

SIPEI

Sistema de Predicción Estadística de Inmisión.



Santiago de Compostela, 10/07/2009

Central Térmica de As Pontes

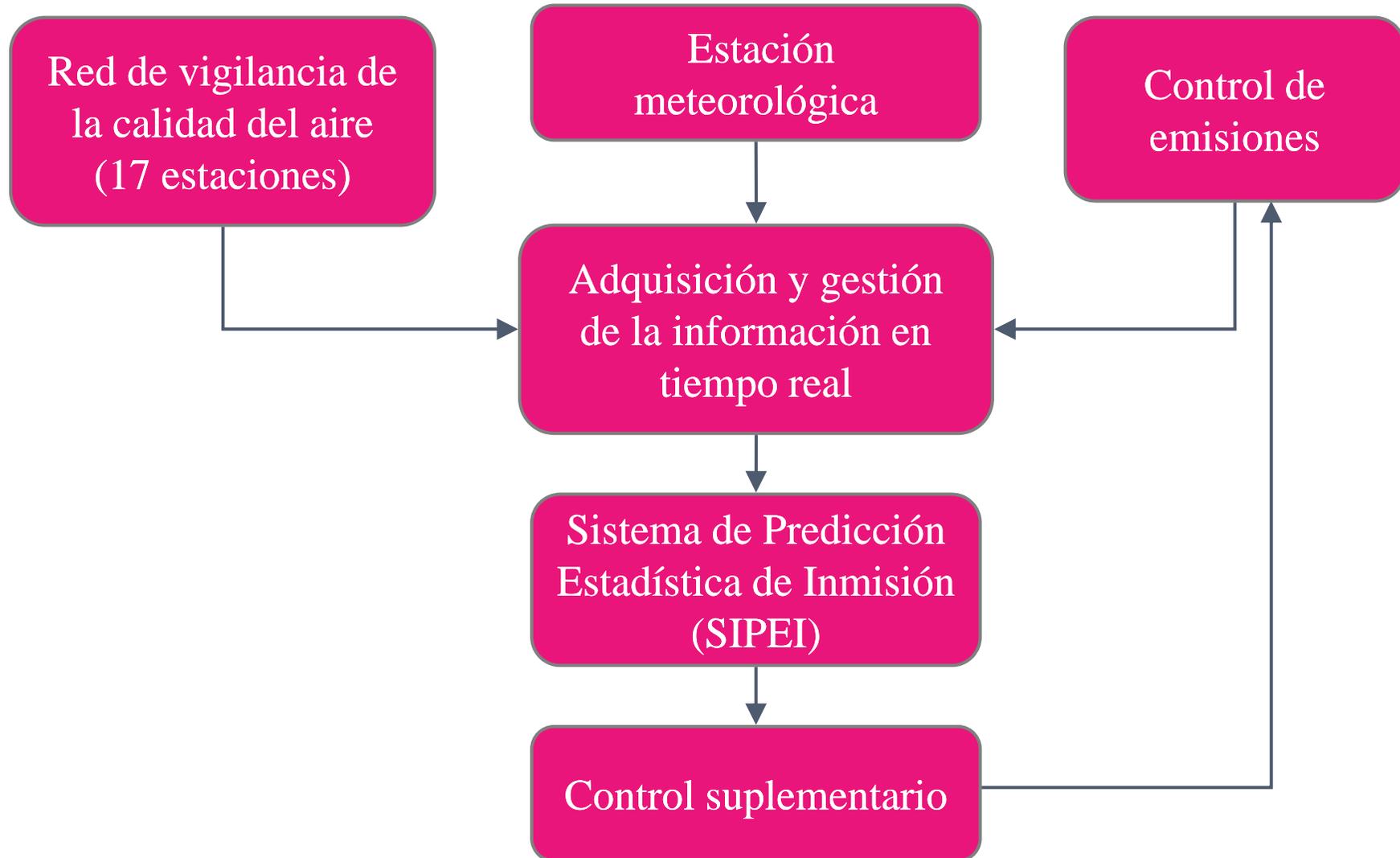
- La Unidad de Producción Térmica (UPT) de As Pontes constituye uno de los centros productivos propiedad de Endesa Generación S.A., situado en el municipio de As Pontes de García Rodríguez, al noroeste de la provincia de A Coruña.



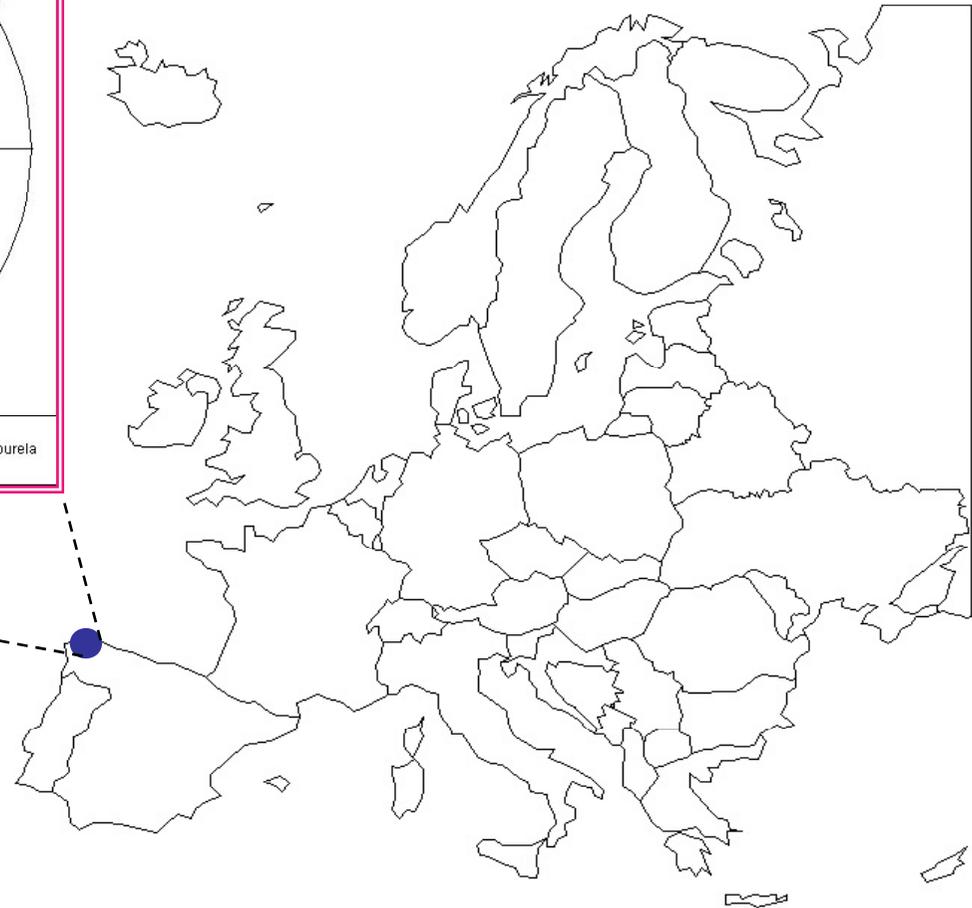
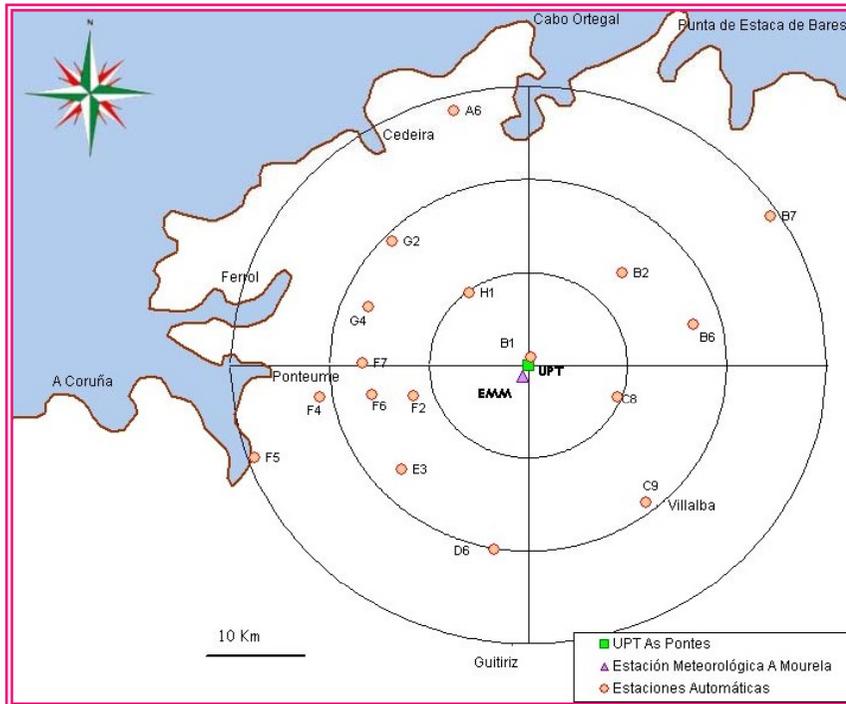
Descripción general

- ❑ Inició su actividad en 1976 con la puesta en marcha de un grupo de generación de energía, disponiendo en la actualidad de cuatro.
- ❑ Fue diseñada para utilizar los lignitos extraídos de la mina a cielo abierto situada en sus proximidades con alto contenido de azufre.
- ❑ En el período 1993-1996 fue transformada con el objetivo de utilizar mezclas de lignito local con carbones subbituminosos de importación caracterizados por sus bajos contenidos en azufre.
- ❑ Actualmente ya se ha finalizado una nueva adaptación para consumir carbón subbituminoso de importación como combustible principal (período de transformación: 2005-2008).

Sistema de Seguimiento y Control de la Calidad Atmosférica



Sistema de Seguimiento y Control de la Calidad Atmosférica



Problema



- ❑ Existe un foco emisor donde se monitoriza en continuo distintas medidas de calidad de aire en su entorno, e interesa predecir valores futuros de la calidad de aire. El principal interés inicial es la predicción de valores futuros de SO₂.
- ❑ Los modelos estadísticos de predicción son una herramienta eficaz para obtener estas predicciones y sugerir una línea de actuación para intentar evitar los episodios de calidad de aire.

Nueva problemática

- ❑ Cercano al emplazamiento de la Central Térmica de As Pontes se ha construido una nueva Central de Ciclo Combinado de gas natural.



Descripción general

- ❑ Consiste en un grupo generador de electricidad formado por dos turbinas de gas y una turbina de vapor (dos focos emisores).
- ❑ Está diseñado para utilizar gasóleo como combustible de emergencia.
- ❑ Como cualquier instalación de generación de energía por combustión debe disponer de un Sistema de Control de las Emisiones que ha sido integrado en el Sistema de Seguimiento y Control de la Calidad Atmosférica .

Nueva problemática

En el período 2007-2008 las emisiones se han visto afectadas por dos actuaciones clave:

□ La Central Térmica actual consume únicamente carbón de importación.



Disminución de las emisiones de SO₂ (alrededor del 95%)

□ La nueva Central de Ciclo Combinado entró en funcionamiento.



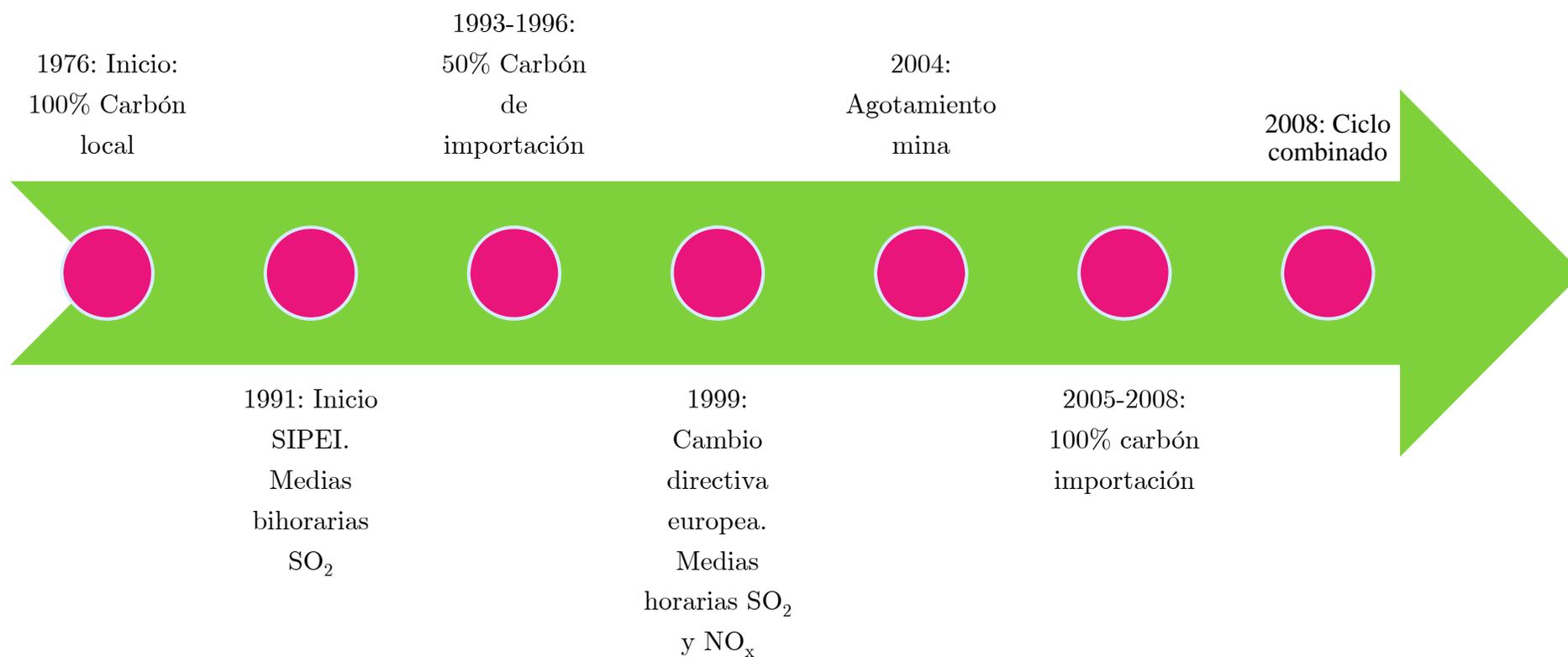
Necesidad de predecir las emisiones de NO_x

Nuevo problema



- ❑ Existen tres focos emisores donde se monitorizan en continuo distintas medidas de calidad de aire en su entorno, e interesa predecir valores futuros de la calidad de aire. El principal interés es la predicción de valores futuros de NO_x y SO_2 .
- ❑ Los modelos estadísticos de predicción vuelven a ser una herramienta clave para evitar los episodios de calidad de aire.

Resumen histórico



Los datos

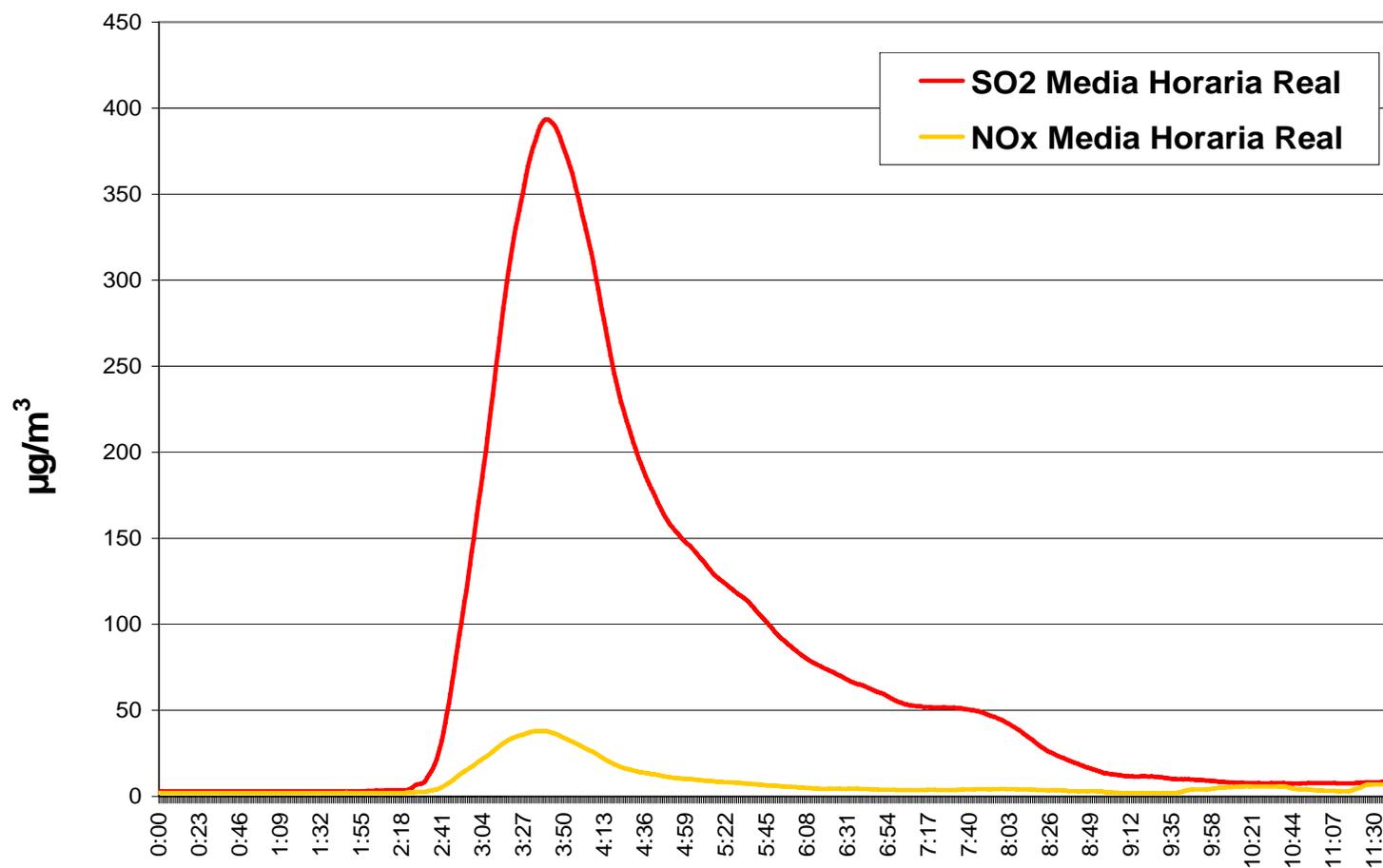
- Actualmente se reciben datos con frecuencia minutal y pentaminutal en tiempo real.
 - *Estaciones de inmisión:* SO₂, NO_x, CO, O₃, PM₁₀ , PM_{2.5} y algunas variables meteorológicas.
 - *Grupos de emisión (Central Térmica):* SO₂, NO_x, partículas, O₂, potencia, caudal y temperatura.
 - *Turbinas de gas (Ciclo Combinado):* SO₂, NO_x, partículas, CO, COV's, O₂, potencia de la turbinas de gas, potencia de la turbina de vapor, caudal y temperatura.
 - *Estación Meteorológica:* temperatura, humedad, velocidad y dirección del viento, radiación solar, ...

Los datos

- ❑ Se va a considerar la media horaria arrastrada, tanto de los valores de SO_2 como de NO_x , para obtener las predicciones de valores futuros.
- ❑ La serie de valores medios horarios de SO_2 tiene un comportamiento muy característico:
 - ❑ Toma valores próximos a cero durante largos períodos de tiempo.
 - ❑ En ocasiones crece de manera brusca y repentina (episodios).
 - ❑ Estos episodios están muy espaciados en el tiempo.
- ❑ La serie del NO_x tiene un comportamiento similar pero a menor escala.

Los datos

Valores de SO₂ y NO_x .12/03/2007



Modelos de predicción



- Las diferentes metodologías utilizadas han aportado soluciones desde varios puntos de vista:

I. Predicción puntual

II. Predicción binaria

III. Predicción espacial

IV. Predicción funcional

I. Predicción puntual

- Este planteamiento nos permite dar como respuesta la predicción del nivel de SO₂ para un instante concreto.

- Modelos utilizados:
 - *Modelo Semiparamétrico.*

 - *Modelo Parcialmente Lineal.*

 - *Modelo de Redes Neuronales.*

Modelo Semiparamétrico

- En los primeros años de desarrollo la frecuencia de envío de datos era pentaminutal y la legislación vigente establecía límites para las medias bihorarias.
- El modelo usa la serie temporal de medias bihorarias arrastradas

$$x_t = \frac{1}{24} \sum_{i=0}^{23} SO_2(t-i)$$

donde $SO_2(t)$ representa la concentración de SO_2 en el instante t (pentaminutal), medida en $\mu\text{g}/\text{m}^3$.

- El modelo semiparamétrico generaliza los modelos Box-Jenkins de la siguiente manera:

$$X_{t+k} = \varphi(X_t, X_{t-l}) + Z_{t+k}$$

Modelo Semiparamétrico

- En cada instante t se estima la función de regresión

$E(X_{t+6} / X_t, X_{t-1})$ con el estimador tipo núcleo Nadaraya-Watson y la matriz histórica.

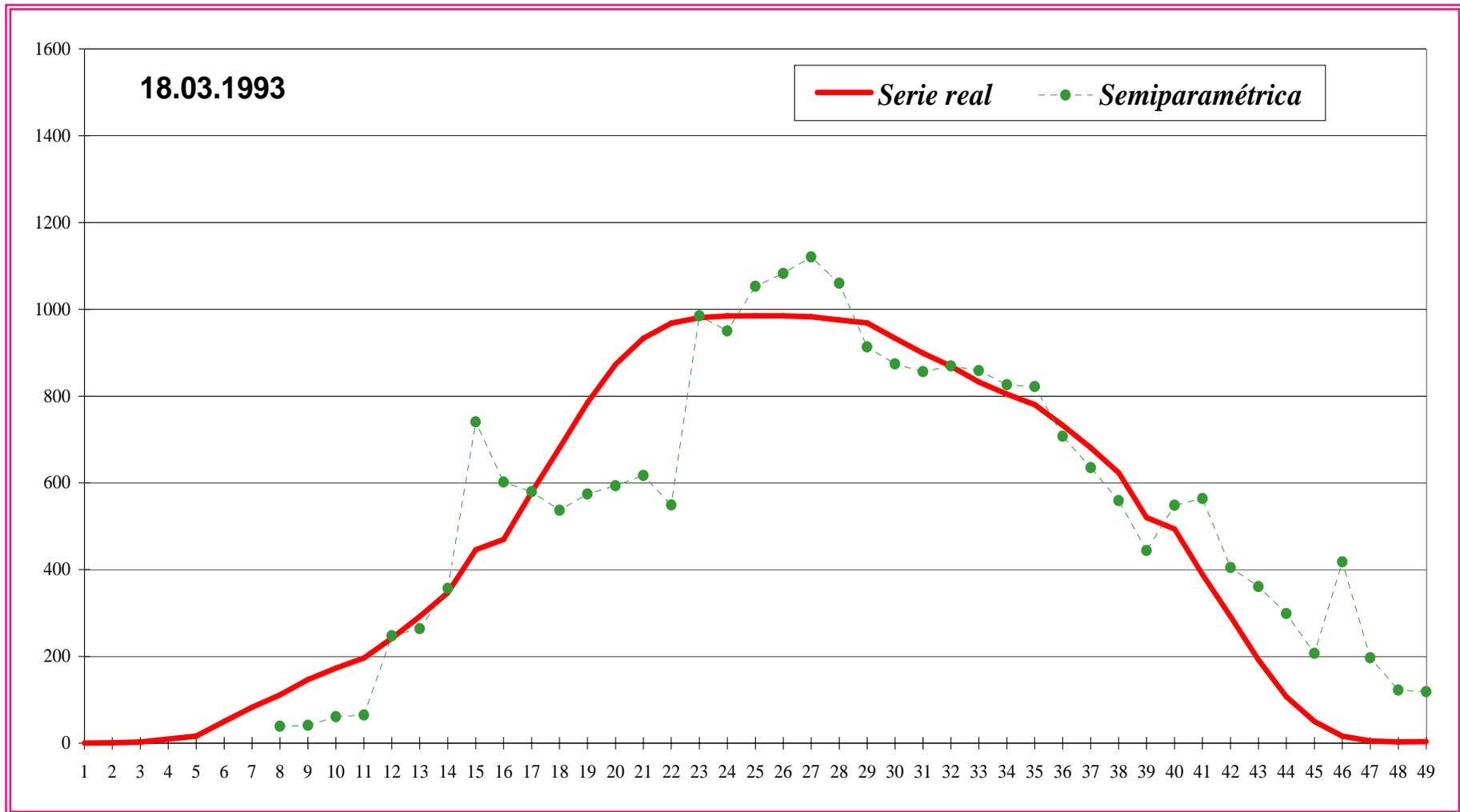
- Se calcula la serie de residuos $\hat{Z}_{t-64}, \dots, \hat{Z}_t$ (6 últimas horas), donde $\hat{Z}_i := X_i - \hat{E}(X_i / X_{i-6}, X_{i-7})$ y se ajusta un modelo ARIMA adecuado.

- Se obtiene la predicción Box-Jenkins de \hat{Z}_{t+6} .

- La predicción final propuesta es:

$$\hat{X}_{t+6} = \hat{E}(X_{t+6} / X_t, X_{t-1}) + \hat{Z}_{t+6}$$

Modelo Semiparamétrico (Matriz Histórica)



Modelo Semiparamétrico



García-Jurado I., González-Manteiga W., Prada-Sánchez J.M., Febrero-Bande M. and Cao R. *Predicting using Box-Jenkins, Nonparametric and Bootstrap Techniques*. Technometrics 1995; 37: 303-310.

Matrices Históricas

- ❑ La clave del buen funcionamiento de todos los modelos de predicción diseñados está en un **mecanismo de memoria** diseñado en los primeros años de desarrollo.
- ❑ La serie temporal de interés está formada en su mayor parte por valores próximos a cero y pocos valores de episodios de inmisión.
- ❑ Se estratifican los datos disponibles de SO_2 según un rango de valores que representan razonablemente a los episodios pasados con objeto de mantener la **información interesante** de los episodios.
- ❑ A lo largo de los años este concepto se ha adaptado a las distintas técnicas estadísticas utilizadas.

Matrices Históricas

~~Serie temporal~~

Matriz histórica :

Un número grande de registros.

Divididos en estratos.

Se asigna un rango de valores de X_{t+k} a cada clase.

Cada vector se asigna a la clase a la que pertenece X_{t+k} .

$$\left(\underline{X_{t-l}^1}, \underline{X_t^1}, \underline{X_{t+k}^1} \right)$$

$$\left(\underline{X_{t-l}^i}, \underline{X_t^i}, \underline{X_{t+k}^i} \right)$$

⋮

$$\left(\underline{X_{t-l}^M}, \underline{X_t^M}, \underline{X_{t+k}^M} \right)$$

Matrices Históricas



Prada-Sánchez, J.M. and Febrero-Bande M.
*Parametric, Non-Parametric and Mixed
approaches to prediction of sparsely
distributed pollution incidents: a case study.*
Journal of Chemometrics **1997**; 11: 13-32.

Modelos Parcialmente Lineales

- ❑ La información utilizada por los modelos semiparamétricos es el pasado de la propia serie.
- ❑ Se puede introducir información adicional: variables meteorológicas y variables de emisión.
- ❑ Los modelos parcialmente lineales utilizan estas variables y amplían el horizonte de predicción a una hora.
- ❑ Se consideran datos de la forma (V_l, Z_l, Y_l) donde V_l es un vector de variables exógenas, $Z_l = (X_l, X_{l-3})$ e $Y_l = X_{l-12}$, siendo $\{X_l\}$ la serie de medias bihorarias de SO_2 .
- ❑ Se asume que se ajustan al modelo parcialmente lineal:

$$Y_l = V_l' \beta + \varphi(Z_l) + \varepsilon_l$$

Modelos Parcialmente Lineales



Prada-Sánchez J.M., Febrero-Bande M., Cotos-Yáñez T., González-Manteiga W., Bermúdez-Cela J.L. and Lucas-Domínguez T. *Prediction of SO_2 pollution incidents near a power station using partially linear models and a historical matrix of predictor-response vectors.* *Environmetrics* 2000; 11: 209-225.

Modelos de Redes Neuronales

- ❑ La entrada en vigor de la Directiva Europea 1999/CE/30, provoca un cambio en la serie de interés, de medias bihorarias a medias horarias.

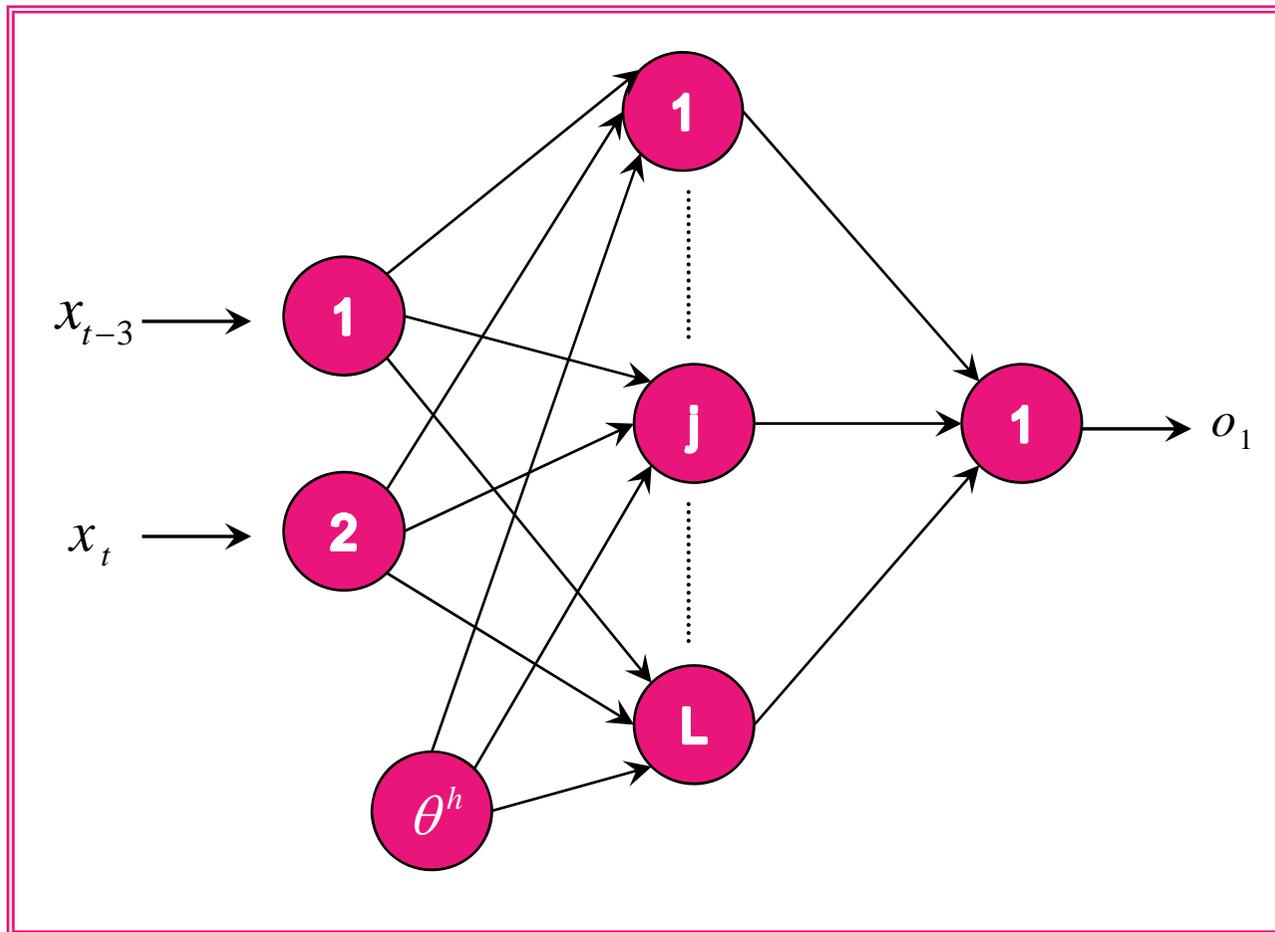
$$x_t = \frac{1}{12} \sum_{i=0}^{11} SO_2(t - i)$$

donde $SO_2(t)$ representa la concentración de SO_2 en el instante t (pentaminutal), medida en $\mu\text{g}/\text{m}^3$.

- ❑ Se adaptó el modelo semiparamétrico para trabajar con la nueva serie pero se observó un considerable aumento de la variabilidad.
- ❑ Para mejorar las predicciones se desarrollaron modelos de redes neuronales.

Modelos de Redes Neuronales

- La topología de la red diseñada es la siguiente:



Modelos de Redes Neuronales

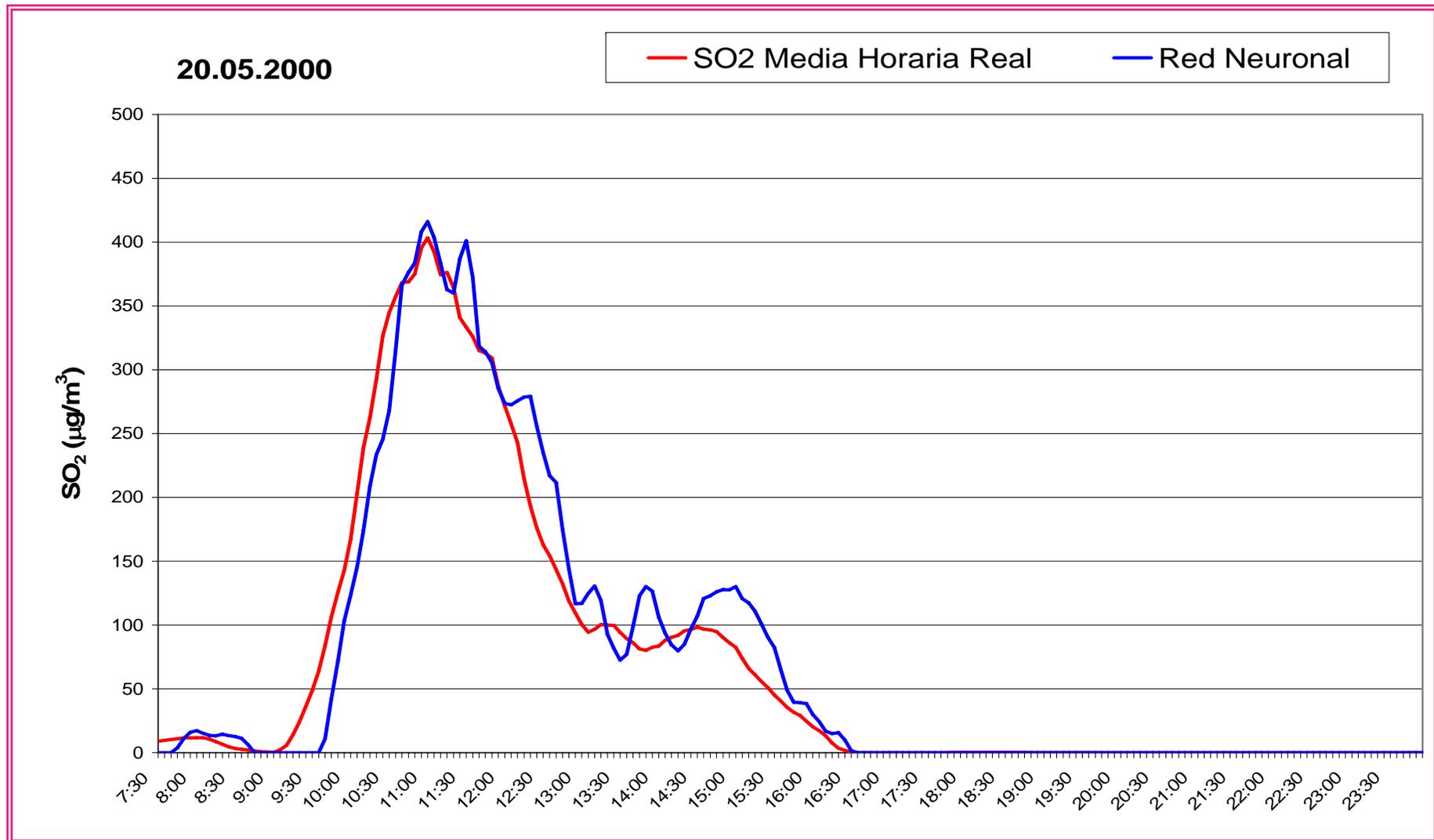
- El predictor dado por la red neuronal tiene la siguiente expresión:

$$\hat{x}_{t+6} = o_1 = \sum_{j=1}^L \omega_1^o f_j^h \left(\theta_j^h + \omega_{j1}^h x_{t-3} + \omega_{j2}^h x_t \right)$$

donde

- $f_j^h(z) = \frac{1}{1 + e^{-z}}$ para todos los nodos j de la capa oculta.
- Los pesos $\{\omega_{j1}^h, \omega_{j2}^h; j = 1, \dots, L\}$ y las tendencias $\{\theta_j^h; j = 1, \dots, L\}$ serán determinados durante el proceso de entrenamiento.
- El número L de nodos de la capa oculta se determinará en el entrenamiento como el valor cuya red neuronal proporcione mejores resultados, tras haber entrenado redes idénticas con distintos valores de L .

Modelos de Redes Neuronales



Modelos de Redes Neuronales



Fernández de Castro B.M., Prada-Sánchez J.M., González-Manteiga W., Febrero-Bande M., Bermúdez-Cela J.L. and Hernández Fernández J.J. *Prediction of SO_2 levels using neural networks.* Journal of the Air and Waste Management Association **2003**; 53: 532-538.

II. Predicción binaria

- El objetivo de estos modelos es estimar la probabilidad de que la serie de medias bihorarias de SO₂ supere un cierto nivel con una hora de antelación.

- Se pretende predecir:

$$p(Z_t) = p(X_{t+12} > r \mid Z_t)$$

donde $Z_t = (X_t, X_t - X_{t-3})$

- Para ello se van a utilizar *modelos aditivos generalizados con función link desconocida (G-GAM)*

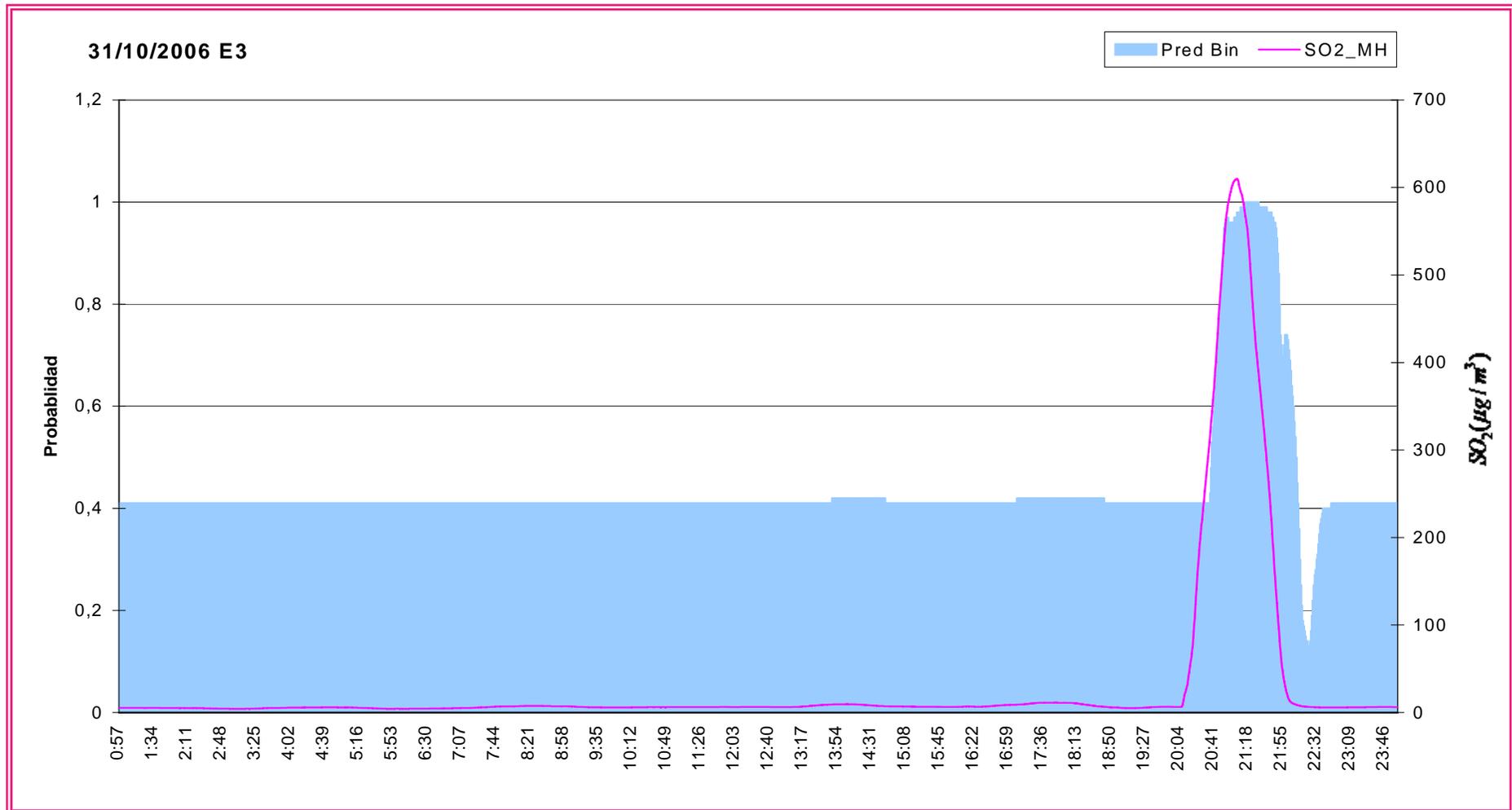
Modelos G-GAM

- Sea Y una variable con respuesta binaria, y $Z=(1, Z_1, \dots, Z_p)$ el vector de covariables asociado.
- Se denota $p(Z) = p(Y = 1 | Z)$.
- El *modelo G-GAM* tiene la siguiente expresión:

$$p(Z) = H(\eta_z) = H\left(\beta_0 + \sum_{j=1}^p f_j(Z_j)\right)$$

donde f_j son funciones suaves desconocidas y la función link H es monótona creciente.

Modelos G-GAM



Modelos G-GAM



Roca-Pardiñas, J., González-Manteiga W., Febrero-Bande M., Prada-Sánchez J.M. and Cadarso-Suárez C. *Predicting binary time series of SO_2 using generalized additive models with unknown link function.* Environmetrics **2004**; 15: 729-742.

Modelos GAM con interacción de segundo orden

- También se han considerado modelos GAM incluyendo términos de interacción de segundo orden

$$p(Z) = H(\eta_z) = H\left(\beta_0 + \sum_{j=1}^p f_j(Z_j) + \sum_{1 \leq j < k \leq p} f_{jk}(Z_j, Z_k)\right)$$

donde f_j son funciones unidimensionales desconocidas, las

$\{f_{jk}\}_{1 \leq j < k \leq p}$ son un conjunto de funciones bidimensionales

desconocidas y H es la función link monótona y conocida.

- Se ha visto que el GAM con interacciones detecta el inicio de los episodios con mayor antelación de lo que lo hace el GAM puro.

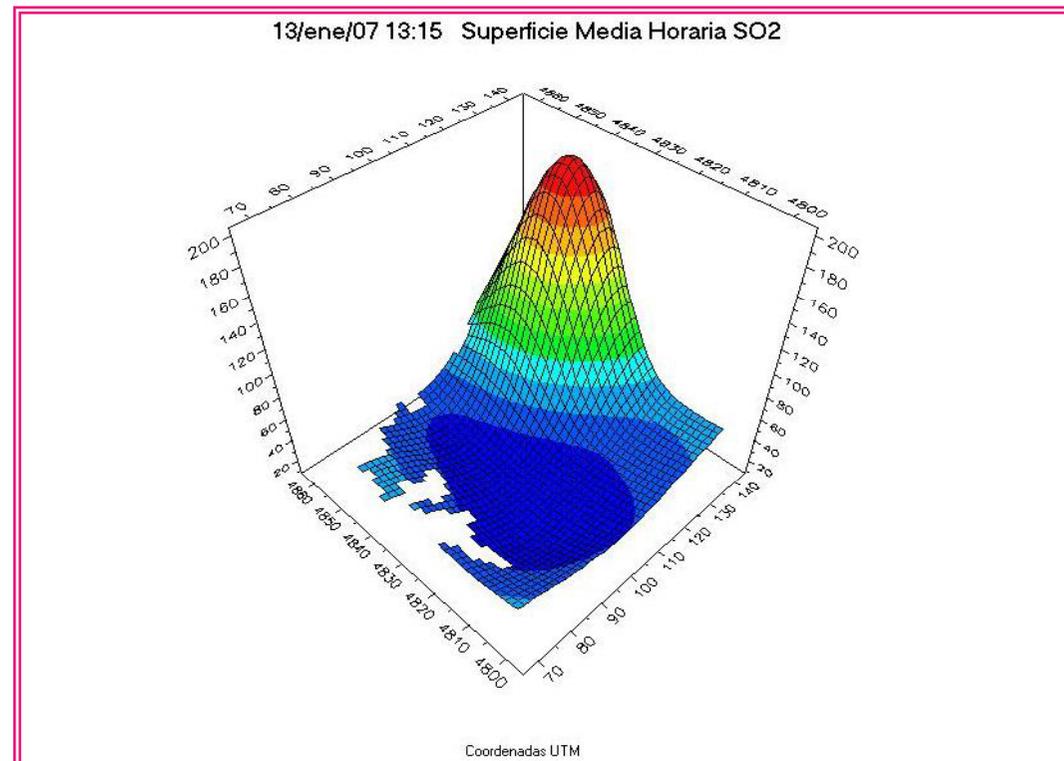
Modelos GAM con interacción de segundo orden



Roca-Pardiñas, J., Cadarso-Suárez C. and González-Manteiga W. *Testing for interactions in generalized additive models: Application to SO₂ pollution data.* Statistics and Computing **2005**; 15: 289-299.

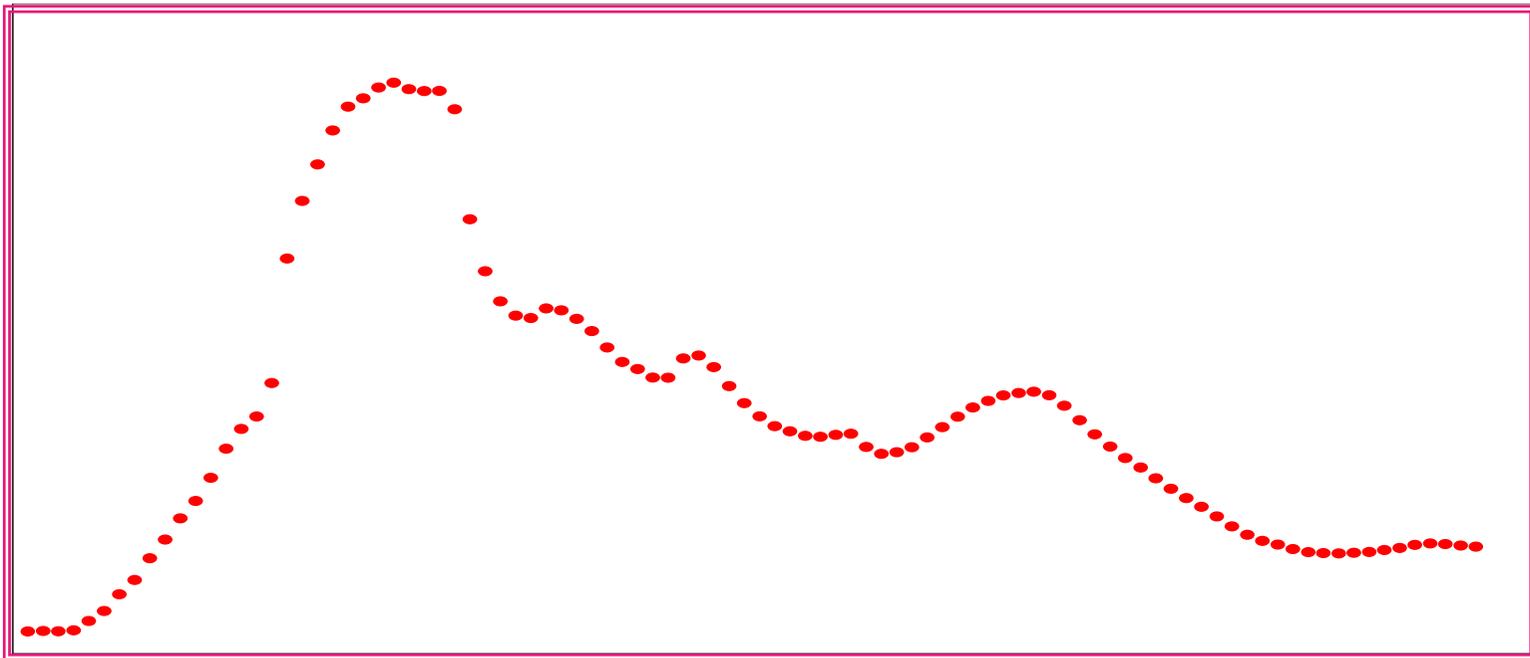
III. Predicción espacial

- ❑ Los modelos de predicción espacial nos permiten construir una superficie de predicción de niveles de SO₂ para el entorno de la instalación.



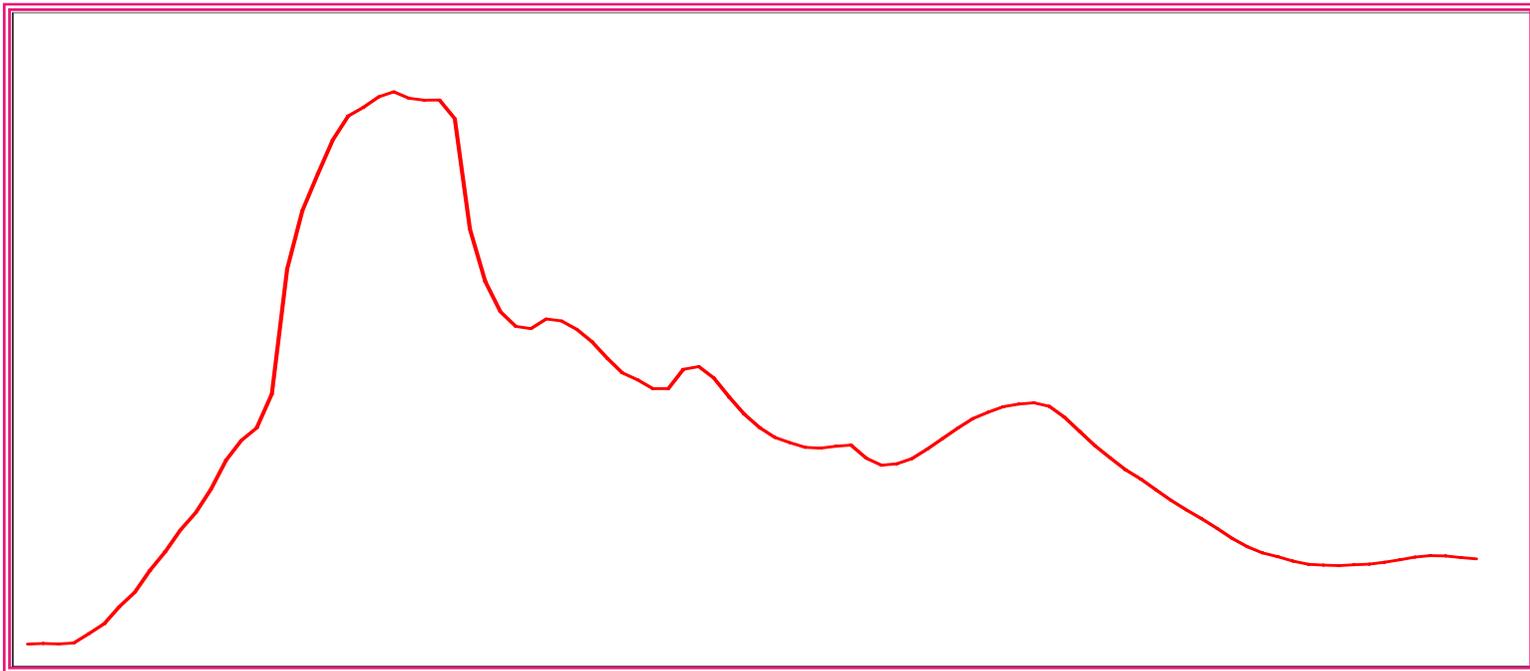
IV. Predicción funcional

□ Idea central:



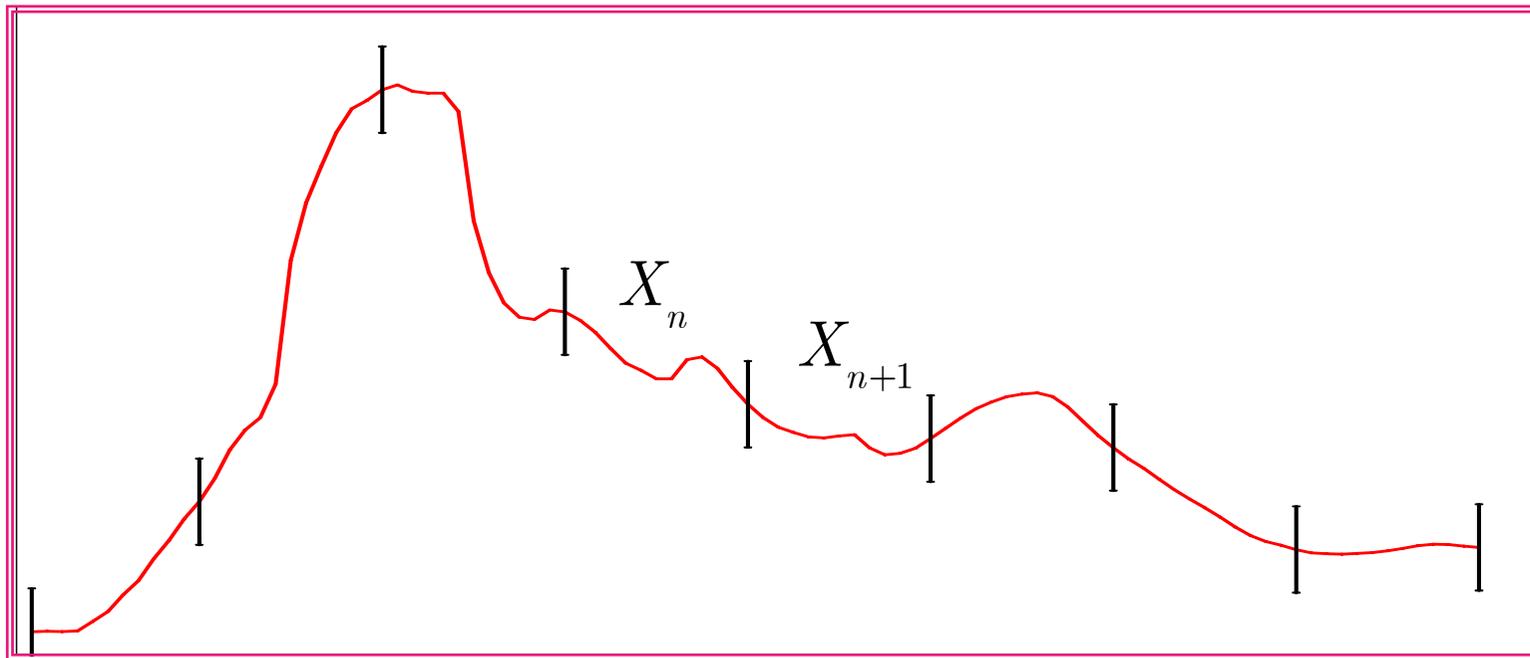
Modelos funcionales

□ Idea central:



Modelos funcionales

□ Idea central:



Modelos funcionales

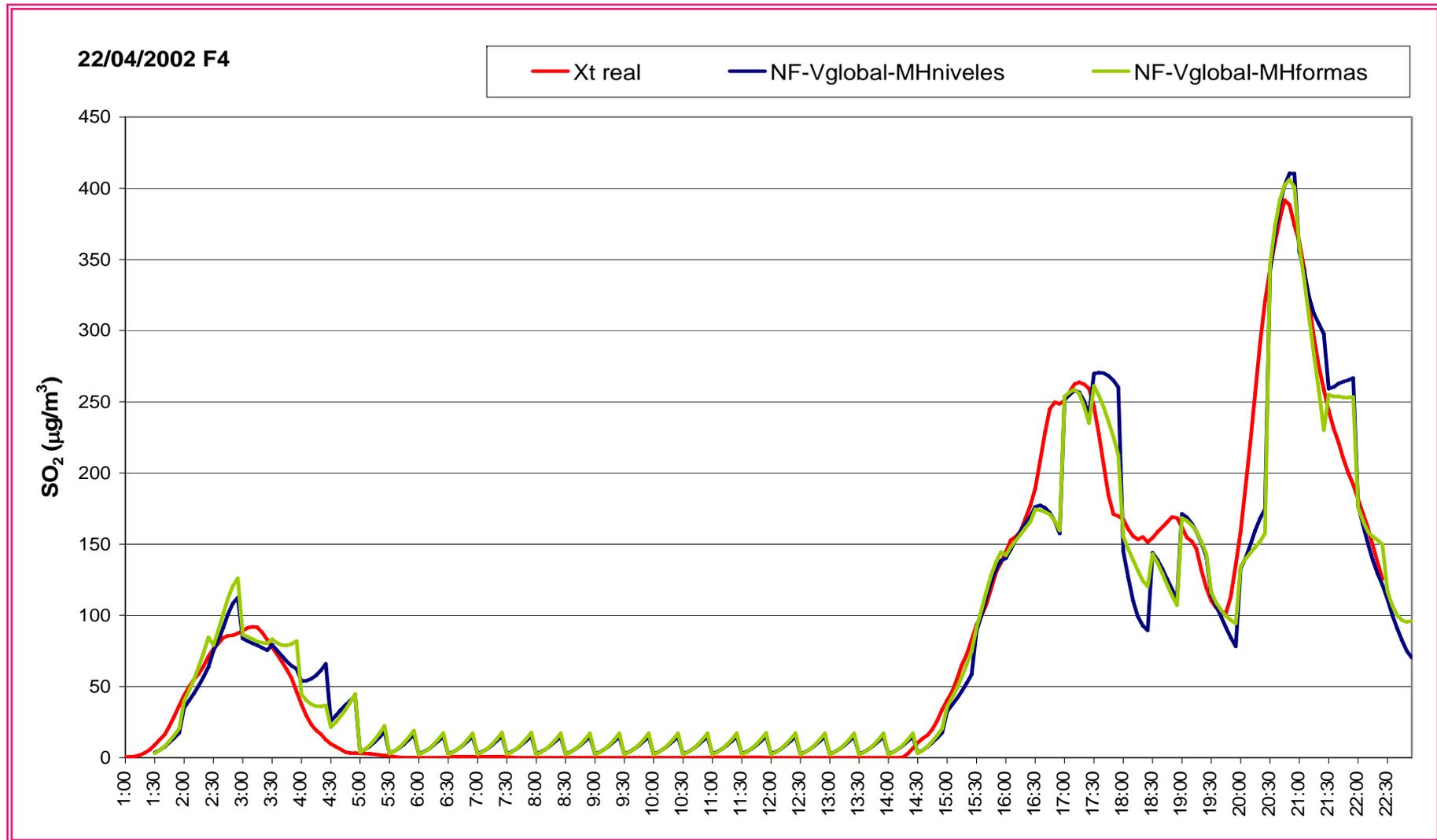
- Se considera la serie de datos como observaciones de un proceso estocástico en tiempo continuo que modeliza los niveles medios horarios de SO_2 .
- Se consideran porciones de dicho proceso estocástico que representan media hora.
- En consecuencia se consideran variables aleatorias que toman valores en

$$H = L^2([0, 6])$$

de la forma:

$$X_n(u) = x(6n + u)$$

Modelos funcionales

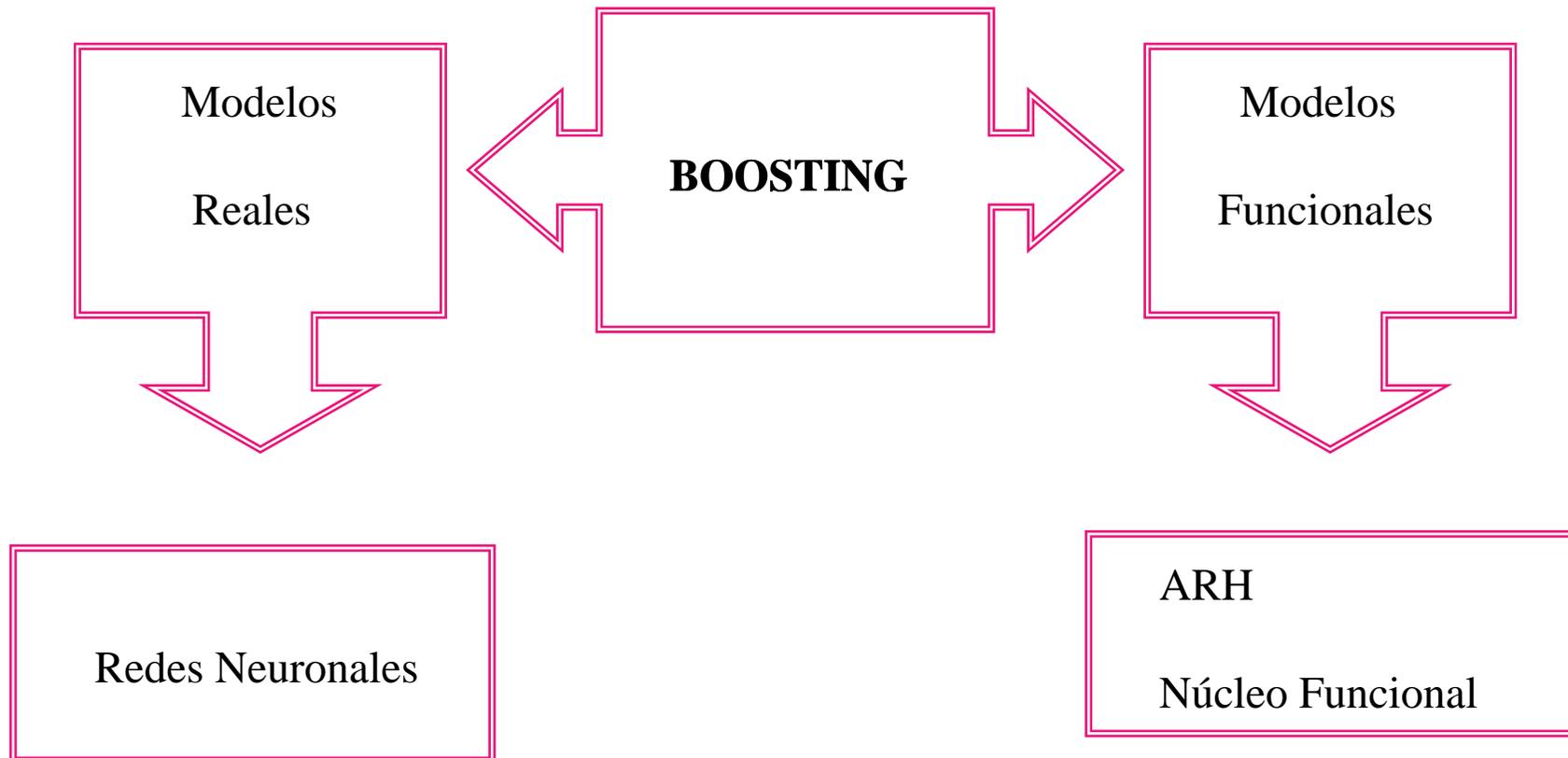


Modelos funcionales



Fernández de Castro B.M., Guillas S. and González-Manteiga W. *Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels.* Technometrics **2005**. 47: 212-222.

Boosting

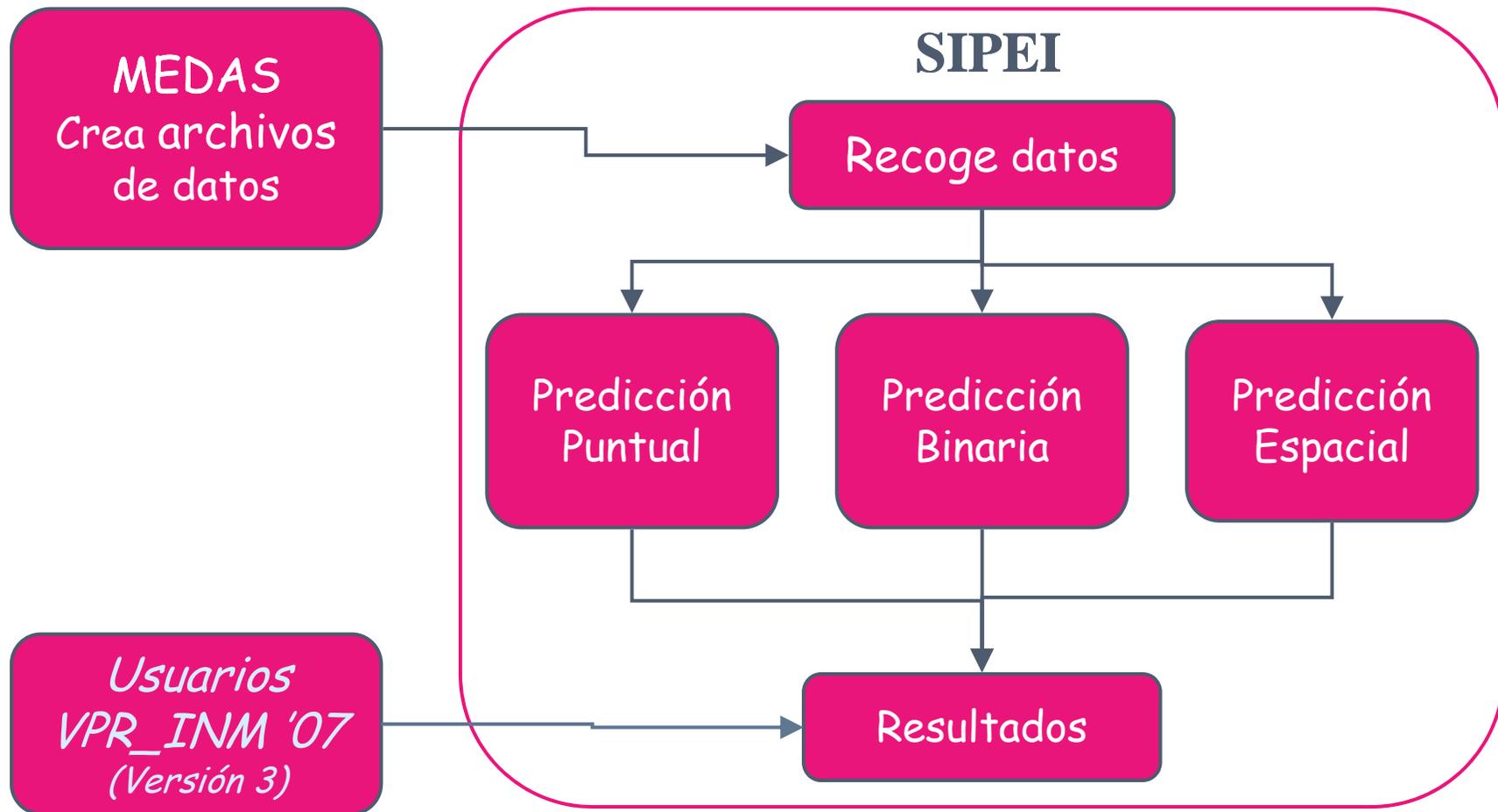


Boosting



Fernández de Castro B.M. and González-Manteiga W.
Boosting for real and functional samples: an application to an environmental problem.
Stochastic Environmental Research and Risk Assessment
2008. 22: 27-37.

Integración de los sistemas de predicción



Nuevas aportaciones a la predicción

- El nuevo objetivo del SIPEI es modelar el NO_x para poder predecir sus valores con media hora de antelación.
- Por otra parte, también será necesario desarrollar metodologías que permitan discernir cuál es el origen de los episodios de alteración de calidad de aire:
 - La Central Térmica actual.
 - El nuevo Ciclo Combinado
 - Otros posibles focos, como por ejemplo el tráfico.

Modelos aditivos

- Se han desarrollado nuevos modelos que obtienen las predicciones de los niveles de SO₂ y NO_x con media hora de antelación.
- Las nuevas predicciones se generan con modelos aditivos para la tendencia: modelos GAM.

$$\hat{X}_{t+30} = \beta_0 + s_1(X_t) + s_2(X_t - X_{t-5})$$

- X_i representa el nivel medio horario del contaminante en el instante i.
- β₀ constante desconocida.
- s₁ y s₂ funciones suaves desconocidas.

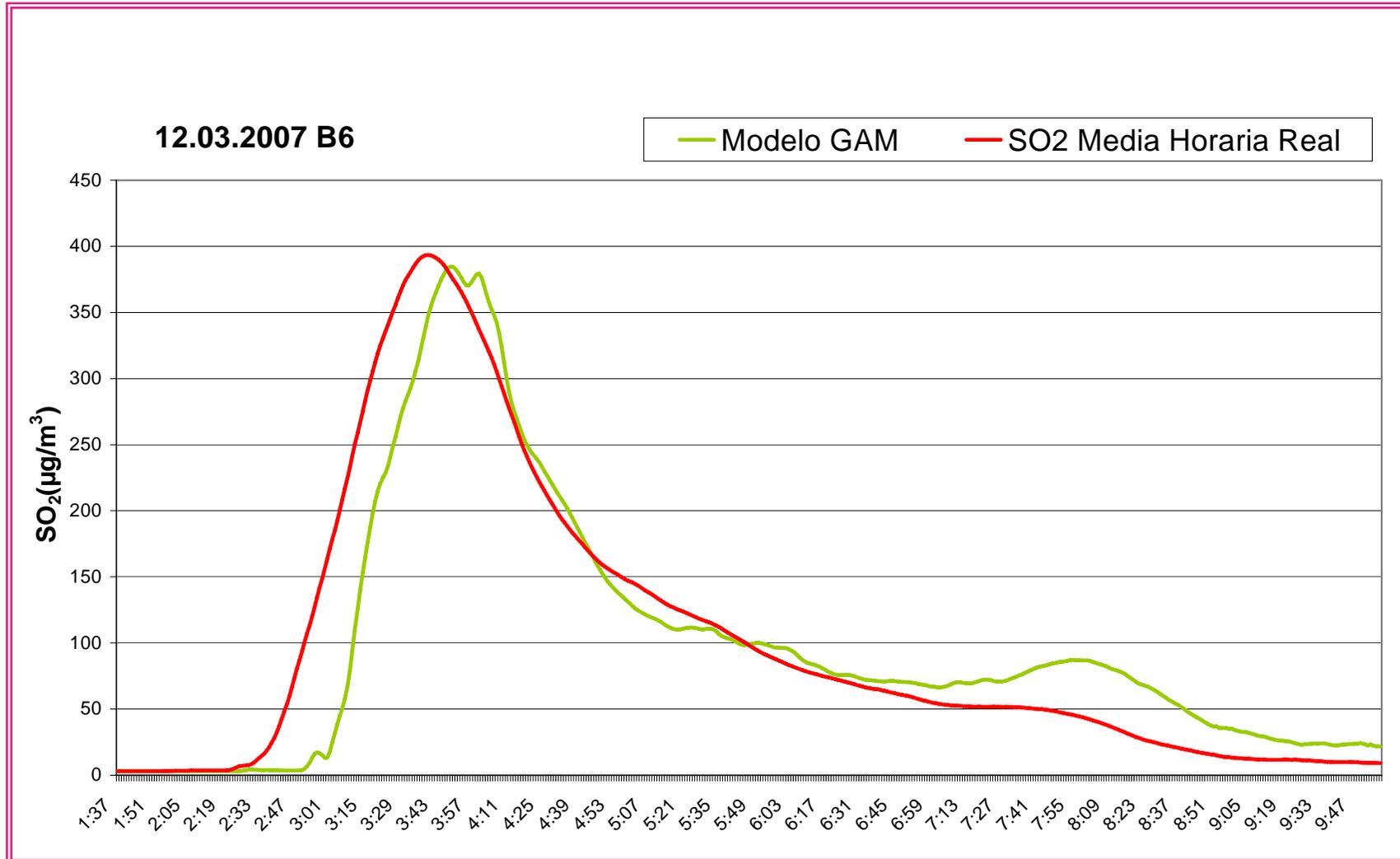
Modelos aditivos

- Se han desarrollado nuevos modelos que obtienen las predicciones de los niveles de SO₂ y NO_x con media hora de antelación.
- Las nuevas predicciones se generan con modelos aditivos para la tendencia: modelos GAM.

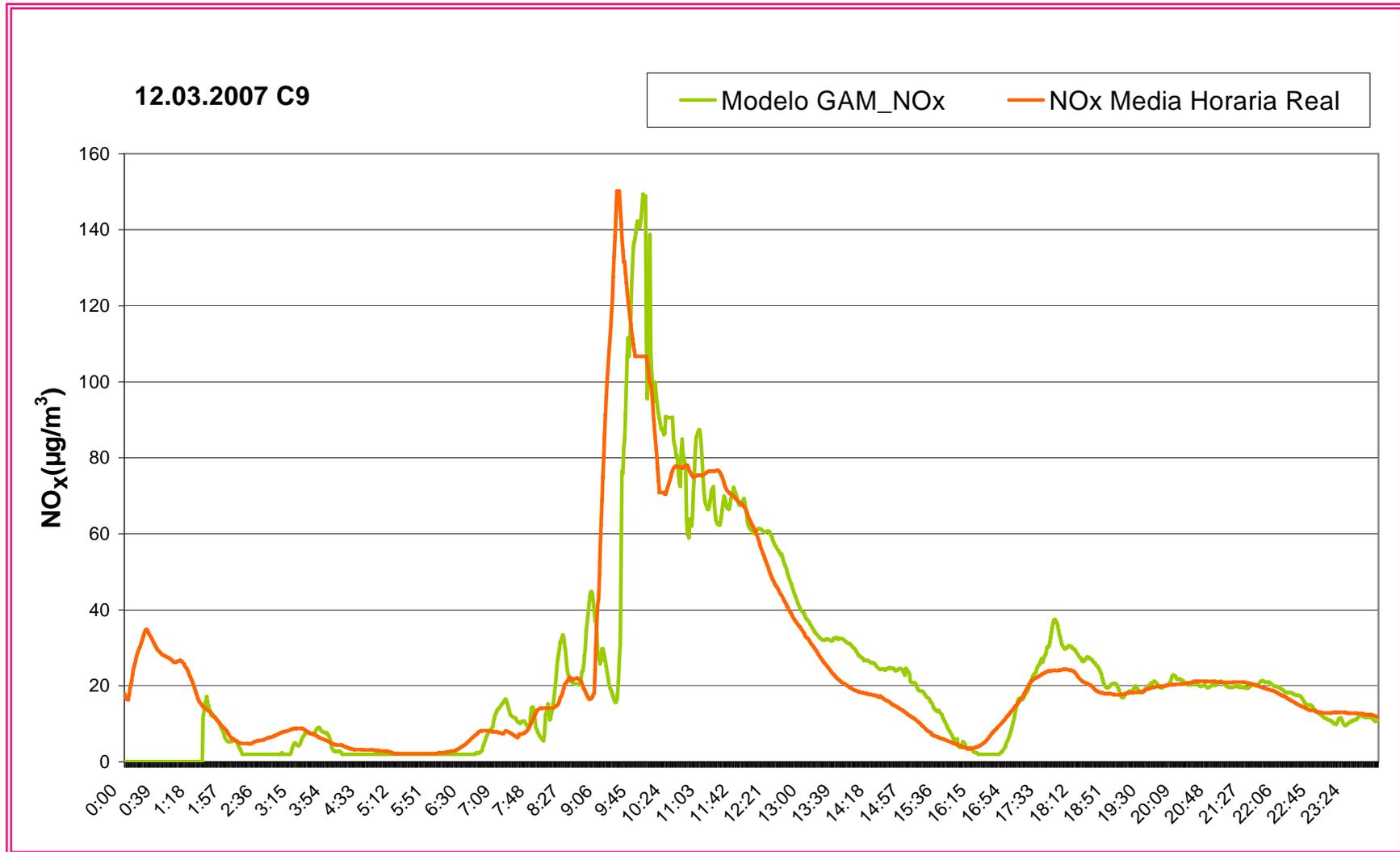
$$\hat{X}_{t+30} = \beta_0 + s_1(X_t) + s_2(X_t - X_{t-5})$$

- X_i representa el nivel medio horario del contaminante en el instante i.
- β₀ constante desconocida.
- s₁ y s₂ funciones suaves desconocidas.

Modelos aditivos



Modelos aditivos



Modelo Semiparamétrico Multidimensional con Cointegración en los errores.

- Sea (Y_l, Z_l) , $l = 0, \pm 1, \pm 2, \pm 3 \dots$, donde Y_l es una serie r-dimensional I(1) con posibilidad de estar cointegrada y Z_l una serie q-dimensional de covariables.
- Se considera el modelo

$$Y_l = \varphi(Z_l) + e_l$$

donde e_l tiene una estructura $VECM(p)$ independiente de Z_l

- La predicción se define por:

$$\dot{Y}_t = \hat{\varphi}_n(Z_t) + \dot{e}_t$$

donde $\hat{\varphi}_n(Z_t)$ es un estimador no paramétrico de φ y \dot{e}_t es la predicción dada por el $VECM$ de la serie residual

$$\hat{e}_t = Y_t - \hat{\varphi}_n(Z_t)$$

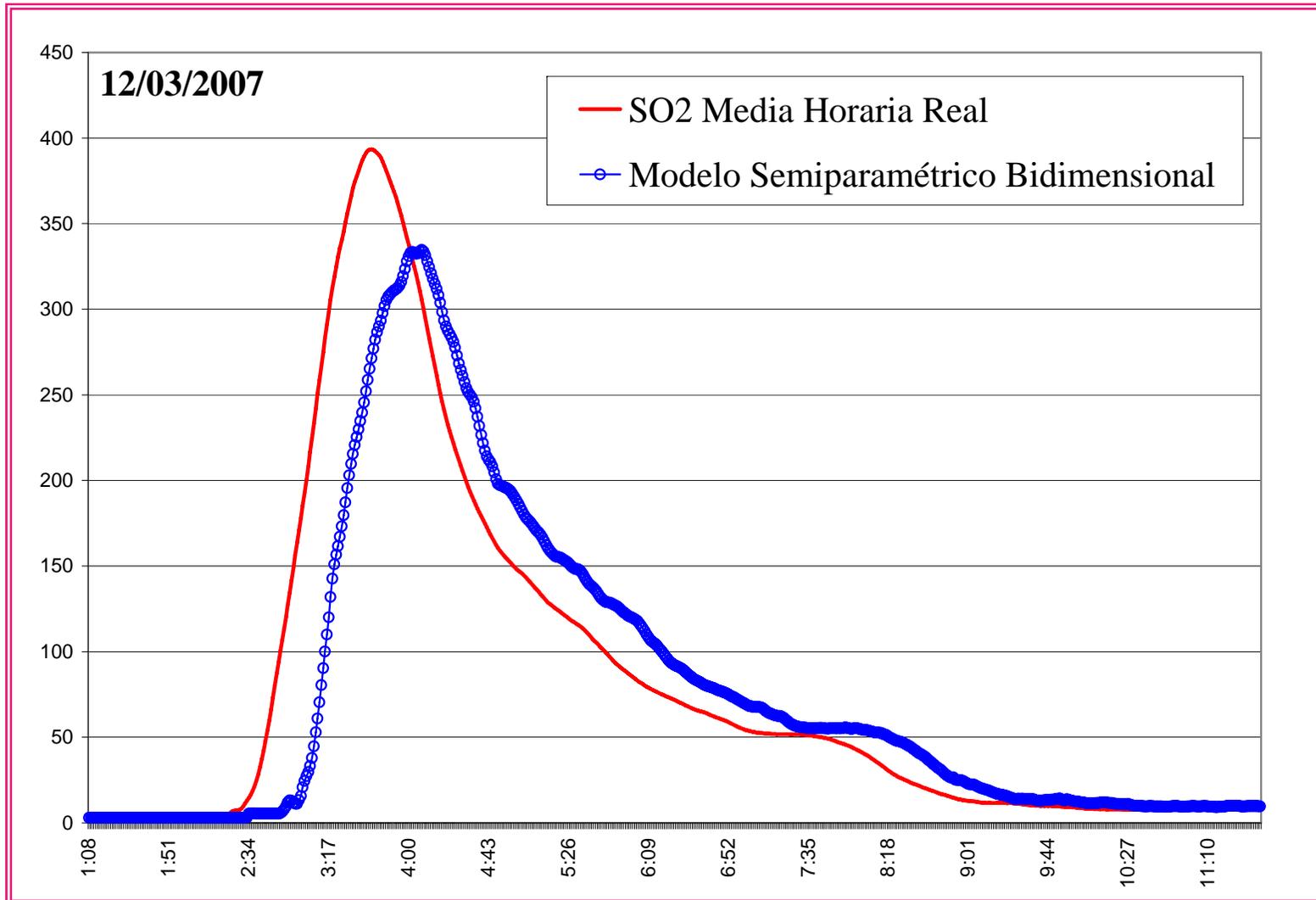
Modelo Semiparamétrico Multidimensional con Cointegración en los errores.

- Sea X_t la serie bidimensional formada por las series de valores medios horarios de SO_2 y NO_x en el instante t .
- En términos de la ecuación anterior: $Y_t = X_{t+k}$ y $Z_t = (X_t, X_t - X_{t-5})$
- Si \hat{X}_i denota los valores observados para instantes pasados ($i \leq t$) y la mejor predicción para instantes futuros ($i > t$), el objetivo es predecir X_{t+30} :
 - Cada instante t , se estima $\varphi(Z_t)$ independientemente para cada componente, de dos formas:
 1. Instante a instante: $\dot{Y}_t = \dot{X}_{t+i} = \hat{\varphi}_1(\hat{Z}_{t+i-1}) + \dot{e}_{t+i}, \quad i = 1, \dots, 30$
 2. A 30 instantes: $\dot{Y}_t = \dot{X}_{t+30} = \hat{\varphi}_{30}(Z_t) + \dot{e}_{t+30}$

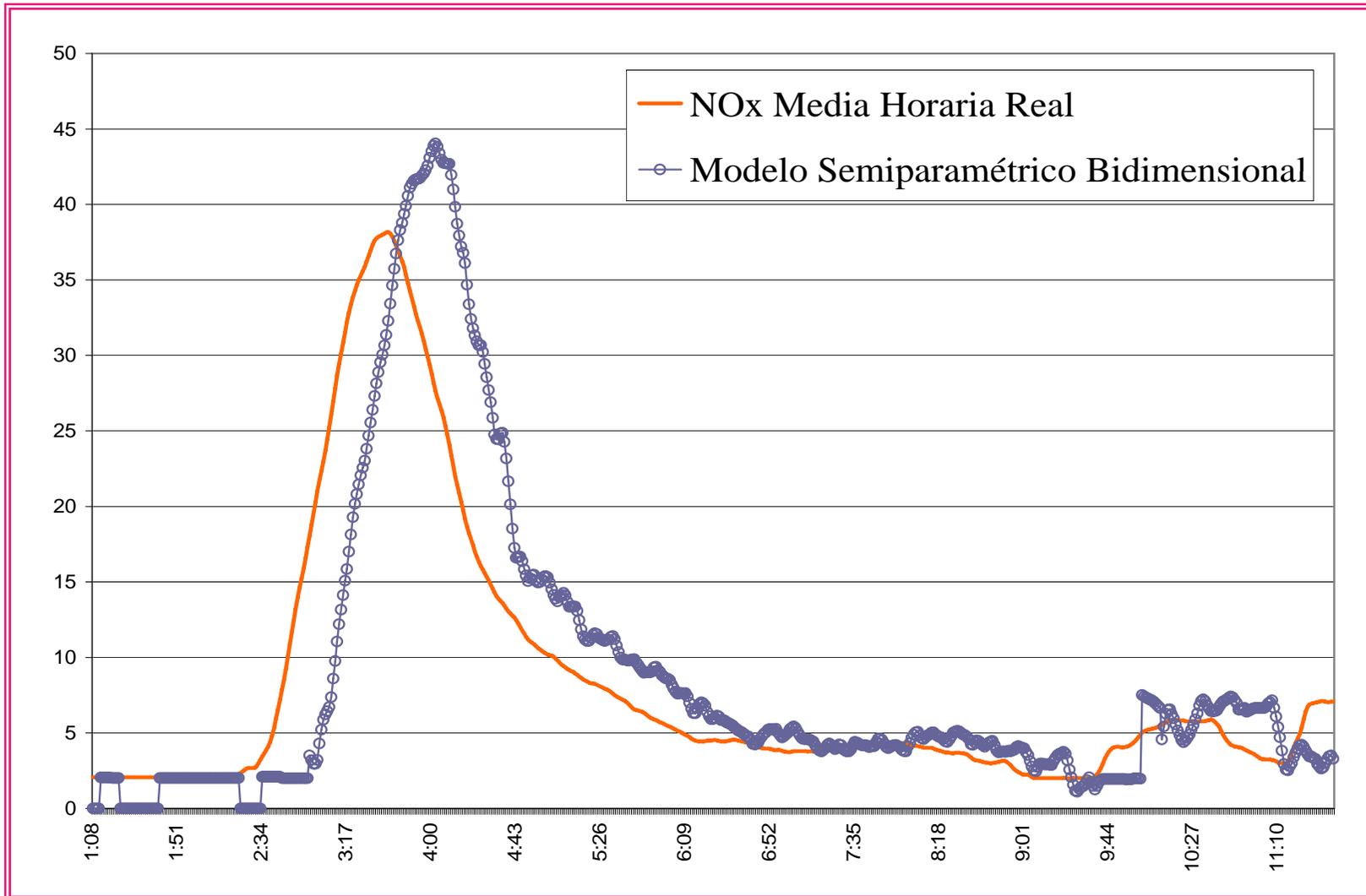
Modelo Semiparamétrico Multidimensional con Cointegración en los errores.

- La serie residual \hat{e}_t se calcula como $\hat{e}_t = Y_t - \hat{\varphi}_j(Z_t)$, $j = 1 \text{ ó } 30$ y se realiza un test de adecuación del modelo (test Ljung-Box) a cada series componente relativas a las últimas 4 horas.
- Si ambas componentes de la serie residual resultan no ser ruido blanco se realiza un test para ver si existe una relación de cointegración (test Johansen). En este caso se ajusta un VECM adecuado y si la serie no está cointegrada se ajusta un VAR. Se obtiene la predicción \dot{e}_{t+30} .
- Las predicciones finales propuestas son:
 1. SPBM1: $\dot{X}_{t+1} = \hat{\varphi}_1(\hat{Z}_t) + \dot{e}_{t+1}, \dots, \dot{X}_{t+30} = \hat{\varphi}_1(\hat{Z}_{t+29}) + \dot{e}_{t+30}$
 2. SPBM30: $\dot{X}_{t+30} = \hat{\varphi}_{30}(\hat{Z}_t) + \dot{e}_{t+30}$

Modelo Semiparamétrico Multidimensional con Cointegración en los errores.



Modelo Semiparamétrico Multidimensional con Cointegración en los errores.

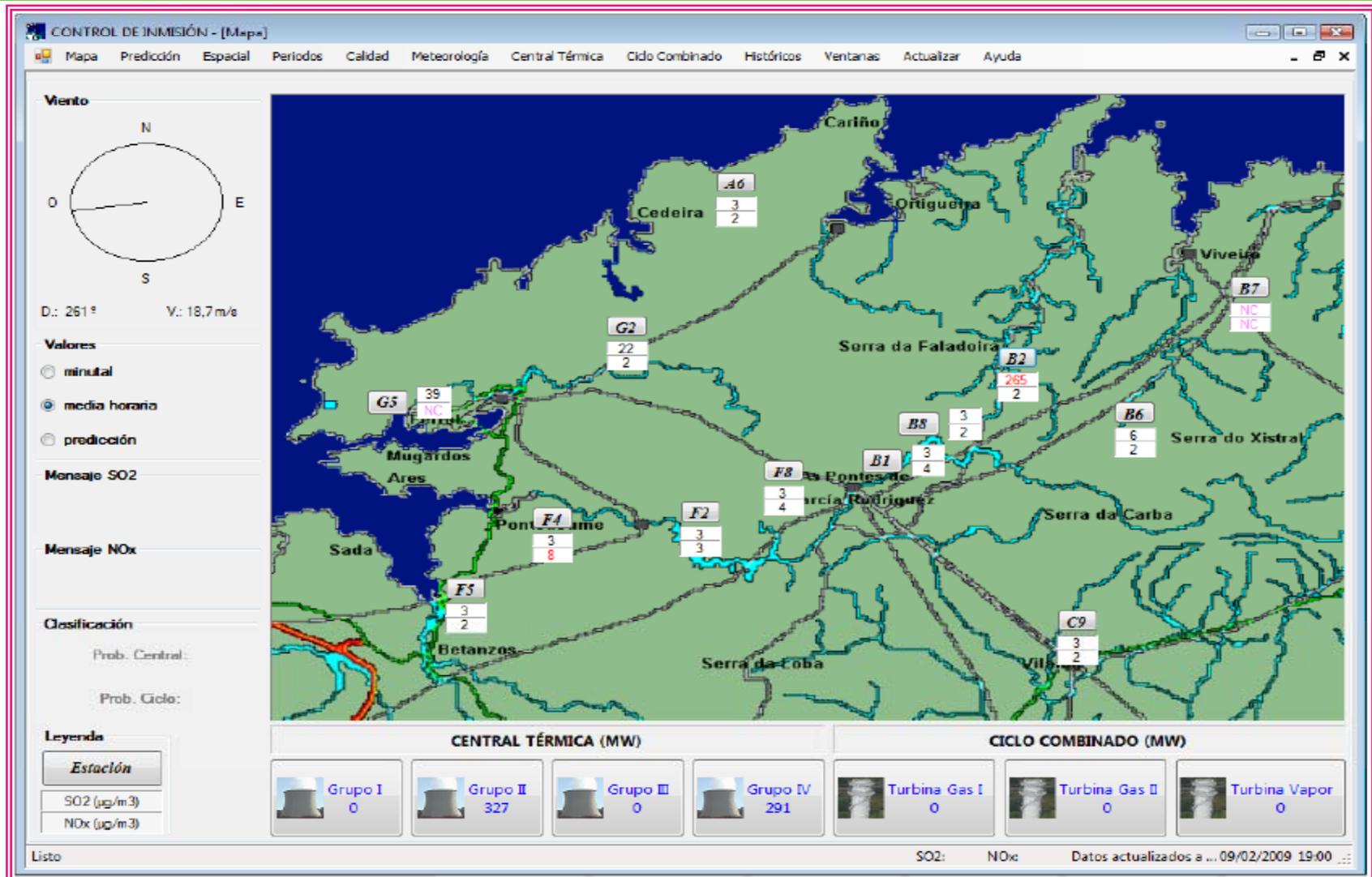


Herramienta de visualización



- Los programas que realizan todas las predicciones funcionan en continuo y están instalados en un ordenador de la sección de Medio Ambiente al que los operadores y el personal de Medio Ambiente se pueden conectar desde sus puestos de trabajo.
- Además, para acceder de forma sencilla a todas las predicciones, SIPEI tiene una herramienta de visualización: *VPR_INM'07 (Versión 3)*

Pantalla principal





María Piñeiro Lamas
maria.pineiro@usc.es