésimo

$$\hat{m}^{-j}(\theta_j;\nu) = \frac{\sum_{i \neq j}^{n} y_i g(\theta_j - \theta_i, 0, \nu)}{\sum_{i \neq j}^{n} g(\theta_j - \theta_i, 0, \nu)}$$

- Se ha realizado la estimación no paramétrica tipo núcleo de la regresión circular-lineal
  - de las medias horarias transformadas mediante la función logaritmo de las concentraciones de SO_2
  - frente a las direcciones horarias de viento.
- En cinco de las estaciones de medición automática, B1, B2, C9, F2 y G2.

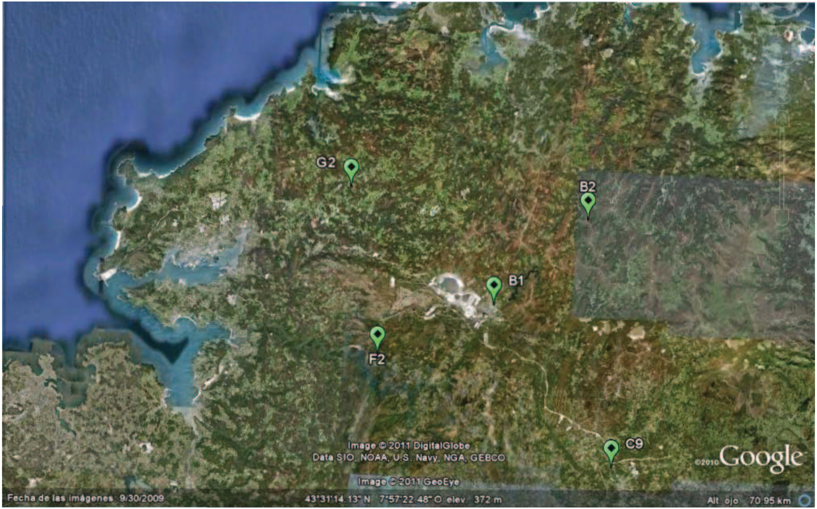


Figura: Estaciones de medida automáticas

- Al analizar los datos obtenidos en las estaciones se observó que también había valores repetidos en las concentraciones de SO_2.
- Perturbación Caso Lineal: Azzalini (1981) propone una perturbación que permite una estimación consistente de la función de distribución

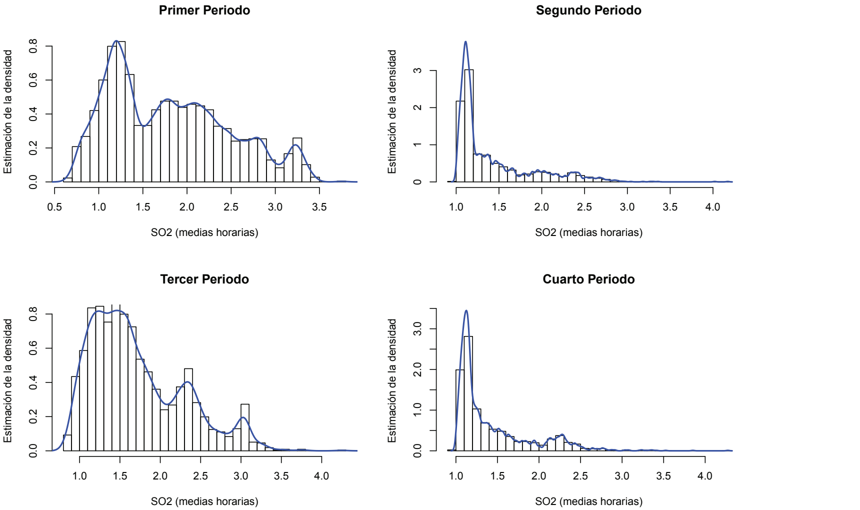
$$\tilde{X}_i = X_i + b\epsilon_i, \epsilon_i \sim Epanech(-\sqrt{5}, \sqrt{5})$$
$$b = 1{,}03\hat{\sigma}n^{-1/3}$$

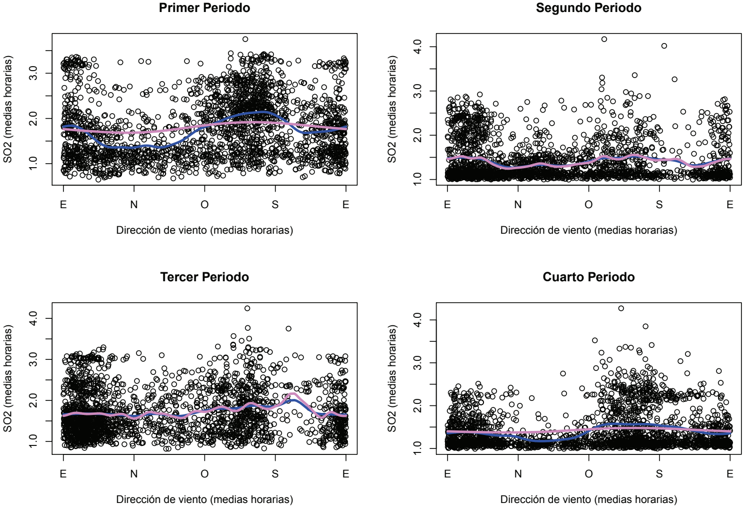
	B1	B2	C9	F2	G2
Primer periodo	3.33	2.87	30.28	37.69	26.42
Segundo periodo	13.07	8.83	7.07	10.73	20.13
Tercer periodo	6.61	25.56	2.33	11.97	18.43
Cuarto periodo	0.00	22.63	4.87	11.24	17.74

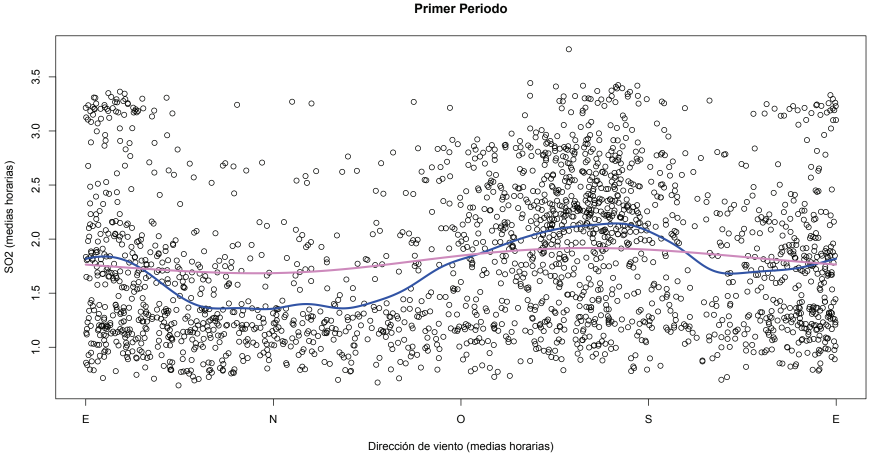
Tabla: Estimaciones del parámetro de suavizado obtenidas mediante validación cruzada.

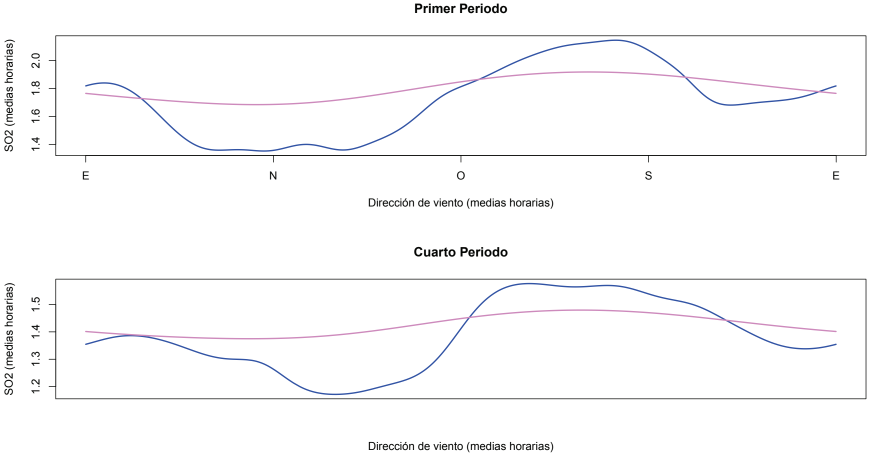
	B1	B2	C9	F2	G2
Primer periodo	592.52	1.04	16.48	1329.96	0.76
Segundo periodo	28.55108	108.28	7.32	0.50	71.05
Tercer periodo	44.07	331.62	7.15	0.51	83.39
Cuarto periodo	48020.99	1.21	11.38	0.51	0.72

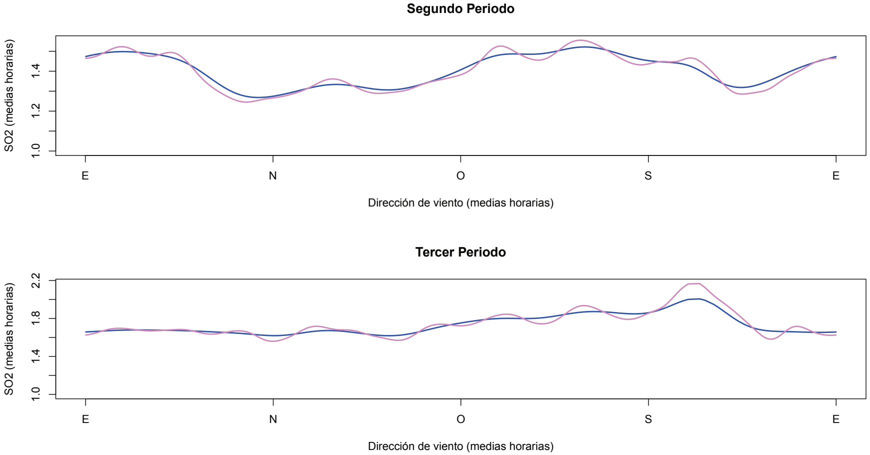
Tabla: Estimaciones del parámetro de suavizado obtenidas mediante la expresión $\kappa = 1/h_s^2$.


Primer Periodo
Segundo Periodo
Tercer Periodo
Cuarto Periodo
Estimación de la densidad
SO2 (medias horarias)


Primer Periodo
Segundo Periodo
Tercer Periodo
Cuarto Periodo
SO2 (medias horarias)
Dirección de viento (medias horarias)
E
N
O
S
E
1.0
2.0
3.0
4.0


Primer Periodo
SO2 (medias horarias)
1.0
1.5
2.0
2.5
3.0
3.5
E
N
O
S
E
Dirección de viento (medias horarias)


Primer Periodo
SO2 (medias horarias)
1.4
1.6
1.8
2.0
E
N
O
S
E
Dirección de viento (medias horarias)
Cuarto Periodo
SO2 (medias horarias)
1.2
1.3
1.4
1.5
Dirección de viento (medias horarias)


Segundo Periodo
SO2 (medias horarias)
1.0
1.2
1.4
E
N
O
S
E
Dirección de viento (medias horarias)
Tercer Periodo
SO2 (medias horarias)
1.0
1.4
1.8
2.2
E
N
O
S
E
Dirección de viento (medias horarias)

1. Análisis descriptivo de dirección de viento en A Mourela
2. Estimación de la función de densidad circular
   - Aplicación Medioambiental
3. Modelo de Möbius de series de tiempo
   - Aplicación Medioambiental
4. Estimación de la regresión circular-lineal
   - Aplicación Medioambiental
5. Conclusiones

- Las estimaciones no paramétricas tipo núcleo de la densidad circular seleccionando el parámetro de suavizado mediante el método de validación cruzada se comportan mejor que si lo hacemos mediante plug-in.
- Las estimaciones no paramétricas tipo núcleo de la función de regresión circular-lineal seleccionando el parámetro de suavizado mediante el método de validación cruzada, también se comportan mejor que al escogerlo mediante plug-in.
- El ajuste proporcionado por el modelo de Möbius de series de tiempo parece razonable.

Panzera A. Di Marzio, M. and C.C. Taylor.
Local polynomial regression for circular predictors.
Statistics and Probability Letters, 79:2066–2075, 2009.

T. D. Downs and K. V. Mardia.
Circular regression.
Biometrika, 89:683–698, 2002.

Watson G.S. Hall, P. and J. Cabrera.
Kernel density estimation whith spherical data.
Biometrika, 74:751–762, 1987.

G. Hughes.
Multivariate and Times Series Models for Circular Data with Applications to Protein Conformational Angles.
PhD thesis, The University of Leeds, Department of Statistics., 2007.

K.V. Mardia and P.E Jupp.
Directional Statistics.
Wiley; New York, 2000.

C.C. Taylor.
Automatic bandwith selection for circular density estimation.
Computational Statistics and Data Analysis, 52:3493–3500, 2008.