

Estimación y predicción de la velocidad del viento mediante funciones ortogonales empíricas y modelos de volatilidad estacional



Sergio Baena Mirabete
M. Pilar Muñoz

Departamento de Estadística e Investigación Operativa
Universitat Politècnica de Catalunya
Barcelona

Índice

- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ Análisis espacial
- ❑ Antecedentes
- ❑ Esquema metodológico
 - ❑ Reducción de la dimensión (EOFs)
 - ❑ Modelos de volatilidad estacional e implementación en R
- ❑ Aplicación: estimación y predicción de la velocidad del viento
- ❑ Conclusiones
- ❑ Futuras líneas de trabajo

Objetivo y motivación

- ❑ **Objetivo:** implementación, estimación y predicción de modelos espacio-temporales para la velocidad del viento en la zona del Estrecho de Gibraltar.
- ❑ **Motivación:** predecir la velocidad del viento con precisión es de vital importancia para los productores de electricidad eólica.
 - ❑ La energía eólica es un recurso abundante, renovable, limpio y ayuda a reducir las emisiones de efecto invernadero.
 - ❑ Su principal inconveniente es su intermitencia.
- ❑ Por encima de un cierto umbral de velocidad (*peak*), no se ponen en marcha los molinos de viento. Estimar la **volatilidad** permite evaluar el riesgo de no producción.
- ❑ Otras aplicaciones: evolución de las partículas contaminantes en el aire (PM10),...

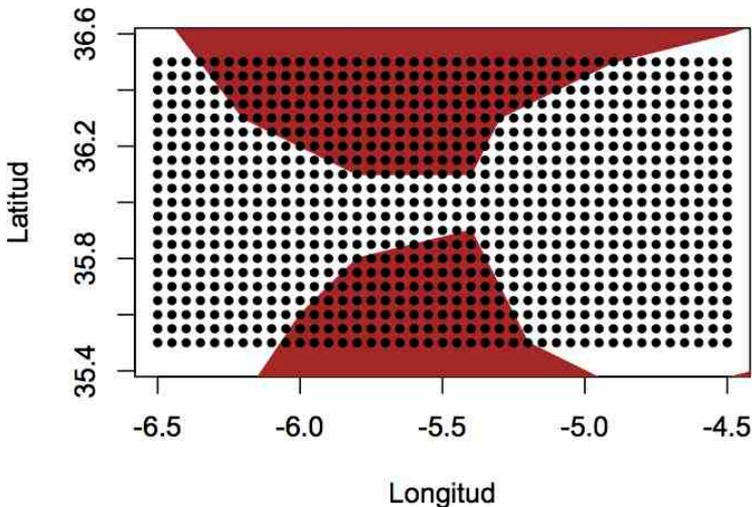
Índice

- Objetivo / Motivación
- **Descripción de los datos**
- Análisis espacial
- Antecedentes
- Esquema metodológico
 - Reducción de la dimensión (EOFs)
 - Modelos de volatilidad estacional e implementación en R
- Aplicación: estimación y predicción de la velocidad del viento
- Conclusiones
- Futuras líneas de trabajo

Descripción de los datos

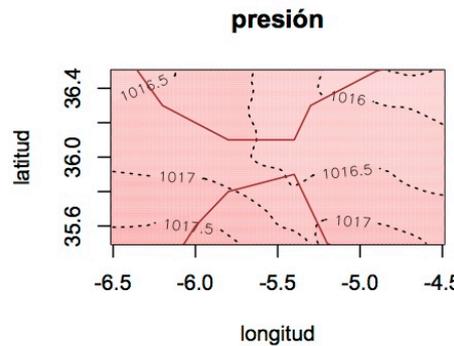
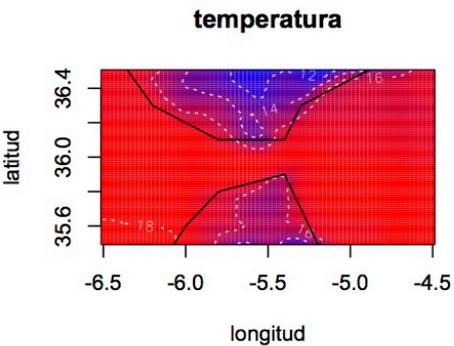
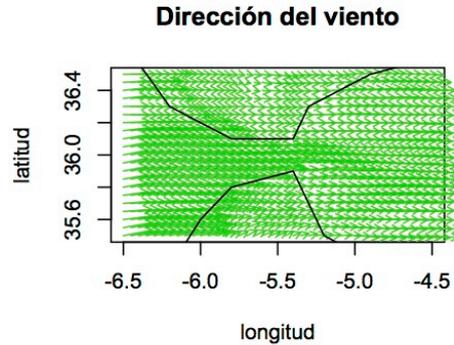
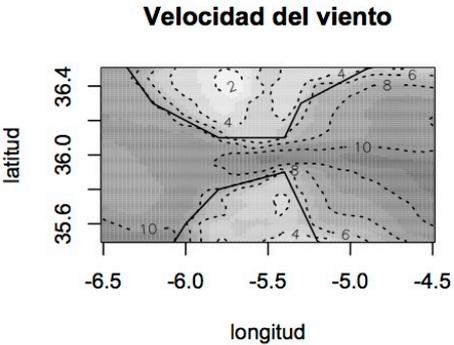
- ❑ Los datos utilizados provienen de un modelo numérico de predicción meteorológica (HIRLAM): no son datos observados.
- ❑ Práctica habitual: *post-processing* (datos numéricos + observados)
- ❑ Área geográfica de estudio: zona del Estrecho de Gibraltar (longitud: W4°30' a W6°30' / latitud: N35°3' a N36°5').

Área geográfica de estudio

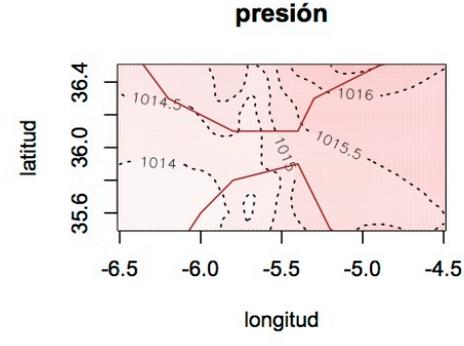
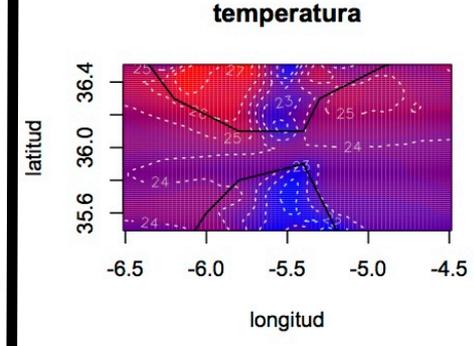
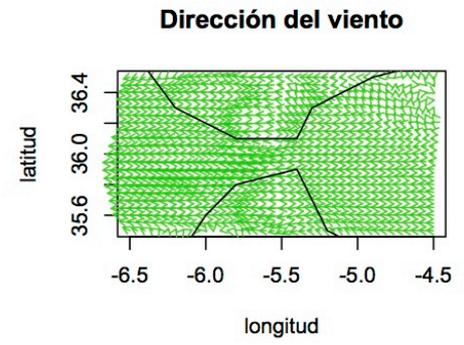
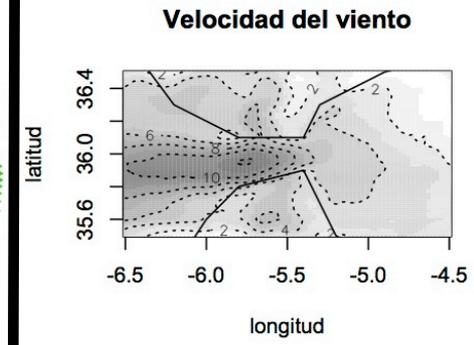


- ❑ 861 localizaciones espaciales.
- ❑ Cada punto es una serie temporal: 861x5.824.
- ❑ Datos desde 1 enero 2009 hasta 31 diciembre 2010.
- ❑ En total, más de 5 millones de datos:
 - ❑ Velocidad y dirección del viento
 - ❑ Presión
 - ❑ Temperatura, entre otras...

Descripción de los datos



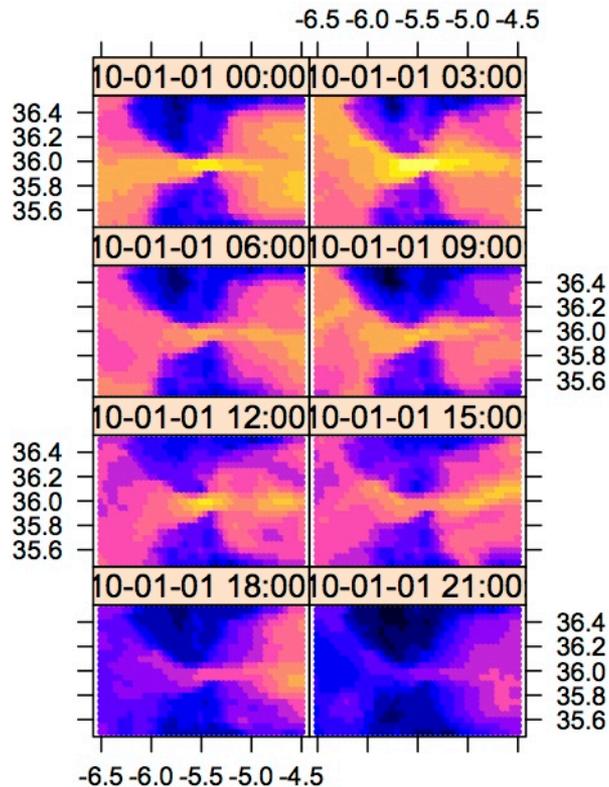
1 Enero 2010 6:00 horas



1 Julio 2010 6:00 horas

Descripción de los datos

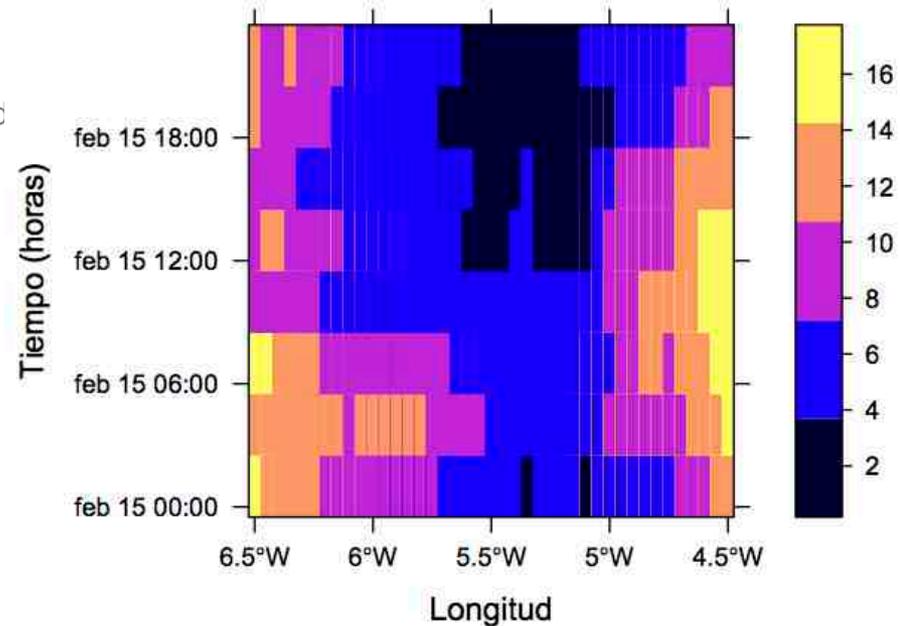
Velocidad del viento: 1-Enero-2010



Tramos de velocidad

- [0.1349, 1.108]
- (1.108, 2.081]
- (2.081, 3.054]
- (3.054, 4.027]
- (4.027, 5]
- (5, 5.973]
- (5.973, 6.946]
- (6.946, 7.919]
- (7.919, 8.892]
- (8.892, 9.864]
- (9.864, 10.84]
- (10.84, 11.81]
- (11.81, 12.78]
- (12.78, 13.76]
- (13.76, 14.73]

Diagrama Hovmöller: 15-Febrero-2010



Relaciona en un mismo gráfico la dimensión espacial (longitud) y la temporal. En la zona del atlántico se observa una velocidad del viento mayor en las primeras horas del día.

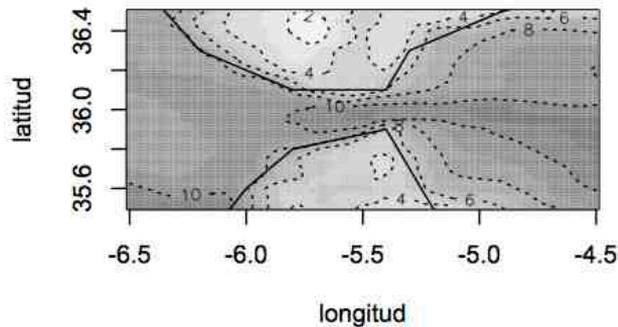
Índice

- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ **Análisis espacial**
- ❑ Antecedentes
- ❑ Esquema metodológico
 - ❑ Reducción de la dimensión (EOFs)
 - ❑ Modelos de volatilidad estacional e implementación en R
- ❑ Aplicación: estimación y predicción de la velocidad del viento
- ❑ Conclusiones
- ❑ Futuras líneas de trabajo

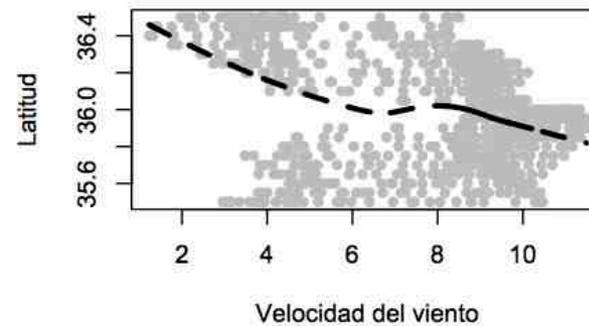
Análisis espacial

- Tiempo fijado (foto): 1 de enero de 2010 a las 6:00 horas

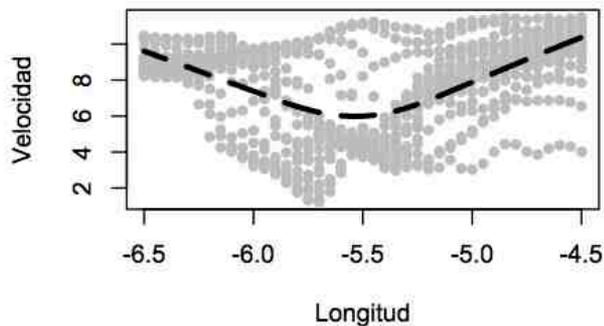
Velocidad del viento (1-Enero-2010 6:00 h)



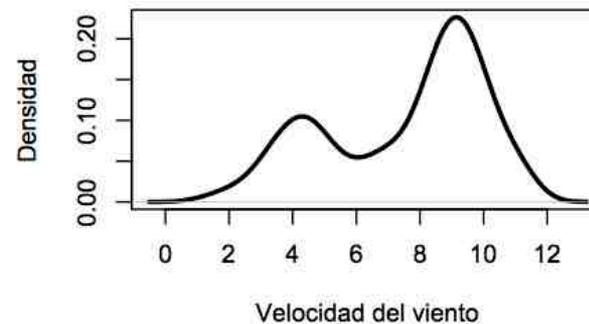
Velocidad del viento y latitud



Velocidad del viento y longitud

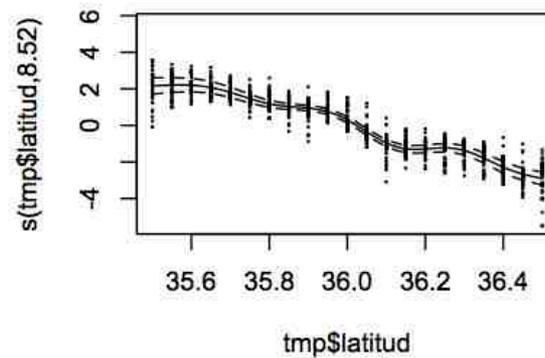
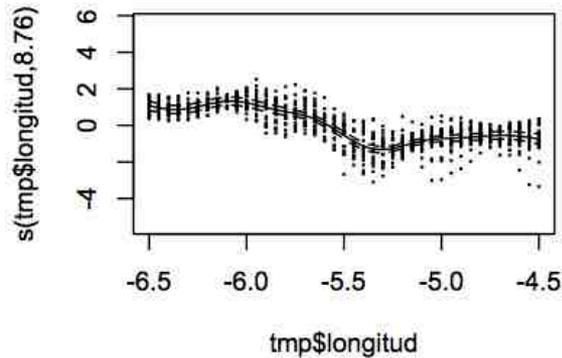


Estimación función de densidad (Kernel)

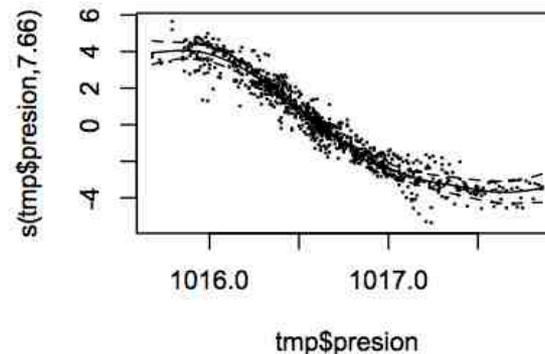
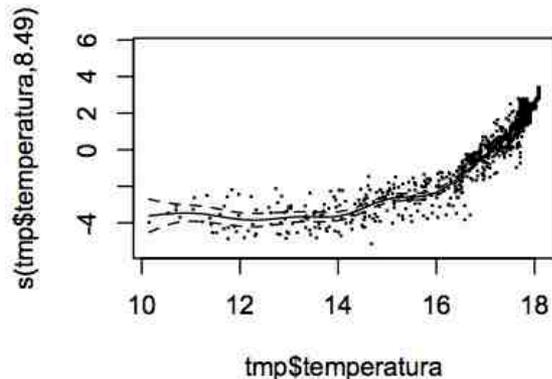


Análisis espacial

▣ Análisis de tendencia en los datos: modelo GAM



Existe una relación entre la velocidad del viento y las coordenadas espaciales

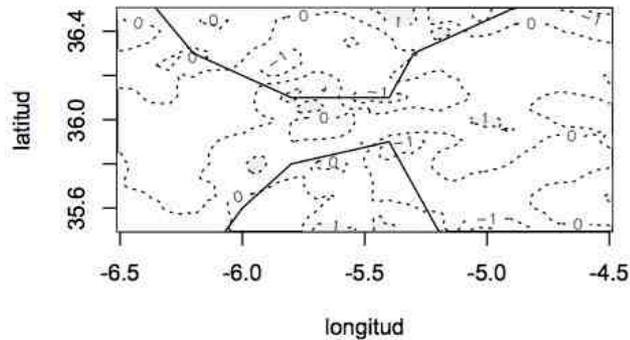


Existe una relación entre la velocidad del viento vs la presión (-) y temperatura (+)

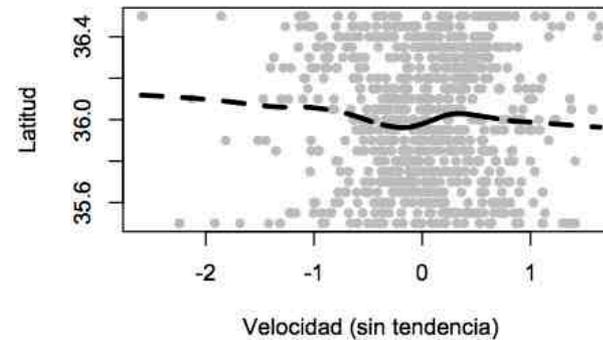
Análisis espacial

- Datos sin tendencia: residuos del modelo GAM

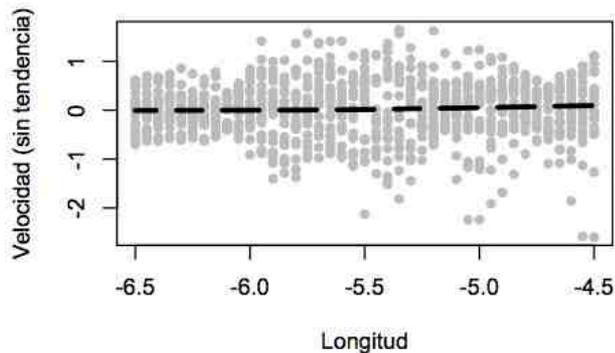
Velocidad del viento (sin tendencia)



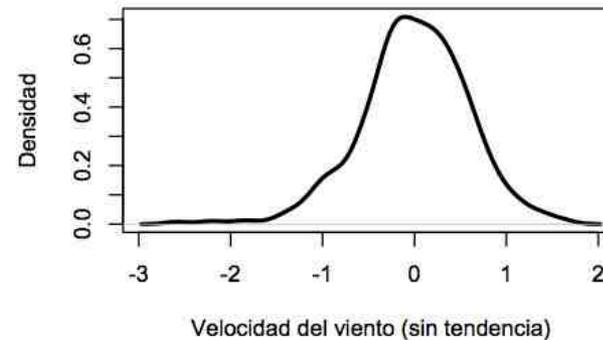
Velocidad del viento (sin tendencia) y latitud



Velocidad del viento (sin tendencia) y longitud



Estimación función de densidad (Kernel)



Análisis espacial

- **Semivariograma:** función que relaciona la variancia del incremento observado de la velocidad del viento entre dos puntos versus la distancia existente entre ambos puntos.
- Datos sin tendencia: para distancias pequeñas, variabilidad baja (continuidad en los datos).

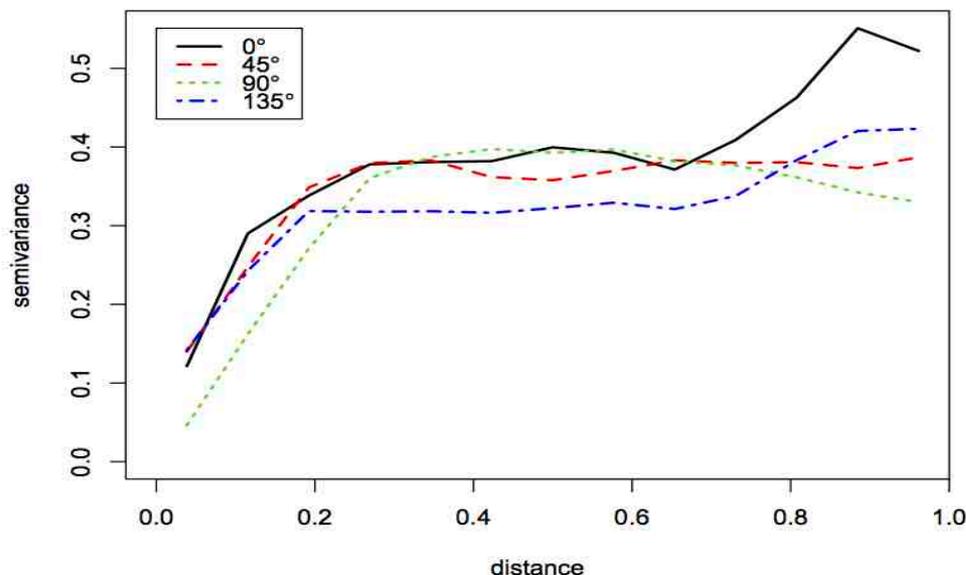


Foto espacial para el 1 de enero de 2010 a las 6 horas...
¿Y la dimensión temporal?

La predicción de la velocidad del viento tiene que tener en cuenta también la evolución en el tiempo

Índice

- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ Análisis espacial
- ❑ **Antecedentes**
- ❑ Esquema metodológico
 - ❑ Reducción de la dimensión (EOFs)
 - ❑ Modelos de volatilidad estacional e implementación en R
- ❑ Aplicación: estimación y predicción de la velocidad del viento
- ❑ Conclusiones
- ❑ Futuras líneas de trabajo

Antecedentes

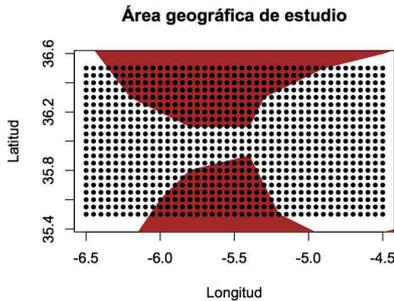
- ❑ Predicción meteorológica: basada en modelos numéricos a partir de ecuaciones diferenciales parciales (EDPs)
 - ❑ Alta resolución: dan resultados a pequeña escala (p.e cada 12 kms) pero necesitan muchos puntos en la malla (coste computacional alto)
 - ❑ Baja resolución: a gran escala (p.e cada 200 kms), bien computacionalmente pero malas predicciones locales.
- ❑ Década 60 y 70, la NWS estima un modelo estadístico con variables observadas + predicciones numéricas (Post-Procesado): mejora de las predicciones en áreas pequeñas.
- ❑ Este trabajo se engloba dentro de las técnicas de post-procesado:
 - ❑ Implementación de modelos de volatilidad estacional multiplicativos (SGARCH) para la estimación y predicción de la velocidad del viento. Lau (1999) y Doshi (2011) trabajan sobre estos modelos desde un punto de vista teórico.
 - ❑ Otros enfoques: PGARCH (Bollerslev, 1996) o introducción de una variable dummy en la ecuación de volatilidad para tratar la estacionalidad (Gil, 2008)

Índice

- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ Análisis espacial
- ❑ Antecedentes
- ❑ **Esquema metodológico**
 - ❑ **Reducción de la dimensión (EOFs)**
 - ❑ **Modelos de volatilidad estacional e implementación en R**
- ❑ Aplicación: estimación y predicción de la velocidad del viento
- ❑ Conclusiones
- ❑ Futuras líneas de trabajo

Esquema metodológico

Malla de puntos zona del Estrecho de Gibraltar desde enero de 2009 a diciembre de 2010



Proceso espacio-temporal
 $X(t,s)$ $s=1,\dots,861$ $t=1,\dots,5824$

Descomposición a partir de EOFs:
 $y_k(t)$: k -ésima componente principal (CP).
 u_k : k -ésima EOF (espacial)

$$X(t,s) = \sum_{k=1}^M Y_k(t)u_k(s)$$

Modelización temporal 1ª seis CPs
 (92,50% variancia total):
 $Y_1(t), Y_2(t), \dots, Y_6(t)$
 $Y_k(t)$: modelo de volatilidad estacional SARMA-SGARCH

Validación Modelo sobre $Y_k(t)$:
 - Análisis residuos: ACF /PACF,...
 - Verificación supuestos del modelo

Predicción temporal $\hat{Y}_k(t+h)$ $h=1,\dots,T$
 T : horizonte de predicción, p.e: 4

Predicción espacio-temporal (en los puntos de la malla)

$$\hat{X}(t+h,s) = \sum_{k=1}^M \hat{Y}_k(t+h)u_k(s)$$

Validación gráfica errores espacio-temporales $E(t+h,s)$

$$E(t+h,s) = \hat{X}(t+h,s) - X(t+h,s)$$

Reducción de la dimensión (EOF)

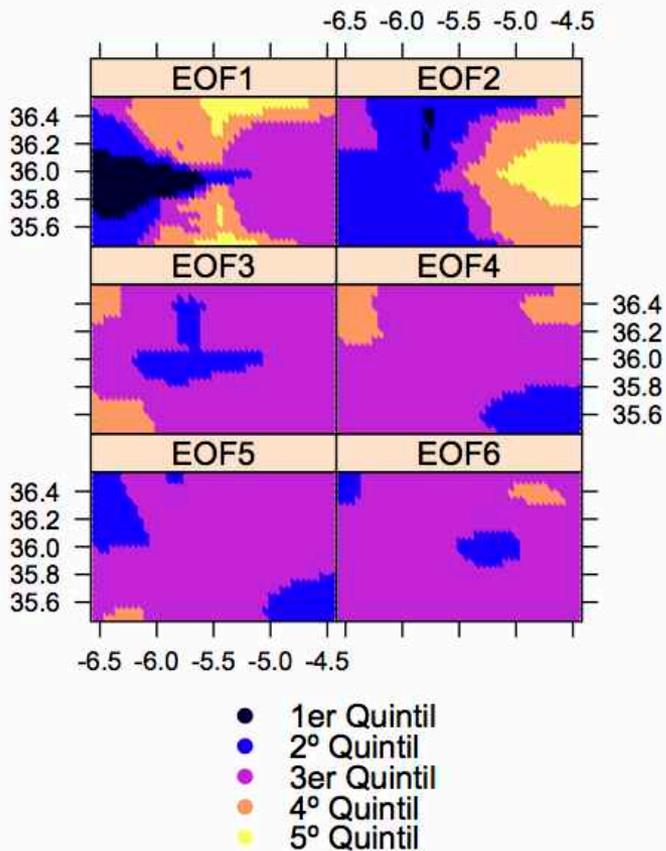
- ❑ **Funciones ortogonales empíricas (EOF):** descomponer un proceso continuo espacio-temporal $X(t,s)$, que representa la velocidad del viento, a partir de un conjunto de funciones espaciales $u_k(s)$ y temporales $y_k(t)$:

$$X(t,s) = \sum_{k=1}^M y_k(t) u_k(s) \quad M=861 \text{ localizaciones}$$

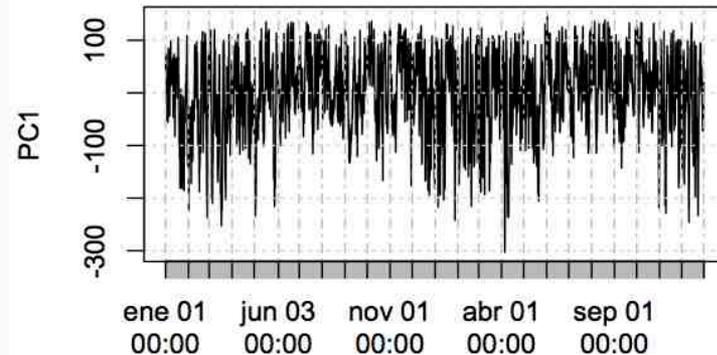
- ❑ $u_k(s)$: patrón espacial (EOF) que corresponde al k-ésimo vector propio de la matriz de covariancias.
- ❑ $Y_k(t)$: patrón temporal (componente principal). Es la proyección de la matriz de datos original (centrada) sobre la matriz de EOFs.
 - ❑ Primer PC explica el 65% de la variancia total. Los 6 primeros el 92,5%.

Reducción de la dimensión (EOF)

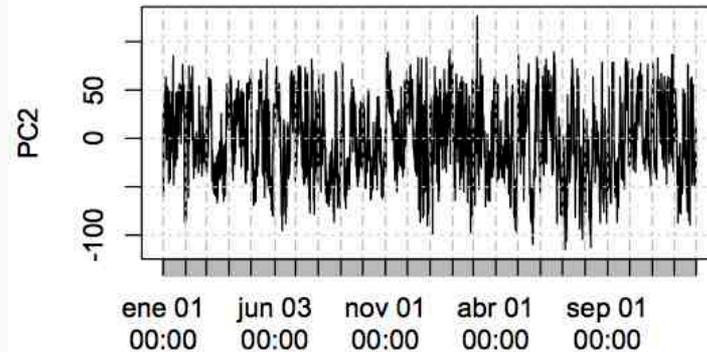
EOF: componentes espaciales



Primera Componente Principal (PC1)



Segunda Componente Principal (PC2)



Modelo de volatilidad estacional

- ❑ La componente temporal observada para la velocidad del viento sigue un patrón común a muchos fenómenos atmosféricos.
- ❑ Comportamiento de su pasado inmediato tiene un efecto sobre el presente y de carácter estacional.
- ❑ Variabilidad no constante: existen *clusters* de volatilidad.
- ❑ Dicha volatilidad tiene un comportamiento estacional.
- ❑ Colas pesadas en la distribución de la velocidad del viento: el supuesto de normalidad no aplicable a los datos.
- ❑ Necesidad de extender los modelos de volatilidad condicionada incorporando un polinomio estacional multiplicativo y flexibilizando la distribución del error (T-student).

Modelo de volatilidad estacional

- **Nuestra aportación:** Implementación en R de rutinas que estiman, simulan y predicen un SARMA-SGARCH con distribución del error Gaussiano o T-student.
- Extiende la especificación SARMA-GARCH (Engle & Bollerslev 1986) incorporando un polinomio estacional multiplicativo en la volatilidad.
- Especificación del modelo SARMA(P,Q)x(P',Q')-SGARCH(p,q)x(p',q'):

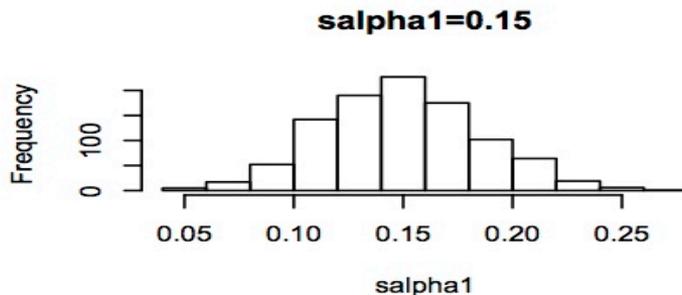
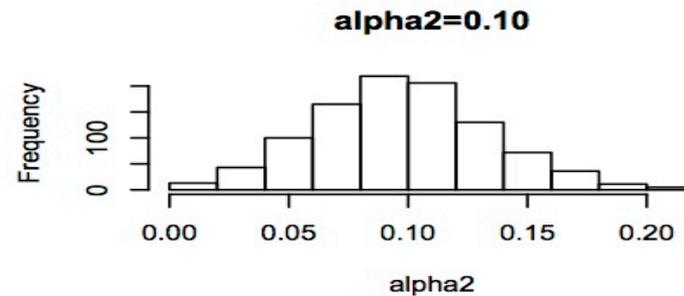
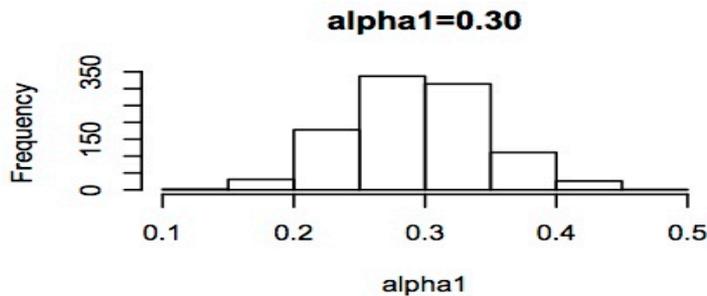
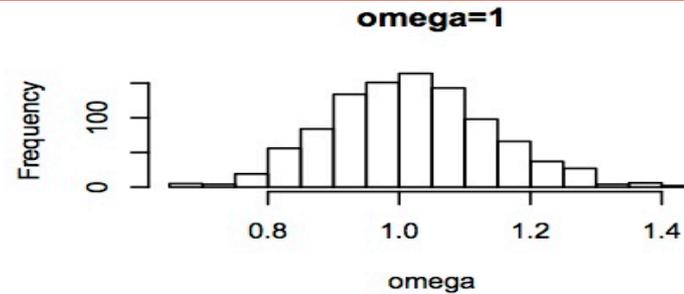
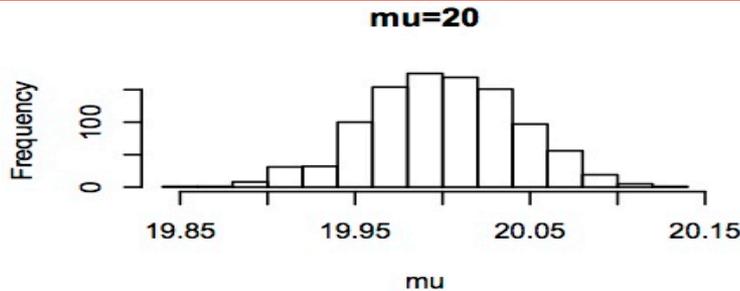
$$\Phi(B)\Phi_s(B)Y_t = \mu + \Theta(B)\Theta_s(B)a_t$$

$$a_t = \sigma_t \varepsilon_t$$

$$\Omega_s(B)\Omega(B)\sigma_t^2 = w + [\Psi_s(B)\Psi(B) - 1]a_t^2$$

- Estimación de los parámetros y grados de libertad de la t-student maximizando la función de verosimilitud condicionada (Tsay, 2010)

Modelo de volatilidad estacional



$$Y_t = 20 + a_t$$

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = 1 + \left[\left(1 + 0.15B^8 \right) \left(1 + 0.30B + 0.10B^2 \right) - 1 \right] a_t^2$$

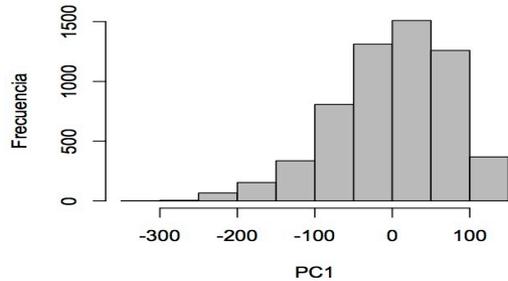
Simulación (1000 x 1000)
de un modelo
SGARCH(0,2)x(0,1)^{s=8}

Índice

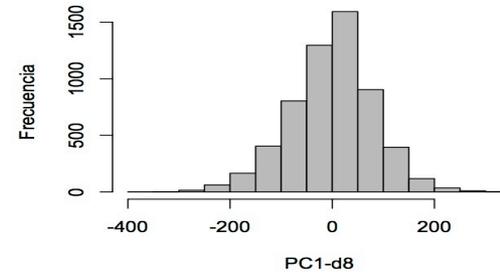
- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ Análisis espacial
- ❑ Antecedentes
- ❑ Esquema metodológico
 - ❑ Reducción de la dimensión (EOFs)
 - ❑ Modelos de volatilidad estacional e implementación en R
- ❑ **Aplicación: estimación y predicción de la velocidad del viento**
- ❑ Conclusiones
- ❑ Futuras líneas de trabajo

Aplicación: estimación y predicción de la velocidad del viento

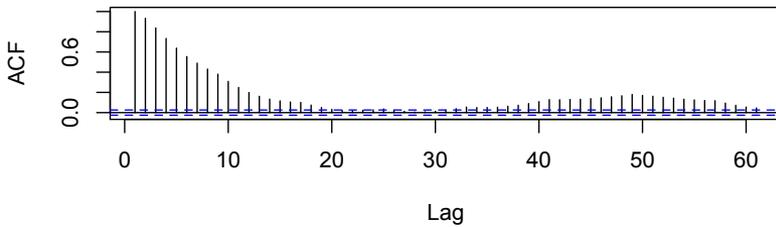
Distribución 1ª componente principal (PC1)



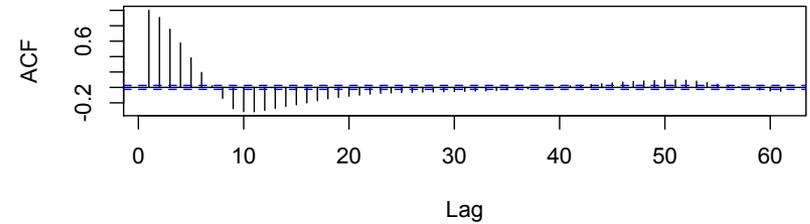
Distribución PC1 diferenciada de orden 8



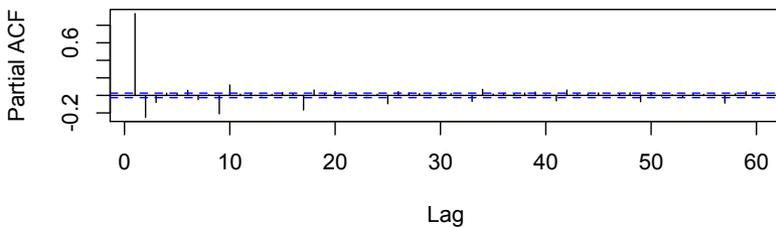
ACF PC1



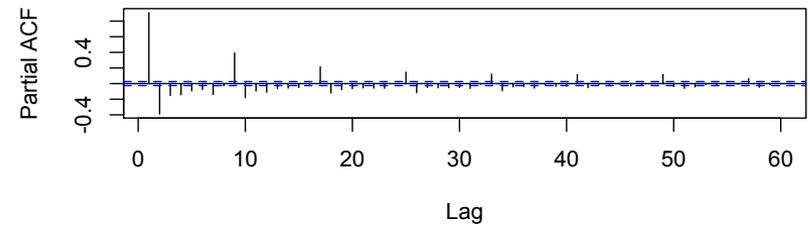
ACF PC1-d8



PACF PC1



PACF PC1-d8



Aplicación: estimación y predicción de la velocidad del viento

- Modelo PC1-d8: **SARMA(1,7)x(0,3)^{s=8}-SGARCH(0,4)x(0,5)^{s=8}.**

$$\Phi(B)Y_t = \mu + \Theta(B)\Theta_s(B)a_t$$

$$a_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = w + [\Psi_s(B)\Psi(B) - 1]a_t^2$$

$$\varepsilon_t \sim t - student(gl)$$

donde

$$\Phi(B) = (1 + \phi_1 B)$$

$$\Theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \dots + \theta_7 B^7)$$

$$\Theta_s(B) = (1 + \theta_1^s B^8 + \theta_2^s B^{16} + \theta_3^s B^{24})$$

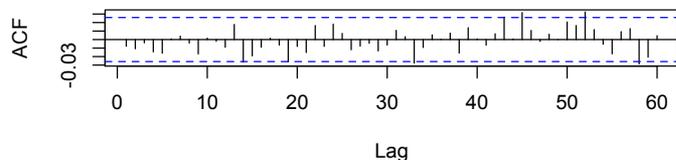
$$\Psi(B) = (1 + \alpha_1 B + \alpha_2 B^2 + \alpha_3 B^3 + \alpha_4 B^4)$$

$$\Psi_s(B) = (1 + \alpha_1^s B^8 + \alpha_2^s B^{16} + \dots + \alpha_5^s B^{40})$$

Parámetros	Estimador	Error estándar	Estadístico t
mu	0,2746	0,1198	2,2932
ar1	0,7690	0,0361	21,2799
ma1	0,4682	0,0389	12,0291
ma2	0,3354	0,0462	7,2524
ma3	0,2621	0,0473	5,5438
ma4	0,1877	0,0463	4,0539
ma5	0,1269	0,0423	2,9957
ma6	0,1344	0,0368	3,6497
ma7	0,1166	0,0330	3,5393
sma1	-0,7975	0,0315	-25,3018
sma2	-0,0860	0,0273	-3,1458
sma3	-0,0788	0,0144	-5,4756
w	179,4784	17,8260	10,0683
alpha1	0,1368	0,0191	7,1581
alpha2	0,0549	0,0152	3,6090
alpha3	0,0312	0,0123	2,5345
alpha4	0,0481	0,0154	3,1343
salpha1	0,1254	0,0191	6,5690
salpha2	0,0620	0,0158	3,9265
salpha3	0,0326	0,0140	2,3341
salpha4	0,0542	0,0153	3,5435
salpha5	0,0426	0,0138	3,0889
gl	9,0432	0,9422	9,5981

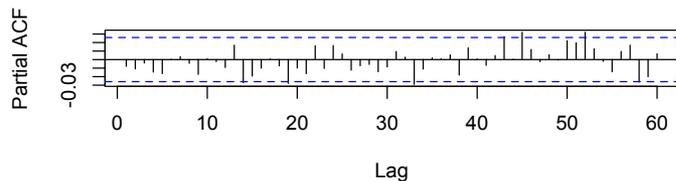
Aplicación: estimación y predicción de la velocidad del viento

ACF residuos del modelo

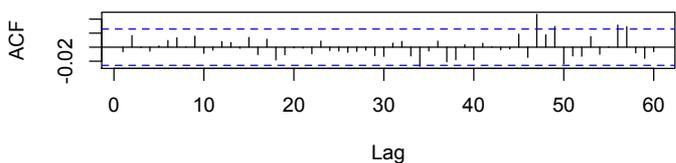


Test Box-Ljung sobre los residuos:
• No rechaza H_0 de ruido blanco (p-valor:0,84)

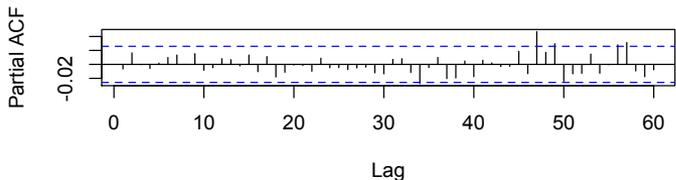
PACF residuos del modelo



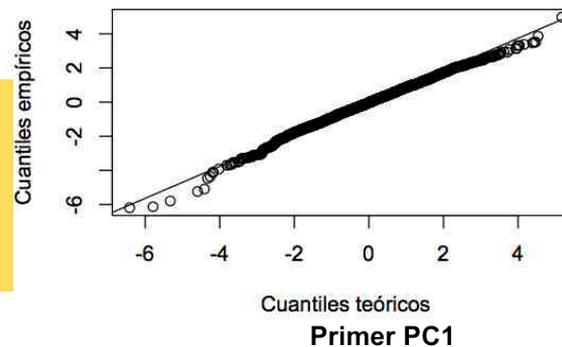
ACF residuos al cuadrado



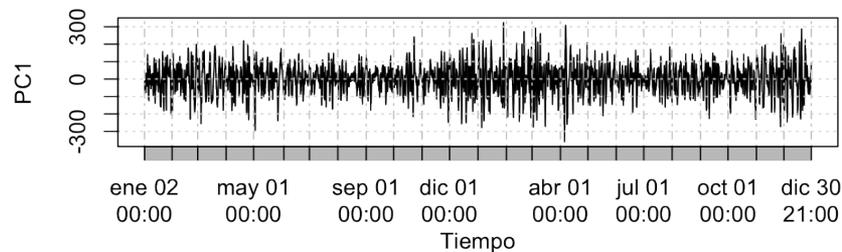
PACF residuos al cuadrado



QQ-Plot Distribución t-student (9 g.l.)

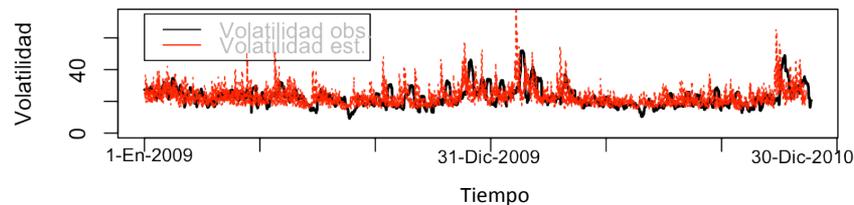


Los datos se ajustan a los cuantiles teóricos de la T-student (9)



Volatilidad observada: suavizada con media móvil de 40 observ.

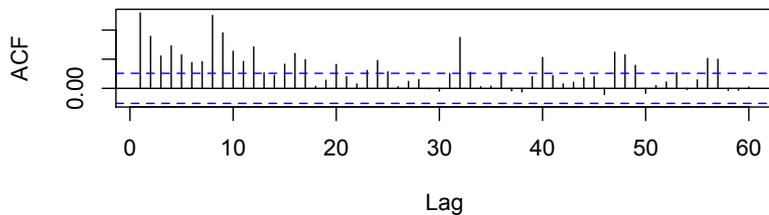
Volatilidad estimada y observada



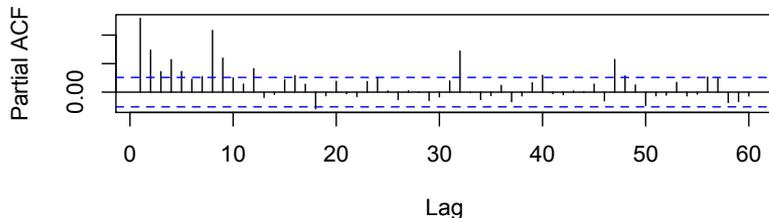
Aplicación: estimación y predicción de la velocidad del viento

- ❑ ACF / PACF residuos al cuadrado sin especificar un proceso SGARCH en la volatilidad: autocorrelaciones significativas y con un patrón estacional de periodicidad igual a 8.
- ❑ El supuesto de normalidad no se ajusta a los datos: colas pesadas: T-student con 9 grados de libertad presenta mejores resultados.

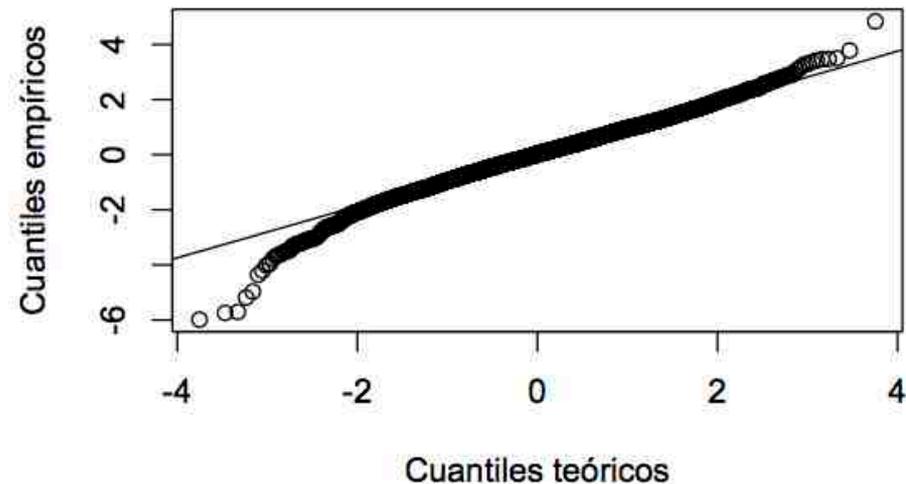
ACF residuos al cuadrado modelo sin SGARCH



PACF residuos al cuadrado modelo sin SGARCH



QQ-Plot Distribución Normal



Aplicación: estimación y predicción de la velocidad del viento

- Predicción de la velocidad del viento en $t+1$ (+3 horas) para cada punto de la malla de datos.
- Pasos:
 - 1) Predicción en $t+1$ de las primeras 6 componentes principales: $y_k(t+1)$.
 - 2) Se obtiene la predicción en $t+1$ en cada localización espacial, $X(t+1, s)$:

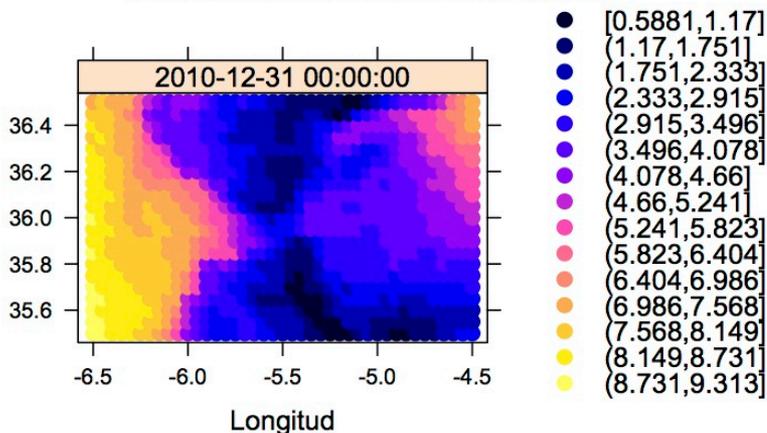
$$X(t+1, s) = \sum_{k=1}^M y_k(t+1) u_k(s)$$

donde $u_k(s)$ es la matriz de EOFs espaciales correspondientes a los vectores propios de la matriz de covariancias.

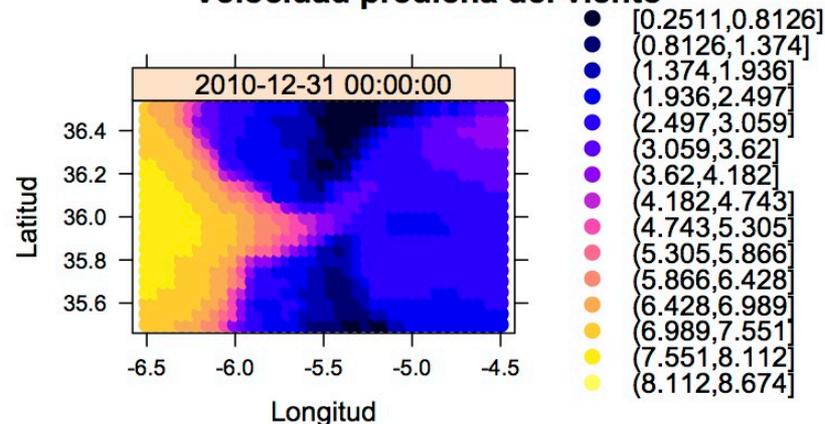
- El día 31 de diciembre de 2010 (8 observaciones) no han entrado en la estimación del modelo: se utiliza para comparar la predicción vs la observación real.

Aplicación: estimación y predicción de la velocidad del viento

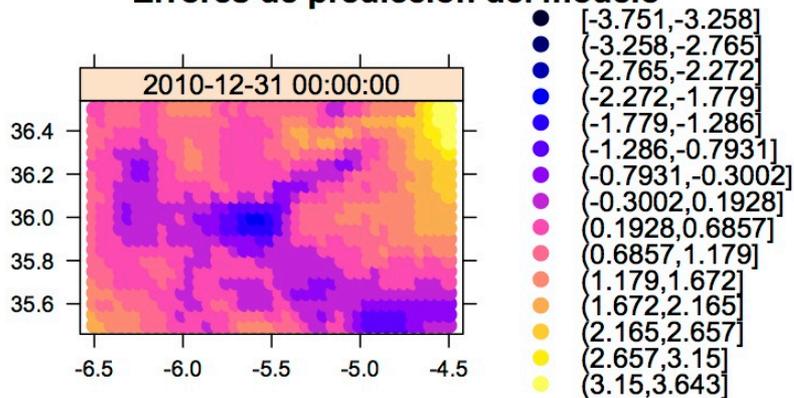
Velocidad observada del viento



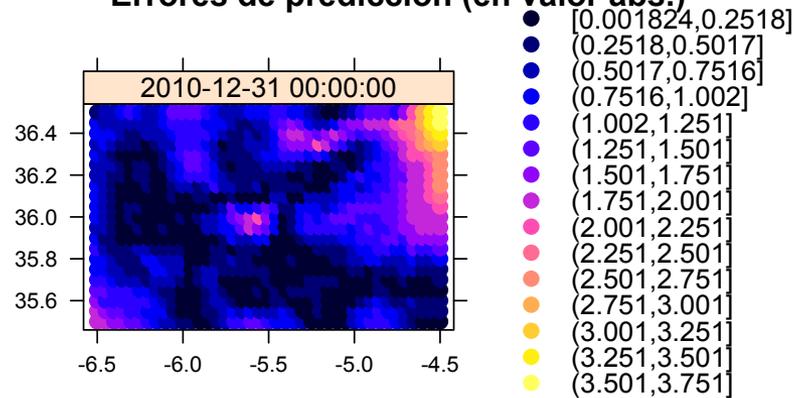
Velocidad predicha del viento



Errores de predicción del modelo

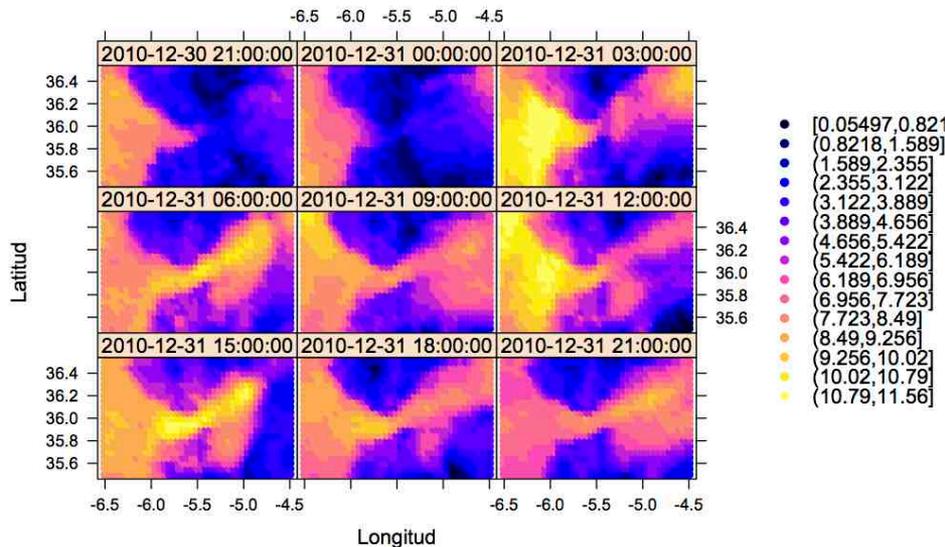


Errores de predicción (en valor abs.)

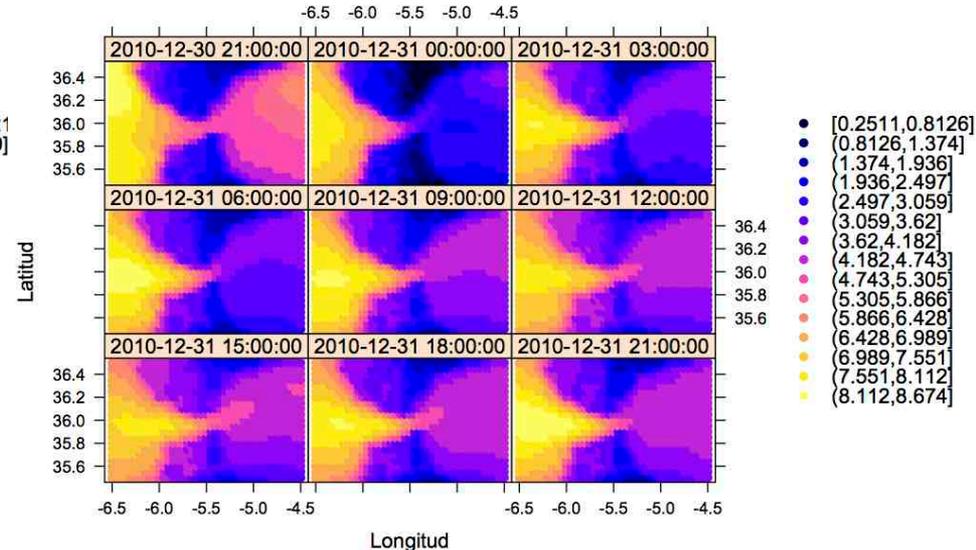


Aplicación: estimación y predicción de la velocidad del viento

Velocidad observada del viento



Predicción velocidad del viento



- ❑ La predicción a 3 horas (día 31 00 horas): errores de predicción bajo.
- ❑ A medida que aumenta el horizonte de predicción, las predicciones van empeorando paulatinamente: los errores se van acumulando.
- ❑ En la práctica, se debe encontrar un equilibrio entre el coste de ir actualizando la información versus el beneficio de predicciones más precisas.

Índice

- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ Análisis espacial
- ❑ **Antecedentes**
- ❑ Esquema metodológico
 - ❑ Reducción de la dimensión (EOFs)
 - ❑ Modelos de volatilidad estacional e implementación en R
- ❑ Aplicación: estimación y predicción de la velocidad del viento
- ❑ **Conclusiones**
- ❑ Futuras líneas de trabajo

Conclusiones

- Estimación modelo estadístico: predicciones más precisas que es vital, por ejemplo, en la planificación energética (minimiza el déficit eólico o el exceso).
- Cálculo de las EOFs: reducción de la dimensión de los datos (inicialmente matriz 861x5.824!!) y separación componente temporal y espacial.
→ ha facilitado el procesamiento de la información y la modelización.
- Componente temporal:
 - Patrón estacional en la media condicionada del proceso ($s=8$).
 - Colas pesadas (no gaussiano): la t-student mejores resultados.
 - Fuerte volatilidad (variancia no constante) y de carácter estacional
- La implementación en R y estimación de modelos de volatilidad estacional (SARMA-SGARCH) ha permitido una correcta validación del modelo a partir de los errores (y errores al cuadrado) así como estimar el proceso de volatilidad de los datos.

Índice

- ❑ Objetivo / Motivación
- ❑ Descripción de los datos
- ❑ Análisis espacial
- ❑ Antecedentes
- ❑ Esquema metodológico
 - ❑ Reducción de la dimensión (EOFs)
 - ❑ Modelos de volatilidad estacional e implementación en R
- ❑ Aplicación: estimación y predicción de la velocidad del viento
- ❑ Conclusiones
- ❑ **Futuras líneas de trabajo**

Futuras líneas de trabajo

- ❑ Se ha constatado que la velocidad del viento viene explicada por otras variables (presión, temperatura,...): explorar técnicas EOF multivariantes (Kutzbach 1967 o Bretherton et al. 1992).
- ❑ Se ha calculado la correlación cruzada entre la volatilidad estimada entre las dos primeras componentes: existen correlaciones significativas. Explorar modelos de volatilidad multivariantes (Meyer & Yu 2006).
- ❑ En el trabajo actual, no se ha realizado una predicción espacial fuera de los puntos observados de la malla: aplicar técnicas de Kriging a partir de la componente espacial analizada en el presente estudio.
- ❑ Librería en R con la implementación del SARMA-SGARCH que se ha realizado en el actual trabajo.

Q & A

