

**MODELOS DE CLASIFICACIÓN  
ESTADÍSTICA BASADOS EN  
INDICADORES DE SO<sub>2</sub> Y NO<sub>x</sub>  
EN LA U.P.T. DE AS PONTES**

**Proyecto Fin de Máster**

**Máster en Técnicas Estadísticas**



Los Programas Oficiales de Postgrado (POP) son estudios definidos dentro del Espacio Europeo de Educación Superior (EEES) como conducentes a la obtención de los títulos de Máster y Doctor.

En este contexto, la Universidad de Santiago de Compostela, conjuntamente con las de A Coruña y Vigo, recoge en su oferta académica para el bienio 2007-2009 el POP Interuniversitario en Estadística e Investigación Operativa.

Dentro de este programa se incluye el Máster en Técnicas Estadísticas, estructurado en tres cuatrimestres con una carga lectiva de 30 créditos ECTS cada uno. De éstos, se establece que 10 créditos de los correspondientes al último cuatrimestre se dedicarán a la realización de un Proyecto Fin de Máster o Trabajo Tutelado, que estará autorizado y dirigido por un profesor del POP.

El presente documento recoge bajo el título “**Modelos de Clasificación Estadística basados en Indicadores de SO<sub>2</sub> y NO<sub>x</sub> en la U.P.T. de As Pontes**” el trabajo realizado por D. Francisco Manuel Prieto Magdalena como Proyecto Fin de Máster para el citado Máster en Técnicas Estadísticas.

D. Wenceslao González Manteiga y D. Manuel Febrero Bande, ambos catedráticos del Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela y profesores del citado POP, autorizan y dirigen el presente proyecto.

Este trabajo fue cofinanciado por la Dirección Xeral de Investigación, Desenvolvemento e Innovación de la Consellería de Innovación e Industria de la Xunta de Galicia a través del programa IN.CI.TE (Código PGIDIT07PXIB207031PR).

Santiago de Compostela, 26 de junio de 2009

Fdo.: D. Francisco Manuel Prieto Magdalena

Fdo.: D. Wenceslao González Manteiga

Fdo.: D. Manuel Febrero Bande

*A todas las personas que en este tiempo han estado o han pasado por la Sección de Medio Ambiente de la U.P.T. de As Pontes, que tanto me han ayudado y tan bien me han acogido, y a todas las que forman parte de este proyecto, por sus aportaciones y su colaboración, y en especial a María y a Adela... Gracias!*

## ÍNDICE

<b>INTRODUCCIÓN.....</b>	<b>2</b>
<b>1. LA U.P.T. DE AS PONTES .....</b>	<b>4</b>
1.1. HISTORIA.....	4
1.2. MARCO LEGISLATIVO DE REFERENCIA .....	7
1.3. SISTEMAS DE CONTROL .....	10
<b>2. MODELOS DE CLASIFICACIÓN.....</b>	<b>13</b>
2.1. DESCRIPCIÓN DEL PROBLEMA .....	13
2.2. MODELO INICIAL.....	15
2.3. NUEVOS MODELOS PROPUESTOS .....	18
2.3.1. Justificación.....	18
2.3.2. Modelos.....	21
2.3.3. Esquema .....	29
<b>3. APLICACIÓN A DATOS REALES.....</b>	<b>31</b>
3.1. CONSIDERACIONES PREVIAS.....	31
3.2. UN PRIMER EJEMPLO .....	33
3.3. UN SEGUNDO EJEMPLO .....	42
<b>4. OTROS MODELOS EN ESTUDIO.....</b>	<b>47</b>
<b>REFERENCIAS .....</b>	<b>49</b>
<b>ANEXO I: METODOLOGÍA.....</b>	<b>50</b>
<b>ANEXO II: CÓDIGO R.....</b>	<b>55</b>

## INTRODUCCIÓN

Últimamente se viene hablando mucho del calentamiento global del planeta como un problema de especial gravedad que podría afectar enormemente al conjunto de la sociedad en los años venideros.

En este sentido, se fue extendiendo la creencia de que el problema guarda una relación directa con el descontrolado incremento experimentado en las últimas décadas en los niveles de emisión de agentes contaminantes a la atmósfera, principalmente dióxido de carbono (CO<sub>2</sub>).

Sin embargo, ni éste es un problema nuevo ni es el único que afecta negativamente al Medio Ambiente.

Numerosos colectivos sociales y científicos viene alertando desde mucho tiempo atrás de lo nocivas que pueden llegar a ser, tanto para el ecosistema como para la propia salud humana, las actividades derivadas del desarrollo industrial y la explotación de los recursos naturales existentes.

Gracias a esto se ha ido prestando más atención a lo que realmente estaba ocurriendo y se le ha ido dando más importancia.

Son cada vez más las personas y colectivos que están aunando esfuerzos y alzando la voz para conseguir concienciar y hacer reaccionar a aquellos en cuyas manos está tomar o exigir que se tomen medidas al respecto.

Así, algunas administraciones han ido dictando en los últimos tiempos normas referidas a la vigilancia y control de este tipo de actividades con el objetivo de garantizar un desarrollo sostenible y más respetuoso con el Medio Ambiente.

Por su parte las empresas, ya sea por conciencia o por obligación, han ido incluyendo entre sus planes medidas encaminadas a prevenir y controlar el impacto medioambiental derivado de su propia actividad.

Sin embargo, a pesar del trabajo de concienciación que se está realizando y de las medidas que se están adoptando, no siempre es posible prevenir y evitar que se produzcan situaciones dañinas con el Medio Ambiente.

En estos casos, es de vital importancia su detección y control para intentar que se cause el menor daño posible, así como el determinar el origen y la causa de las mismas para poder exigir responsabilidades al respecto y evitar que tales situaciones se vuelvan a producir.

La comunidad científica tiene mucho que aportar tanto en el análisis de la situación como en la definición de medidas a adoptar, así como en la implantación efectiva de las mismas.

No en vano, el Medio Ambiente es un área de investigación muy importante para numerosos sectores científicos. Entre ellos las matemáticas y, en particular, la Estadística.

Un claro ejemplo se puede encontrar en el ambicioso programa de colaboración que se viene desarrollando desde hace varios años entre la Sección de Medio Ambiente de la Unidad de Producción Térmica de As Pontes y el Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela.

Esta colaboración se ha basado fundamentalmente en el desarrollo de modelos estadísticos que permitan predecir la aparición de episodios de alteración de la calidad del aire.

Sin embargo, en la U.P.T. de As Pontes se han producido últimamente importantes cambios que han motivado la aparición de nuevos intereses encaminados, además de a predecir, a determinar el origen de un episodio que se pudiese llegar a producir.

Este trabajo es el resultado de los más recientes estudios realizados dentro de este programa y que van dirigidos a la determinación del origen más probable de las emisiones que motivaron la aparición de un episodio de contaminación.

A continuación se presenta la situación y características de la U.P.T. de As Pontes, lo que permitirá conocer y comprender el marco de referencia de este estudio. Posteriormente se describe la problemática a abordar y se detallan tanto los modelos existentes como los nuevos que se proponen. Después, se ilustra con algunos ejemplos el resultado de su aplicación en datos reales y, finalmente, se citan otros modelos en estudio.

## 1. LA U.P.T. DE AS PONTES

### 1.1. HISTORIA

El municipio gallego de As Pontes de García Rodríguez alberga uno de los más importantes yacimientos mineros de lignito pardo de la Península Ibérica.

Tras su descubrimiento se inició la construcción en sus proximidades de una central térmica para el aprovechamiento y explotación del mineral en la producción de energía eléctrica.

La instalación, que tenía en la Mina y en la Central Térmica sus principales activos, es propiedad de Endesa Generación, S.A. y representa uno de los complejos mineroeléctricos más importantes del continente.

De la mina, que es a cielo abierto, se extraía fundamentalmente lignito pardo para abastecer de combustible fósil a la central. Este mineral tiene unas características entre las que destacan negativamente su elevada humedad y contenidos en azufre, así como su bajo poder calorífico. Pero tenía la ventaja de su bajo coste de extracción y traslado.

La U.P.T. de As Pontes inició su actividad en la producción de energía eléctrica en el año 1976 con la puesta en funcionamiento del primero de los cuatro grupos que conformaban el diseño de la Central Térmica. Cada uno de estos grupos cuenta con una potencia nominal eléctrica aproximada de 350 MW y funciona de manera independiente a los demás, aunque están dispuestos en dos fases: los grupos I y II en una y el III y IV en otra, estando cada una de ellas gestionada y controlada de manera diferenciada.

Durante varios años, la actividad de la instalación se basó en la extracción del lignito de la mina, traslado a la central, almacenamiento y posterior combustión en los diferentes grupos para la producción de energía eléctrica. Funcionando la Central Térmica, de esta manera, exclusivamente con carbón local procedente de la mina.

Con el paso de los años, las previsiones de agotamiento del yacimiento y las características del lignito, tanto en lo que a poder calorífico se refiere como a la emisión de contaminantes que provocaba su combustión, debido a su propia composición, llevó a los responsables de la instalación a trazar una hoja de ruta, previo acuerdo con las diferentes administraciones públicas competentes en la materia, que permitiese en los siguientes años ir incorporando otros tipos de carbón. De esta forma, se garantizaría la explotación de la mina hasta su agotamiento dentro de unos plazos razonables, al tiempo que se podría ir trabajando en los planes de regeneración de la misma, respetando además el cumplimiento de las nuevas exigencias que en materia normativa se estaban aprobando.

La elección de carbones subbituminosos de importación para mezclar con el lignito pardo local obligaba a transformar los grupos de la central para conseguir una mejora en la eficiencia del proceso.

Esta operación se inició en el año 1993 con la transformación del primer grupo y se dio por concluida en el año 1996, año en el que los cuatro grupos estaban ya en condiciones adecuadas a la nueva realidad.

Gracias a esta adaptación se lograrían alcanzar importantes reducciones en las emisiones de contaminantes a la atmósfera, siendo las más significativas, debido al bajo contenido en azufre del nuevo carbón, las referidas a dióxido de azufre ( $\text{SO}_2$ ), con reducciones globales superiores al 40%.

Sin embargo, en los planes de actuación figuraba una segunda fase en la que se pasaría de utilizar una mezcla de carbones a utilizar únicamente carbón de importación, lo que supondría el cierre definitivo de la mina local, previsto para enero de 2008.

Esta medida requeriría de una nueva transformación en los grupos, la última hasta la fecha, que permitiría a la central funcionar de la manera más eficiente posible quemando exclusivamente carbón subbituminoso importado.

Los nuevos trabajos de adaptación se desarrollaron entre los años 2005 y 2008 y concluyeron con la puesta en funcionamiento del último grupo transformado, que fue, curiosamente, el grupo I.

Tras esta modificación se lograrían nuevas reducciones en las emisiones de  $\text{SO}_2$ , cercanas al 96%, así como de óxidos de nitrógeno ( $\text{NO}_x$ ), con un 17%, dióxido de carbono ( $\text{CO}_2$ ), casi un 9%, y partículas en suspensión, con algo más del 46%.

Paralelamente a esta última actuación se estaban acometiendo ya en las inmediaciones de la Central Térmica los trabajos de construcción de una nueva central, también propiedad de la misma empresa, y que estaría integrada en el conjunto de la U.P.T. de As Pontes. Se trataba de una central de ciclo combinado, que funcionaría fundamentalmente quemando gas natural para producir energía eléctrica.

El Ciclo Combinado proyectado estaría compuesto por dos turbinas de gas y una de vapor. Cada una de las turbinas de gas funcionaría de manera independiente y tendría una potencia nominal eléctrica aproximada de 270 MW, que junto con la turbina de vapor supondría para el conjunto de la central una potencia aproximada de 800 MW. Y esto a su vez, unido a la capacidad de la Central Térmica, supondría para el total de la instalación una potencia superior a los 2200 MW.

Tras varios años de trabajos dedicados al estudio, diseño y construcción, en el primer semestre del año 2008 se alcanzó el ansiado hito de su puesta en funcionamiento.

En la actualidad, la Mina está cerrada. En ella se han iniciado ya los trabajos de regeneración de acuerdo con los planes aprobados. Planes que contemplan el llenado del hueco dejado por el lignito pardo con aguas procedentes de la lluvia y los ríos cercanos para formar un impresionante lago artificial.

La Central Térmica, completamente transformada, funciona a pleno rendimiento utilizando única y exclusivamente carbón subbituminoso de importación.

El Ciclo Combinado, superadas las fases de prueba y puesta a punto, se encuentra ya oficialmente en producción.

En la Figura 1.1 se puede apreciar la situación y tamaño de la Mina con respecto a la Central Térmica, el Ciclo Combinado y al propio municipio de As Pontes de García Rodríguez.



**Figura 1.1** Situación de la U.P.T. de As Pontes.

## 1.2. MARCO LEGISLATIVO DE REFERENCIA

La actividad desarrollada en la U.P.T. de As Pontes produce inevitablemente emisiones de sustancias contaminantes a la atmósfera. Estas emisiones deben ser controladas para garantizar el cumplimiento, en cada momento, de la normativa legal vigente que le sea de aplicación.

En la actualidad, las condiciones que buscan garantizar la calidad de aire con respecto a la incidencia que tiene la emisión de ciertas sustancias a la atmósfera y el control de las propias emisiones están regulados en las siguientes normas:

- *Real Decreto 1073/2002, de 18 de octubre, sobre evaluación y gestión de la calidad del aire ambiente en relación con el dióxido de azufre, dióxido de nitrógeno, óxidos de nitrógeno, partículas, plomo, benceno y el monóxido de carbono.*

Que adapta a la legislación nacional los contenidos de las siguientes directivas europeas:

- *Directiva 1999/30/CE del Consejo, de 22 de abril de 1999, relativa a los valores límite de dióxido de azufre, dióxido de nitrógeno y óxidos de nitrógeno, partículas y plomo en el aire ambiente.*
- *Directiva 2000/69/CE del Parlamento Europeo y del Consejo, de 16 de noviembre de 2000, sobre los valores límite para el benceno y el monóxido de carbono en el aire ambiente.*
- *Real Decreto 430/2004, de 12 de marzo, por el que se establecen nuevas normas sobre limitación de emisiones a la atmósfera de determinados agentes contaminantes procedentes de grandes instalaciones de combustión, y se fijan ciertas condiciones para el control de las emisiones a la atmósfera de las refinerías de petróleo.*

Que adapta a la legislación nacional los contenidos de la siguiente directiva europea:

- *Directiva 2001/80/CE del Parlamento Europeo y del Consejo, de 23 de octubre de 2001, sobre limitación de emisiones a la atmósfera de determinados agentes contaminantes procedentes de grandes instalaciones de combustión.*

El Real Decreto 1073/2002 establece unos valores límite y fija los denominados Valores Umbrales de Evaluación, que permiten establecer el tipo de medición a realizar, en función de los valores históricos obtenidos en cada emplazamiento.

Hay dos tipos de Valores Umbrales de Evaluación:

- **Umbral de Evaluación Superior (UES):** el nivel marcado para cada contaminante, por debajo del cual puede utilizarse una combinación de mediciones y técnicas de modelización para evaluar la calidad del aire ambiente. Por encima del umbral de evaluación superior, se debe proceder a realizar mediciones en continuo.
- **Umbral de Evaluación Inferior (UEI):** el nivel marcado para cada contaminante, por debajo del cual es posible limitarse al empleo de técnicas de modelización o de estimación objetiva para evaluar la calidad del aire ambiente.

En lo que al alcance de este trabajo se refiere, únicamente son de interés los valores límite y umbrales que el citado Real Decreto establece para el SO<sub>2</sub>, el NO<sub>2</sub> y el NO<sub>x</sub>.

### Valores límite del SO<sub>2</sub>

PROTECCIÓN SALUD HUMANA		PROTECCIÓN ECOSISTEMAS
VALOR LÍMITE HORARIO (VLH)	VALOR LÍMITE DIARIO (VLD)	VALOR LÍMITE ANUAL (VLA)
<b>350 µg/m<sup>3</sup></b> No debe superarse más de 24 veces/año	<b>125 µg/m<sup>3</sup></b> No debe superarse más de 3 veces/año	<b>20 µg/m<sup>3</sup></b> (año civil e invierno)

**Tabla 1.1** Valores límite del SO<sub>2</sub> para la protección de la salud y de los ecosistemas.

### Valores umbrales del SO<sub>2</sub>

PROTECCIÓN SALUD HUMANA		PROTECCIÓN ECOSISTEMAS	
UES	UEI	UES	UEI
<b>75 µg/m<sup>3</sup></b> No debe superarse más de 3 veces/año	<b>50 µg/m<sup>3</sup></b> No debe superarse más de 3 veces/año	<b>12 µg/m<sup>3</sup></b> (valor invernal)	<b>8 µg/m<sup>3</sup></b> (valor invernal)

**Tabla 1.2** Valores Umbrales de Evaluación del SO<sub>2</sub>.

**Valores límite del NO<sub>2</sub>**

PROTECCIÓN SALUD HUMANA	
VALOR LÍMITE HORARIO (VLH)	VALOR LÍMITE ANUAL (VLA)
<p><b>200 µg/m<sup>3</sup></b> No debe superarse más de 18 veces/año</p>	<p><b>40 µg/m<sup>3</sup></b></p>

**Tabla 1.3** Valores límite del NO<sub>2</sub> para la protección de la salud.

**Valores umbrales del NO<sub>2</sub>**

PROTECCIÓN SALUD HUMANA			
(VLH)		(VLA)	
UES	UEI	UES	UEI
<p><b>140 µg/m<sup>3</sup></b> No debe superarse más de 18 veces/año</p>	<p><b>100 µg/m<sup>3</sup></b> No debe superarse más de 18 veces/año</p>	<p><b>32 µg/m<sup>3</sup></b></p>	<p><b>26 µg/m<sup>3</sup></b></p>

**Tabla 1.4** Valores Umbrales de Evaluación del NO<sub>2</sub>.

**Valor límite del NO<sub>x</sub>**

PROTECCIÓN VEGETACIÓN
VALOR LÍMITE ANUAL (VLA)
<p><b>30 µg/m<sup>3</sup></b></p>

**Tabla 1.5** Valor límite del NO<sub>x</sub> para la protección de la vegetación.

**Valores umbrales del NO<sub>x</sub>**

PROTECCIÓN VEGETACIÓN	
UES	UEI
<p><b>24 µg/m<sup>3</sup></b></p>	<p><b>19,5 µg/m<sup>3</sup></b></p>

**Tabla 1.6** Valores Umbrales de Evaluación del NO<sub>x</sub>.

### 1.3. SISTEMAS DE CONTROL

Con la puesta en funcionamiento de la U.P.T. de As Pontes se creó una Red de Vigilancia y Control de la Calidad Atmosférica que permitiese llevar un seguimiento y control de los niveles de concentración de agentes contaminantes en el entorno de la instalación.

Esta red estaba compuesta inicialmente por 32 estaciones de medida manuales, que llegaron a ser 40 con el paso de los años, y que se fue rediseñando y mejorando hasta llegar a disponer de otras 17 estaciones de medida automáticas. Quedando éstas como las únicas que se venían utilizando en los últimos años.

Recientemente, debido principalmente al impacto que suponía la instalación del Ciclo Combinado, se hizo necesario abordar una nueva redefinición de la red (Proymasa-Endesa, 2006) en la que se decidió prescindir de 7 de las estaciones existentes y añadir otras 3 de nueva construcción.

Así se llegó a la situación actual, en la que se dispone de una red de 13 estaciones de medida automáticas convenientemente situadas en un radio aproximado de 30 km. alrededor de la instalación.

Las estaciones automáticas transmiten por radio medidas en continuo de la evolución de la calidad del aire y están dotadas todas ellas de analizadores tanto de SO<sub>2</sub> como de NO<sub>x</sub>. Algunas de ellas disponen también de otros equipos como analizadores de monóxido de carbono (CO), ozono (O<sub>3</sub>) o partículas en suspensión, así como sensores de temperatura, dirección y velocidad de viento.

La red se completa con una estación meteorológica central, denominada E. M. de A Mourela, que está situada a poco más de un kilómetro de la U.P.T. y dispone de un mástil de 80 metros de altura con sensores que registran información de temperatura en superficie y a 10, 30 y 80 metros de altura, dirección y velocidad de viento a 10 y 80 metros de altura, radiación solar, precipitación, humedad relativa y presión atmosférica.

En cuanto a emisiones se refiere, para la Central Térmica ya se disponía de un Sistema de Control de Emisiones propio que registra en la actualidad, para cada uno de los grupos, información de los niveles de emisión a la atmosfera de SO<sub>2</sub>, NO<sub>x</sub>, partículas en suspensión, oxígeno (O<sub>2</sub>), temperatura, potencia y caudal.

Para el Ciclo Combinado se instaló un sistema similar de control de emisiones, pero que registra parámetros más acordes a la naturaleza de su actividad. Se registra así para cada turbina de gas información de los niveles de emisión a la atmosfera de SO<sub>2</sub>, NO<sub>x</sub>, partículas en suspensión, monóxido de carbono (CO), compuestos orgánicos volátiles (COV's), oxígeno (O<sub>2</sub>), temperatura, potencia y caudal.

La Red de Vigilancia y Control de la Calidad Atmosférica, la Estación Meteorológica de A Mourela y el Sistema de Control de Emisiones de la Central Térmica conforman lo que se conoce como Sistema de Control Suplementario de la Contaminación Atmosférica, al que ahora se suma el Sistema de Control de Emisiones del Ciclo Combinado.

Uno de los objetivos principales de este sistema es poder adecuar las condiciones de operación de la instalación para tratar de evitar posibles episodios de contaminación.

Toda la información recogida en los diferentes elementos que componen el sistema se recibe, almacena y gestiona en la Unidad Central de Adquisición de Datos y Gestión de la Información.

Parte de esta información, la que resulta de utilidad en cada caso, se pone a su vez a disposición de otros sistemas periféricos, tanto internos como externos, que se dedican al control, tratamiento y aprovechamiento de la misma.

Entre estos sistemas se encuentra el desarrollado por el Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela para la predicción de niveles de contaminación en el entorno de la instalación.

Durante varios años de colaboración se ha venido investigando en este tema y desarrollando modelos estadísticos que ayuden a los responsables de la instalación a tomar decisiones con el objetivo de anticiparse a la aparición de posibles episodios de contaminación causados por las emisiones de la misma.

Así, se ha llegado a disponer en la actualidad de un completo Sistema de Predicción Estadística de Inmisión (SIPEI) que, integrado en el sistema de control, incorpora a éste modelos estadísticos de predicción que permiten anticiparse a posibles episodios de alteración de la calidad del aire facilitando, con media hora de antelación, estimaciones sobre los niveles previstos de  $\text{SO}_2$  y  $\text{NO}_x$  en cada una de las estaciones de la red.

Información detalla de este sistema puede verse en Piñeiro Lamas, María (2007).

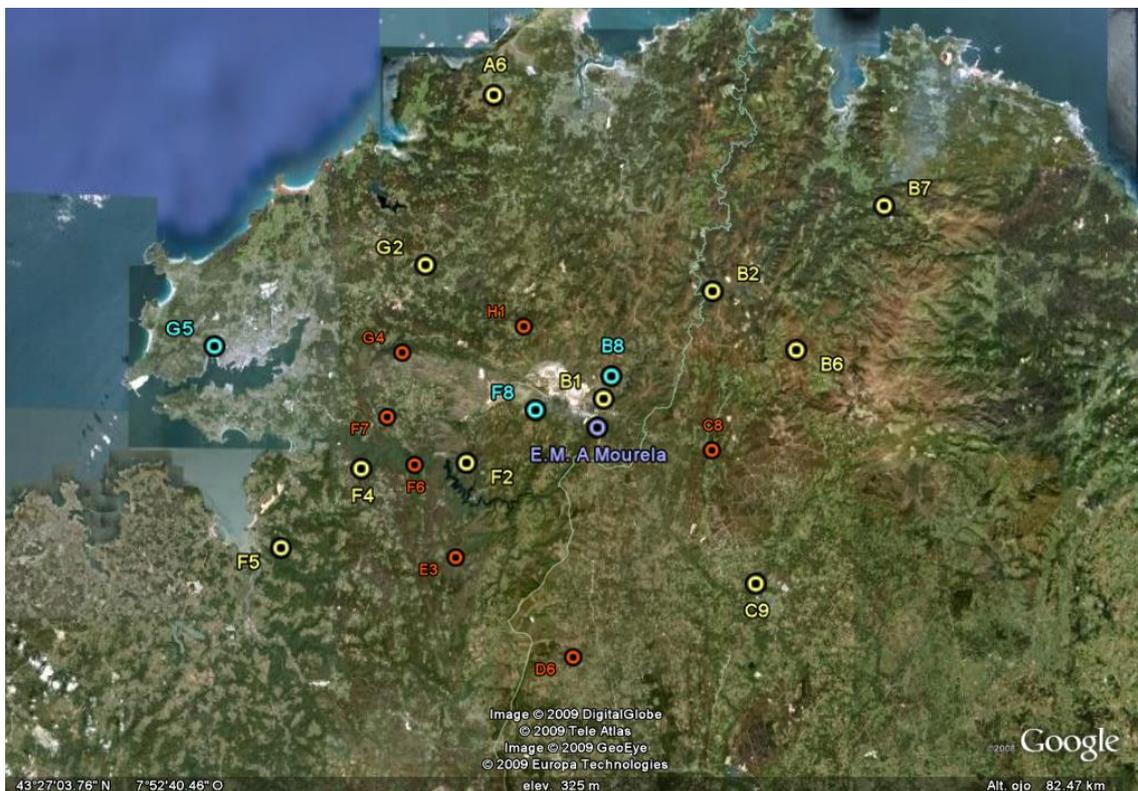
Paralelamente al interés en predecir, en los últimos años ha surgido un importante interés en clasificar el origen de un episodio. Esto es, determinar cuál es el origen de las emisiones que causaron la aparición de un episodio de alteración de la calidad del aire.

Los primeros estudios en esta línea, previos a la construcción del Ciclo Combinado, se centraron únicamente en determinar si el origen de las emisiones podía ser imputable o no a la actividad de la Central Térmica, dado que éste era el único posible foco de emisión con que contaba la instalación.

Sin embargo, con la puesta en funcionamiento del Ciclo Combinado la situación se ha complicado ya que ahora no se trata solo de determinar si el posible origen está o no en la instalación, sino en cuál de las centrales o, incluso, en qué grado puede estar involucrada cada una de ellas.

Esta problemática es la que se aborda en el presente estudio.

La situación y distribución de las estaciones de medida automáticas y de la estación meteorológica de A Mourela se puede ver en la siguiente imagen (Figura 1.2). En ella se han diferenciado en colores las estaciones para destacar el resultado de la última redefinición. Así, aparecen en color amarillo las estaciones que no han cambiado, en naranja las que se han eliminado y en azul las de nueva incorporación.



**Figura 1.2** Distribución de la Red de Vigilancia y Control de la Calidad Atmosférica y Estación Meteorológica de A Mourela.

## 2. MODELOS DE CLASIFICACIÓN

### 2.1. DESCRIPCIÓN DEL PROBLEMA

Cuando surgió el interés inicial por clasificar el origen de un episodio de alteración de la calidad del aire el único posible foco de emisión existente en la U.P.T. de As Pontes era la Central Térmica.

En este contexto, se pretendía desarrollar un modelo sencillo que, en caso de producirse un episodio, permitiese calcular la probabilidad de que éste fuese debido a emisiones procedentes de la central.

Se quería además que el modelo se desarrollase como parte integrada en el sistema de predicción que se tenía. En éste se había establecido una escala de niveles de alarma que advertía sobre la importancia que las predicciones obtenidas daban a la posibilidad de alcanzar o superar los valores límite de  $\text{SO}_2$  y  $\text{NO}_x$  legalmente establecidos.

Al igual que en el caso de la predicción, el modelo se limitaría únicamente a clasificar el origen de los posibles episodios de  $\text{SO}_2$  y de  $\text{NO}_x$  y se haría siempre que los modelos de predicción generasen una alarma. Dándose además prioridad, debido a las características de la propia instalación, a los episodios de  $\text{SO}_2$ .

Planteado el problema, lo primero que había que asumir era que, independientemente del modelo que se desarrollase, los resultados obtenidos serían difícilmente contrastables, por lo que el análisis previo al diseño del mismo debía ser lo más completo posible.

Dada la situación resultaba lógico pensar que, tratándose de determinar si una sustancia detectada en una estación procede o no de la Central Térmica, sería importante considerar a qué distancia y en qué dirección se encontraba ésta. Y esto llevaría, consecuentemente, a considerar la dirección y velocidad del viento. Sin embargo había un problema insalvable, y es que las emisiones de la central se hacen a través de una chimenea de 356 metros de altura y ni se dispone de información sobre las condiciones meteorológicas a esa altura ni, aunque se conociesen, resultaría sencillo predecir el comportamiento que tendría el penacho bajo las mismas.

La segunda cuestión que se tuvo en cuenta, y a la postre fundamental, fue que la Central Térmica emite tanto  $\text{SO}_2$  como  $\text{NO}_x$  en cantidades significativamente elevadas. Esto significa que la presencia de una debe ir acompañada necesariamente de la otra, en mayor o menor medida. Y, de igual modo, la ausencia de una conllevaría la ausencia de la otra.

Otra consideración a realizar a la hora de decidir si una sustancia procede o no de un determinado foco emisor es conocer qué otros posibles focos pueden existir en el entorno con capacidad para generarla.

La respuesta en el caso del  $\text{SO}_2$  era bastante simple: ninguno, al menos que se acercase a los niveles de emisión de la Central Térmica.

La respuesta en el caso del  $\text{NO}_x$  resultó ser algo más compleja. Eran varios los focos ya conocidos, y más los que podrían aparecer. Entre los ya conocidos se encontraban los del tráfico de vehículos a motor, de actividades agropecuarias o de otras industrias presentes en la zona. Entre los que podrían aparecer estarían los asociados a la construcción de nuevas instalaciones industriales. Entre ellas una ya prevista y conocida, el Ciclo Combinado.

A continuación se detalla el modelo inicial desarrollado para dar respuesta al problema aquí expuesto. Para, seguidamente, describir la situación actual, cuyo panorama ha cambiado sustancialmente con la entrada en funcionamiento del Ciclo Combinado y la transformación de la Central Térmica, y presentar los nuevos modelos propuestos para enfrentarse a la misma.

## 2.2. MODELO INICIAL

Una vez descartada la posibilidad de utilizar en el modelo variables meteorológicas como la dirección o velocidad del viento, todos los esfuerzos se centraron en aprovechar al máximo la relación existente entre los niveles de emisión de  $\text{SO}_2$  y  $\text{NO}_x$ .

Considerando tanto las características de estas sustancias como el comportamiento del penacho que las contiene, afectado por procesos de difusión y dilución, era evidente que los niveles en los que ambas sustancias eran emitidas a la atmósfera disminuirían considerablemente a medida que éste se alejase de la central. Sin embargo, era de esperar que ambas disminuyeran en igual proporción, por lo que el ratio entre ambas debería mantenerse. Y fue en esta propiedad en la que se basó el modelo inicial.

Así, si  $s$  y  $n$  fuesen los niveles de inmisión de  $\text{SO}_2$  y  $\text{NO}_x$ , respectivamente, medidos en una determinada estación y  $S_{CT}$  y  $N_{CT}$  los niveles de emisión de la Central Térmica correspondientes, se tendría que

$$\frac{s}{n} = \frac{S_{CT}}{N_{CT}}$$

Es decir, el ratio de inmisión coincidiría con ratio de emisión de la central.

Esto se verificaría si la Central Térmica fuese el único foco de emisión posible de ambas sustancias, lo cual no era cierto, al menos en el caso del  $\text{NO}_x$ .

Sabido esto, resultaría entonces de gran utilidad identificar la procedencia del mismo en la mayor medida que fuese posible. Para ello, aprovechando que se disponía de gran cantidad de datos históricos, se decidió estudiar el comportamiento del  $\text{NO}_x$  en cada zona.

Como resultado de este estudio se obtuvo una matriz de datos históricos, a los que se denominaría  $\text{NO}_x$  de fondo, confeccionada en base a los niveles de  $\text{NO}_x$  registrados en cada una de las estaciones en días en los que no se había producido ningún fenómeno anormal.

Una última cuestión que había que tener en cuenta era que las mediciones de los niveles de inmisión estaban limitadas por unos valores mínimos que había fijados para los propios analizadores, por lo que todas las medidas registradas para  $\text{SO}_2$  y  $\text{NO}_x$  estaban sometidas a las restricciones  $s \geq s_{min} = 3 \mu\text{g}/\text{m}^3$  y  $n \geq n_{min} = 2 \mu\text{g}/\text{m}^3$ , respectivamente.

Teniendo en cuenta todo esto se diseñó el modelo inicial de ratios en el que se consideraba el  $\text{NO}_x$  de fondo, y que se describe a continuación.

## Modelo inicial de ratios

Lo que se hace es lo siguiente:

- Se consideran  $s$  y  $n$ , niveles (media horaria) de inmisión de  $\text{SO}_2$  y  $\text{NO}_x$  en una estación  $e$  en un instante  $t$ .
- Se consideran  $\mathbf{S}_{CT} = (S_{CT,1}, \dots, S_{CT,T})$  y  $\mathbf{N}_{CT} = (N_{CT,1}, \dots, N_{CT,T})$ , niveles de emisión de  $\text{SO}_2$  y  $\text{NO}_x$  de la Central Térmica en  $T$  instantes de tiempo anteriores a  $t$ , tomados a partir de un instante  $t_0$  inicial convenientemente seleccionado.
- Se obtiene  $\mathbf{n}_F^* = (n_{F,1}^*, \dots, n_{F,K}^*)$ , muestra artificial del  $\text{NO}_x$  de fondo en la estación  $e$ , obtenida de simular  $K$  valores de una distribución Normal  $N(\tilde{n}_F, \sigma_F)$ , donde  $\tilde{n}_F$  y  $\sigma_F$  son, respectivamente, la mediana y la desviación típica correspondientes al conjunto de datos históricos del  $\text{NO}_x$  de fondo en esa estación.

Si alguno de los valores de  $\mathbf{n}_F^*$  es inferior a  $n_{min}$ , valor mínimo prefijado, se cambia y se pone éste.

- Se calcula  $\mathbf{n}_{-F}^* = n - \mathbf{n}_F^*$ , muestra artificial del  $\text{NO}_x$  si el de fondo.  
Si alguno de los valores de  $\mathbf{n}_{-F}^*$  es inferior a  $n_{min}$ , valor mínimo prefijado, se cambia y se pone éste.
- Se calculan los ratios de inmisión sin el  $\text{NO}_x$  de fondo  $\mathbf{r}_{-F}^*$  y de emisión de la Central Térmica  $\mathbf{R}_{CT}$ , donde

$$\mathbf{r}_{-F}^* = \mathbf{r}_{-F}^{S*} = \frac{s}{\mathbf{n}_{-F}^*} \quad \text{y} \quad \mathbf{R}_{CT} = \mathbf{R}_{CT}^S = \frac{\mathbf{S}_{CT}}{\mathbf{N}_{CT}}$$

si se trata de una alarma de  $\text{SO}_2$ , o

$$\mathbf{r}_{-F}^* = \mathbf{r}_{-F}^{n*} = \frac{\mathbf{n}_{-F}^*}{s} \quad \text{y} \quad \mathbf{R}_{CT} = \mathbf{R}_{CT}^N = \frac{\mathbf{N}_{CT}}{\mathbf{S}_{CT}}$$

si se trata de una alarma de  $\text{NO}_x$ .

- Se estiman de manera no paramétrica (véase Anexo I.A)  $\hat{f}_{-F}$  y  $\hat{f}_{CT}$ , densidades de los ratios calculados  $r_{-F}^*$  y  $R_{CT}$ , respectivamente y según el caso.
- Se asigna la probabilidad de que el origen esté en la Central Térmica a partir de la comparación de las densidades estimadas

$$\hat{p}_{CT} = \underset{p \in A}{\operatorname{argmin}} \sum_{i=1}^I \left( \hat{f}_{-F}(x_i) - p \hat{f}_{CT}(x_i) \right)^2$$

donde  $I$  es el número de puntos estimados para las densidades y  $A = \left\{ \frac{j}{J} : j = 0, \dots, J \right\}$ , con  $J$  suficientemente grande.

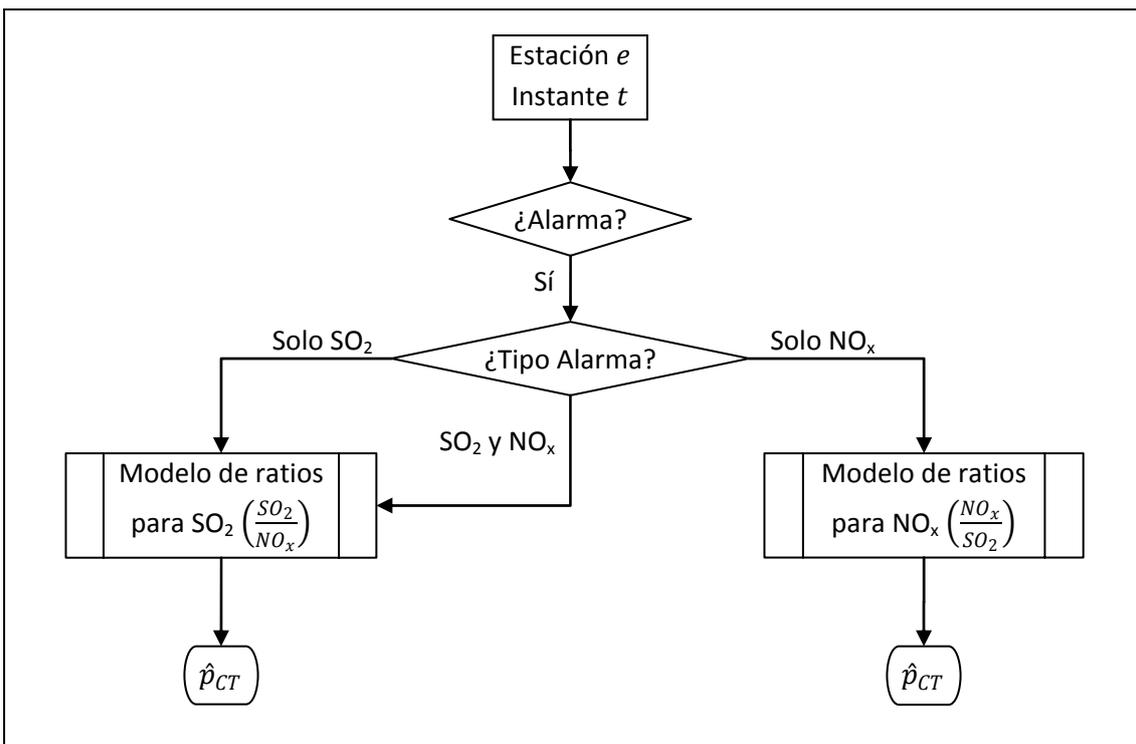


Figura 2.1 Esquema del modelo inicial.

## 2.3. NUEVOS MODELOS PROPUESTOS

### 2.3.1. Justificación

La reciente construcción del Ciclo Combinado en la U.P.T. de As Pontes y la transformación de la Central Térmica para trabajar únicamente con carbón de importación ha supuesto una importante alteración del mapa de emisiones en la zona. Tanto que incluso ha obligado a revisar y redefinir la red de estaciones que se tenía, eliminando siete de las estaciones existentes y añadiendo tres de nueva creación. Y todo ello debido a las enormes diferencias existentes, por una parte, entre las características de esta nueva central y la existente y, por otra, entre las del lignito pardo local y el carbón subbituminoso de importación.

Una primera diferencia se encuentra en las escalas de valores que se manejan para una y otra central, siendo los niveles de emisión de la Central Térmica muy superiores a los del Ciclo Combinado.

Otra diferencia, quizás la de mayor influencia para el entorno, se encuentra en las condiciones en las que se realizan las emisiones, ya que la Central Térmica emite desde una única chimenea de 356 metros de altura y el Ciclo Combinado desde dos, una para cada turbina de gas, de 80 metros cada una.

Finalmente, las diferencias en los carbones supondrán una reducción importantísima en las emisiones de  $\text{SO}_2$ , por lo que el interés prioritario que éste ostentaba hasta la fecha cederá terreno claramente en favor del  $\text{NO}_x$ .

Esta nueva situación obliga irremediablemente a replantear el modelo de clasificación inicial, que se había desarrollado únicamente para la Central Térmica, que era la que había en aquel momento.

Así pues, al problema inicial se añadían ahora nuevos condicionantes que aumentaban enormemente su complejidad y obligaban a revisar y completar el análisis realizado en su día.

Lógicamente la primera opción debía pasar por valorar si el modelo inicial de ratios se podía utilizar también para el Ciclo Combinado. La respuesta fue inmediata y contundente, y fue que no. La actividad normal del Ciclo Combinado genera emisiones significativas de  $\text{NO}_x$ , pero no de  $\text{SO}_2$ , por lo que no tiene sentido calcular un ratio entre ambas. Esto, sin embargo, también tenía su lectura positiva, no sería necesario contemplar la posibilidad de que el origen de un episodio de  $\text{SO}_2$  estuviese en el Ciclo Combinado, reforzando aun más el interés por el  $\text{NO}_x$ .

La segunda opción, en la misma línea, era buscar un sustituto para el  $\text{SO}_2$ , es decir, otra sustancia emitida por el Ciclo Combinado con la que calcular un ratio con el  $\text{NO}_x$ . El resultado de esta búsqueda tampoco dio buenos resultados, ya que no hay ninguna otra sustancia que se pueda utilizar como referencia.

Había entonces que buscar otra alternativa. Y ésta se encaminó precisamente en la dirección que se había tenido que descartar en el modelo inicial, la consideración de variables meteorológicas. Ahora sí que se disponía de información relativa a la dirección y velocidad del viento a la altura de emisión, puesto que la E. M. de A Mourela dispone de sensores a 80 metros, justo la altura de las chimeneas del Ciclo Combinado. Sin embargo, aun conociendo estas condiciones, no resultaría sencillo predecir el comportamiento del penacho bajo las mismas.

Por otra parte, al ser la central de nueva construcción, ni se disponía ni se podría disponer hasta pasado algún tiempo de datos que sirviesen de referencia.

Esta ausencia de datos, unida a la dificultad que suponía poder contrastar los resultados, hacía realmente difícil el desarrollo de un modelo de garantías.

También con los datos de inmisión se tendría una dificultad añadida, la ausencia de datos históricos en las tres nuevas estaciones de medida incluidas en la red. Lo que obligaría a tomar como referencia los datos de otras estaciones cuya situación tuviese el mayor parecido posible con la prevista para cada una de las nuevas. Con esto habría que redibujar el mapa del  $\text{NO}_x$  de fondo existente, al menos temporalmente y a la espera de poder disponer de datos reales de las nuevas estaciones.

En cuanto a la Central Térmica se refiere, el modelo inicial estaba dando resultados satisfactorios por lo que parecía razonable mantenerlo. Eso sí, habría que revisarlo para contemplar, además del  $\text{NO}_x$  de fondo, el procedente del Ciclo Combinado.

Una vez analizado el problema en profundidad había una cosa que estaba bien clara, y era que la situación que se presentaba era francamente incierta y solo con el paso del tiempo, a medida que se fuese disponiendo de datos, se podrían ir despejando algunas de las incógnitas existentes.

Esto llevó a tomar una decisión, firme desde el principio, en cuanto a la base sobre la que se debía desarrollar el nuevo modelo. Éste debía ser lo más flexible posible.

Bajo esta premisa, se diseñó una estructura modular compuesta por varios modelos, cada uno de los cuales estaría dotado de identidad propia, orientado a resolver un problema concreto y de forma que la conexión de todos ellos daría como resultado la respuesta al problema inicial.

De esta forma, a medida que se fuese disponiendo de nuevos datos se podría ir calibrando cada modelo por separado, corrigiéndolo, si fuese necesario, sin que ello afectase a los demás.

Teniendo en cuenta todo esto se diseñaron los nuevos modelos.

Para su desarrollo, partiendo del modelo inicial, se utilizaron ideas basadas en el estudio de datos direccionales (véase en más detalle en Mardia, Kanti V. & Jupp, Peter E., 2000) y se consideraron los resultados de los modelos de difusión utilizados para la redefinición de la red de estaciones (véase informe de Proymasa-Endesa, 2006).

Se describirá en primer lugar cada uno de los modelos desarrollados, para mostrar después el modo en que todos ellos se conectan para conformar el esquema final que dé respuesta al problema planteado.

## 2.3.2. Modelos

### 2.3.2.1. Modelo de históricos

Es un modelo desarrollado exclusivamente para la generación de muestras artificiales del  $\text{NO}_x$  de fondo a partir de los datos históricos de inmisión utilizando una estimación no paramétrica de la densidad.

Lo que se hace es lo siguiente:

- Se considera  $n_F^h$ , datos históricos del  $\text{NO}_x$  de fondo en una estación  $e$ .
- Se considera  $n_{min}$ , valor mínimo prefijado.
- Se calcula  $\rho$ , proporción de datos de  $n_F^h$  cuyo valor es  $n_{min}$ .
- Se considera una mixtura de la forma

$$X = \rho X_1 + (1 - \rho) X_2$$

donde  $X_1$  es una variable que sigue una distribución  $\delta_{n_{min}}$  concentrada en  $n_{min}$  y  $X_2$  es una variable truncada por la izquierda en  $n_{min}$  y cuya densidad se estima de manera no paramétrica (véase Anexo I.A) a partir de los datos de  $n_F^h$  cuyo valor esté por encima de  $n_{min}$ .

- Se simulan  $K$  valores de la variable  $X$ , donde para la parte de la simulación que afecta a la variable  $X_2$  los valores se obtienen aplicando el método Bootstrap suavizado (véase Anexo I.B) con el parámetro ventana obtenido en la estimación de su densidad (véase Anexo I.A).
- Se obtiene la muestra  $n_F^* = (n_{F,1}^*, \dots, n_{F,K}^*)$ , siendo  $n_{F,1}^*, \dots, n_{F,K}^*$  los  $K$  valores obtenidos en la simulación.

### 2.3.2.2. Modelo de ratios

Es una evolución del modelo inicial de ratios en el que además del NO<sub>x</sub> de fondo se considera el del Ciclo Combinado.

Se utiliza para realizar estimaciones para la Central Térmica.

Permite obtener  $\hat{p}_{CT}^s$  y  $\hat{p}_{CT}^n$ , probabilidades estimadas de que el origen del SO<sub>2</sub> y del NO<sub>x</sub>, respectivamente, sea la Central Térmica.

Se asume que no hay presencia de niveles de SO<sub>2</sub> imputables al Ciclo Combinado y, por tanto, su probabilidad es nula. Es decir

$$s_{CC} = 0 \quad \text{y} \quad p_{CC}^s = 0$$

Lo que se hace es lo siguiente:

- Se consideran  $s$  y  $n$ , niveles (media horaria) de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en una estación  $e$  en un instante  $t$ .
- Se consideran  $S_{CT}$  y  $N_{CT}$ , niveles de emisión de SO<sub>2</sub> y NO<sub>x</sub> de la Central Térmica en  $T$  instantes de tiempo anteriores a  $t$ , tomados a partir de un instante  $t_0$  inicialmente convenientemente seleccionado.
- Se considera  $\mathbf{n}_F^*$ , muestra artificial del NO<sub>x</sub> de fondo en la estación  $e$ .  
Si alguno de los valores de  $\mathbf{n}_F^*$  es inferior a  $n_{min}$ , valor mínimo prefijado, se cambia y se pone éste.
- Se considera  $\mathbf{n}_{CC}^*$ , muestra artificial del NO<sub>x</sub> correspondiente al Ciclo Combinado en la estación  $e$ .  
Si alguno de los valores de  $\mathbf{n}_{CC}^*$ , es inferior a cero, se cambia y se pone éste.
- Se calcula  $\mathbf{n}_{-F-CC}^* = n - \mathbf{n}_F^* - \mathbf{n}_{CC}^*$ , muestra artificial del NO<sub>x</sub> sin el de fondo ni del Ciclo Combinado.  
Si alguno de los valores de  $\mathbf{n}_{-F-CC}^*$  es inferior a  $n_{min}$ , valor mínimo prefijado, se cambia y se pone éste.

- Se calculan los ratios de inmisión sin el  $\text{NO}_x$  de fondo ni del Ciclo Combinado  $r_{-F-CC}^*$  y de emisión de la Central Térmica  $R_{CT}$ , donde

$$r_{-F-CC}^* = r_{-F-CC}^{S*} = \frac{s}{n_{-F-CC}^*} \quad \text{y} \quad R_{CT} = R_{CT}^S = \frac{S_{CT}}{N_{CT}}$$

en el caso del  $\text{SO}_2$ , o

$$r_{-F-CC}^* = r_{-F-CC}^{n*} = \frac{n_{-F-CC}^*}{s} \quad \text{y} \quad R_{CT} = R_{CT}^N = \frac{N_{CT}}{S_{CT}}$$

en el caso del  $\text{NO}_x$ .

- Se estiman de manera no paramétrica (véase Anexo I.A)  $\hat{f}_{-F-CC}$  y  $\hat{f}_{CT}$ , densidades de los ratios calculados  $r_{-F-CC}^*$  y  $R_{CT}$ , respectivamente y según el caso.
- Se asigna la probabilidad de que el origen esté en la Central Térmica a partir de la comparación de las densidades estimadas

$$\hat{p}_{CT} = \underset{p \in A}{\operatorname{argmin}} \sum_{i=1}^I \left( \hat{f}_{-F-CC}(x_i) - p \hat{f}_{CT}(x_i) \right)^2$$

donde  $I$  es el número de puntos estimados para las densidades y  $A = \left\{ \frac{j}{J} : j = 0, \dots, J \right\}$ , con  $J$  suficientemente grande.

### 2.3.2.3. Modelo de viento

Es un modelo para la estimación del  $\text{NO}_x$  en el que se usan la dirección y velocidad del viento y la distancia.

Se utilizan para realizar estimaciones para el Ciclo Combinado.

Permite obtener  $\hat{p}_V$ , probabilidad estimada de que las emisiones del Ciclo Combinado puedan incidir en una estación en función del viento.

Lo que se hace es lo siguiente:

- Se considera  $d_e$ , distancia de la estación  $e$  al Ciclo Combinado.
- Se considera  $D_e$ , dirección (ángulo respecto al Norte) de la recta que une el Ciclo Combinado con la estación  $e$ .
- Se consideran  $\mathbf{D}_V = (D_{V,1}, \dots, D_{V,T})$  y  $\mathbf{V}_V = (V_{V,1}, \dots, V_{V,T})$ , dirección (ángulo respecto al Norte) y velocidad del viento en el Ciclo Combinado en  $T$  instantes de tiempo anteriores a  $t$ , tomados a partir de un instante  $t_1$  inicial convenientemente seleccionado.
- Se obtiene  $\mathbf{PV}_V = (PV_{V,1}, \dots, PV_{V,T})$ , proyecciones de las  $T$  velocidades de  $\mathbf{V}_V$  sobre la dirección fijada por  $D_e$  como

$$PV_{V,i} = V_{V,i} \cos(D_{V,i} - D_e)$$

- Se obtiene el conjunto de instantes en que existe una posible incidencia del viento sobre la estación  $e$ . Es decir, se obtienen aquellos índices  $i \in I = \{1, \dots, T\}$  en que

$$PV_{V,i} > 0 \quad \text{ó} \quad V_{V,i} = 0$$

Sea  $I_e = \{i \in I: PV_{V,i} > 0 \text{ ó } V_{V,i} = 0\}$ .

- Para cada  $i \in I_e$  se genera una distribución Ji-Cuadrado  $\chi_{m_i}^2$  sobre la dirección  $D_e$ , siendo los grados de libertad

$$m_i = c_1 \ln(\exp\{c_2/c_1\} + PV_{V,i})$$

con  $c_1$  y  $c_2$  constantes convenientemente seleccionadas.

- Para cada  $i \in I_e$  se simulan  $K$  valores de una distribución  $\chi_{m_i}^2$ , obteniéndose las muestras  $\mathbf{n}_i^* = (n_{i,1}^*, \dots, n_{i,K}^*)$ , siendo  $n_{i,1}^*, \dots, n_{i,K}^*$  los  $K$  valores obtenidos para la distribución  $\chi_{m_i}^2$ .
- Para cada  $i \in I \setminus I_e$  se toman muestras de  $K$  valores nulos,  $\mathbf{n}_i^* = (0, \dots, 0)$ .
- Se considera  $\mathbf{n}_e^* = (n_{i,k}^*)_{i \in I; k \in K}$ , muestra conjunta con todos los valores de todas las muestras.
- Se calcula  $\hat{\alpha}$  evaluando la distribución empírica  $F_n$  de la muestra  $\mathbf{n}_e^*$  en el punto especificado por la distancia  $d_e$

$$\hat{\alpha} = F_n(d_e)$$

- Se calcula  $\hat{p}_V$ , probabilidad de incidencia en la estación  $e$ , como

$$\hat{p}_V = 1 - |2\hat{\alpha} - 1|$$

### 2.3.2.4. Modelo mixto

Es un modelo mixto de dos etapas para la estimación del  $\text{NO}_x$  en el que se usan ratios y proporciones.

Se puede utilizar para realizar estimaciones tanto para la Central Térmica como para el Ciclo Combinado.

Permite obtener  $\hat{p}_{CT}^n$  y  $\hat{p}_{CC}^n$ , probabilidades estimadas de que el origen del  $\text{NO}_x$  sean la Central Térmica y el Ciclo Combinado, respectivamente.

Se asume que no hay presencia de niveles de  $\text{SO}_2$  imputables al Ciclo Combinado y, por tanto, su probabilidad es nula. Es decir

$$s_{CC} = 0 \quad \text{y} \quad p_{CC}^S = 0$$

Lo que se hace es lo siguiente:

#### Etapa 1

- Se consideran  $s$  y  $n$ , niveles (media horaria) de inmisión de  $\text{SO}_2$  y  $\text{NO}_x$  en una estación  $e$  en un instante  $t$ .
- Se consideran  $S_{CT}$  y  $N_{CT}$ , niveles de emisión de  $\text{SO}_2$  y  $\text{NO}_x$  de la Central Térmica en  $T$  instantes de tiempo anteriores a  $t$ , tomados a partir de un instante  $t_0$  inicialmente convenientemente seleccionado.
- Se considera  $n_F^*$ , muestra artificial del  $\text{NO}_x$  de fondo en la estación  $e$ .  
Si alguno de los valores de  $n_F^*$  es inferior a  $n_{min}$ , valor mínimo prefijado, se cambia y se pone éste.
- Se considera una descomposición del  $\text{NO}_x$  en la estación  $e$  del tipo

$$n = n_F + n_{CT} + n_{CC} + n_{OF}$$

donde  $n_F$ ,  $n_{CT}$ ,  $n_{CC}$  y  $n_{OF}$  representan el  $\text{NO}_x$  de fondo y el atribuible a la Central Térmica, al Ciclo Combinado y a otros posibles focos, respectivamente.

- Bajo las hipótesis de que los ratios de emisión e inmisión correspondientes a la Central Térmica coinciden y de que el nivel de SO<sub>2</sub> en inmisión es debido exclusivamente a ésta, es decir  $s = s_{CT}$ , se podrían plantear una serie de equivalencias del tipo

$$\frac{S_{CT}}{N_{CT}} = R_{CT}^s = r_{CT}^s = \frac{S_{CT}}{n_{CT}} = \frac{s}{n_{CT}}$$

Se obtiene así una muestra artificial del NO<sub>x</sub> correspondiente a la Central Térmica

$$n_{CT}^* = \frac{s}{R_{CT}^s}$$

- Con las muestras artificiales del NO<sub>x</sub> de fondo y de la Central Térmica se obtiene una conjunta del Ciclo Combinado y otros posibles focos

$$n_{CC+OF}^* = n_{-F-CT}^* = n - n_F^* - n_{CT}^*$$

Si alguno de los valores de  $n_{CC+OF}^*$ , es inferior a cero, se cambia y se pone éste.

## **Etapas 2**

- Se considera  $n$ , nivel (media horaria) de inmisión de NO<sub>x</sub> en la estación  $e$  en el instante  $t$ .
- Se considera  $n_{CT}^*$ , muestra artificial del NO<sub>x</sub> correspondiente a la Central Térmica.
- Se considera  $n_{CC+OF}^*$ , muestra del NO<sub>x</sub> conjunto del Ciclo Combinado y otros posibles focos.
- Se considera  $\hat{p}_V$ , probabilidad estimada de que las emisiones del Ciclo Combinado puedan incidir en la estación  $e$  en función del viento.

- Se obtiene

$$\mathbf{n}_{CC}^* = \hat{p}_V \mathbf{n}_{CC+OF}^*$$

- Se asigna la probabilidad de que el origen esté en la Central Térmica o el Ciclo Combinado a partir de las proporciones

$$\hat{p}_{CT}^n = \text{Mediana} \left( \frac{\mathbf{n}_{CT}^*}{n} \right)$$

$$\hat{p}_{CC}^n = \text{Mediana} \left( \frac{\mathbf{n}_{CC}^*}{n} \right)$$

### 2.3.3. Esquema

Los cuatro modelos presentados: el de históricos, el de ratios, el de viento y el mixto, éste en sus dos etapas, se combinan para conformar el esquema (Figura 2.2) que permitirá obtener la probabilidad de que los niveles de  $\text{SO}_2$  y  $\text{NO}_x$  detectados en una estación tengan su origen en la Central Térmica o en el Ciclo Combinado, dando respuesta así al problema planteado.

A diferencia de lo que se hacía en el modelo inicial, ahora los cálculos no se realizan únicamente cuando hay alarma en el sistema de predicción, ni en función del tipo de la misma. En su lugar se introduce un nuevo criterio, más flexible, basado en el establecimiento de niveles mínimos de referencia tanto para el  $\text{SO}_2$  como para el  $\text{NO}_x$  (representados en el esquema por  $s_0$  y  $n_0$ , respectivamente) que determinan la necesidad o no de realizar determinados cálculos.

Se suman a éstos otros tres criterios dicotómicos como son el determinar si hay o no  $\text{SO}_2$ , si hay o no  $\text{NO}_x$  y si funciona o no el Ciclo Combinado. Criterios que se fijarán en función de las restricciones y necesidades del sistema.

Otra consideración a realizar a la vista del esquema propuesto es que la probabilidad estimada para la Central Térmica en el caso del  $\text{NO}_x$  se obtiene del modelo de ratios, cuando se podría obtener también en la segunda etapa del modelo mixto. Esto se hace para aprovechar la información que proporciona la presencia de  $\text{SO}_2$ , la cual es utilizada en el modelo de ratios pero no en el mixto.



### 3. APLICACIÓN A DATOS REALES

#### 3.1. CONSIDERACIONES PREVIAS

Los ejemplos que se presentan a continuación son el resultado de aplicar los modelos propuestos a datos reales registrados en determinados días en los que se han producido sucesos de interés, por lo ilustrativo de los mismos.

Para ello se ha seguido el esquema propuesto (Figura 2.2) utilizando en cada uno de los modelos los parámetros oportunos (véase Tabla 3.1).

Además de éstos, se ha fijado

- $s_0 = 10$ , que es el 10% de  $100 \mu\text{g}/\text{m}^3$ , valor establecido como primer nivel de alarma para  $\text{SO}_2$ .

Nótese que el valor límite horario es  $350 \mu\text{g}/\text{m}^3$  (véase Tabla 1.1)

- $n_0 = 10$ , que es el 20% de  $50 \mu\text{g}/\text{m}^3$ , valor establecido como primer nivel de alarma para  $\text{NO}_x$ .

Nótese que el valor límite horario es  $200 \mu\text{g}/\text{m}^3$  (véase Tabla 1.3)

	Modelo de históricos	Modelo de ratios	Modelo de viento	Modelo mixto
$n_{min}$	2	2	--	2
$K$	120	--	120	--
$T$	--	120	30	120
$t_0$	--	$t - 120$	--	$t - 120$
$t_1$	--	--	$t - 15 - d_e/\tilde{V}_V$	--
$c_1$	--	--	3.6	--
$c_2$	--	--	2	--
$I$	--	512	--	--
$J$	--	1000	--	--

**Tabla 3.1** Parámetros utilizados en los modelos, donde  $t$  es el instante actual,  $d_e$  la distancia a la estación  $e$  (en km) y  $\tilde{V}_V$  es la mediana de las velocidades de viento (en m/s) en los  $T$  instantes de referencia considerados.

La implementación de los modelos se ha realizado utilizando lenguaje R, para lo que se ha desarrollado una rutina específica cuyo código se puede ver en el Anexo II.

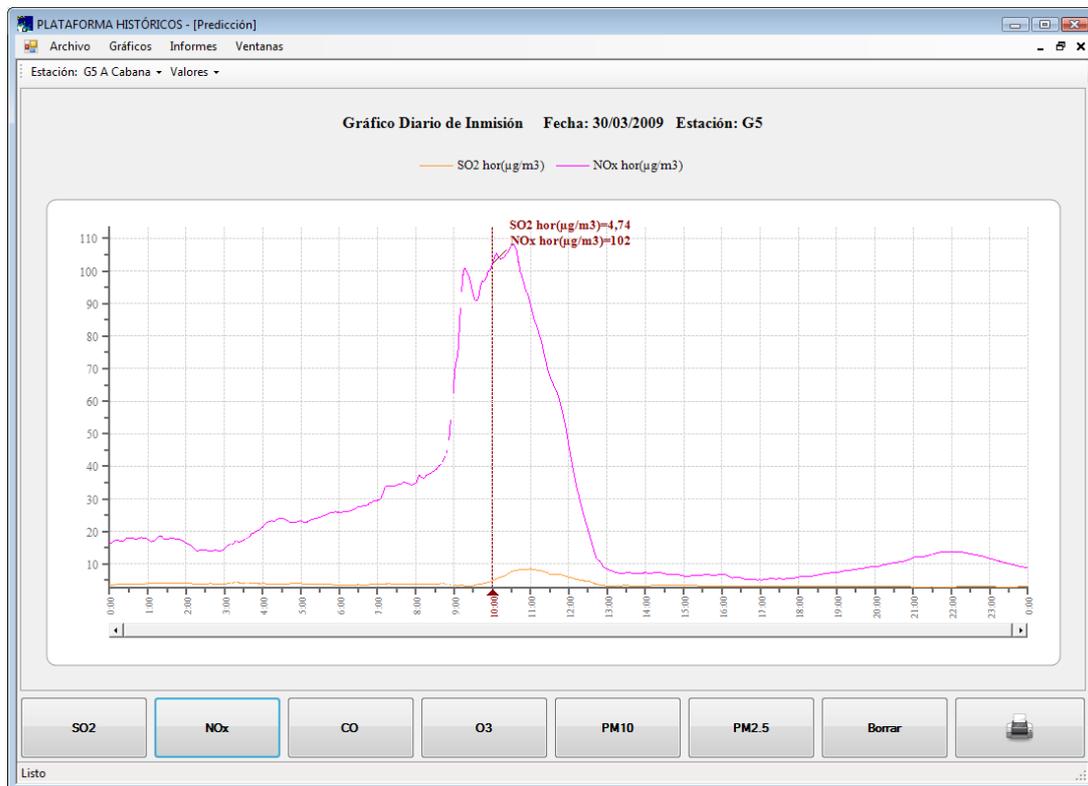
Para ilustrar los ejemplos se ha utilizado la aplicación de visualización VPR INM 2007, que forma parte del sistema de predicción estadística SIPEI.

Es importante tener en cuenta que, tanto en lo que a ésta como a los cálculos se refiere, las direcciones que se consideran representan el ángulo que forman respecto a la dirección Norte y en el sentido de las agujas del reloj. Esto afecta tanto a la dirección del viento como a en la que se encuentra cada estación respecto al Ciclo Combinado. Así, por ejemplo, si el viento procede del suroeste, y sopla por tanto hacia el nordeste, tendrá una dirección de  $225^\circ$ . Y si una estación está situada justo al nordeste del Ciclo Combinado, su dirección será también  $225^\circ$ .

### 3.2. UN PRIMER EJEMPLO

En este ejemplo se ilustra un instante, correspondiente a las 10:00 horas del día 30 de marzo de 2009, en el que se había producido una alarma para NO<sub>x</sub> en la estación G5 (Figura 3.1).

Se trata de un ejemplo muy ilustrativo dado que durante los instantes previos habían estado soplando vientos suaves procedentes del este, con pequeñas variaciones tanto nordeste como sureste, lo que permite disponer de diferentes niveles de incidencia sobre las estaciones. Además se tiene en varias de ellas niveles de NO<sub>x</sub> por encima de  $n_0$ , que es cuando se aplican los modelos de clasificación.



**Figura 3.1** Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en G5 (en µg/m<sup>3</sup>).

A continuación se presentan los datos registrados ese día relativos a la dirección (Figura 3.2) y velocidad (Figura 3.3) del viento a 80 metros de altitud.

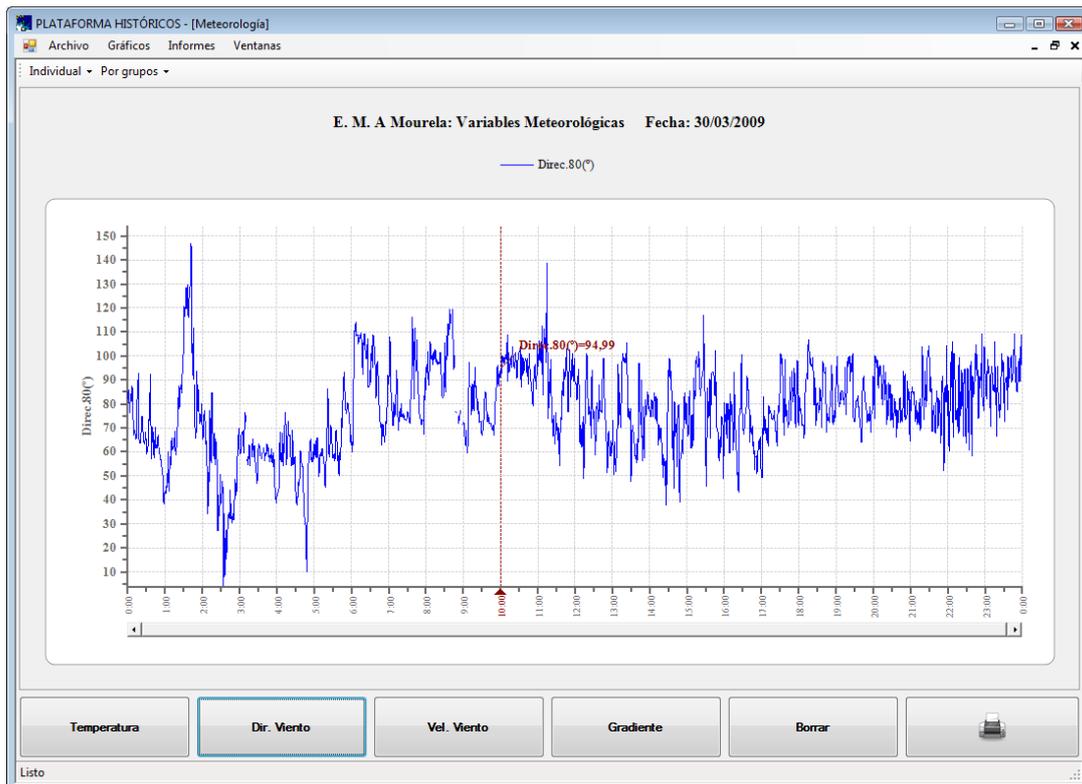


Figura 3.2 Dirección del viento a 80 metros (en grados).

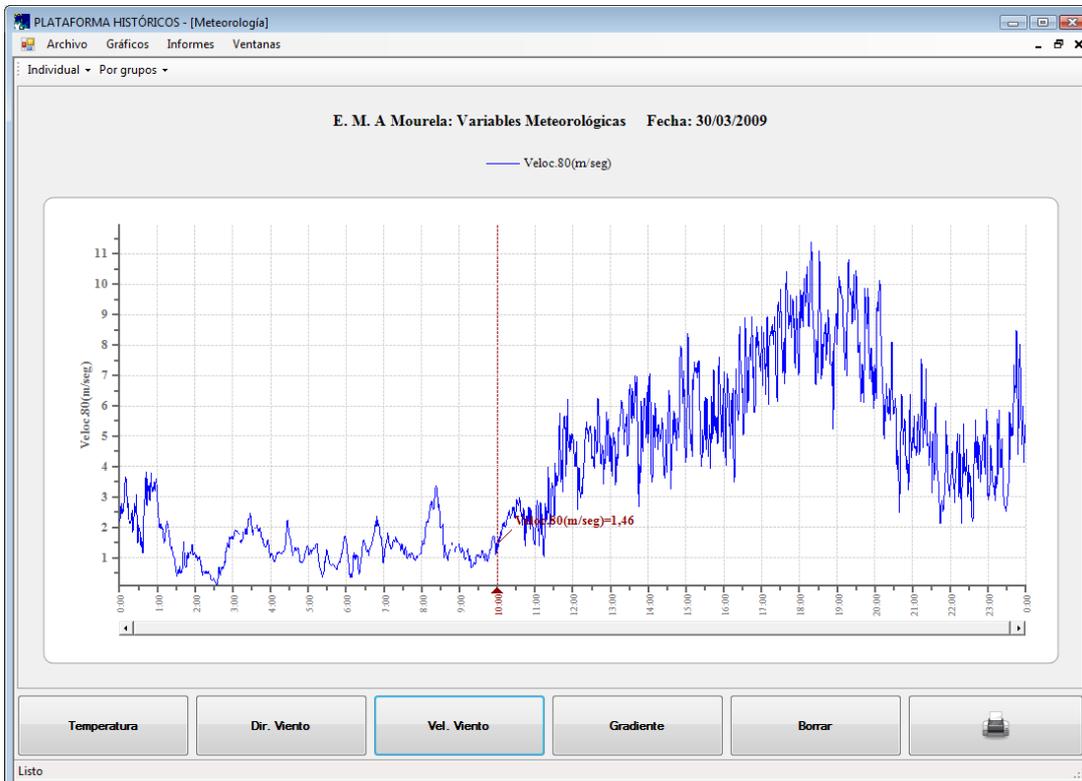


Figura 3.3 Velocidad del viento a 80 metros (en m/s).

La situación registrada por el sistema de predicción para ese instante se puede ver en la Figura 3.4.

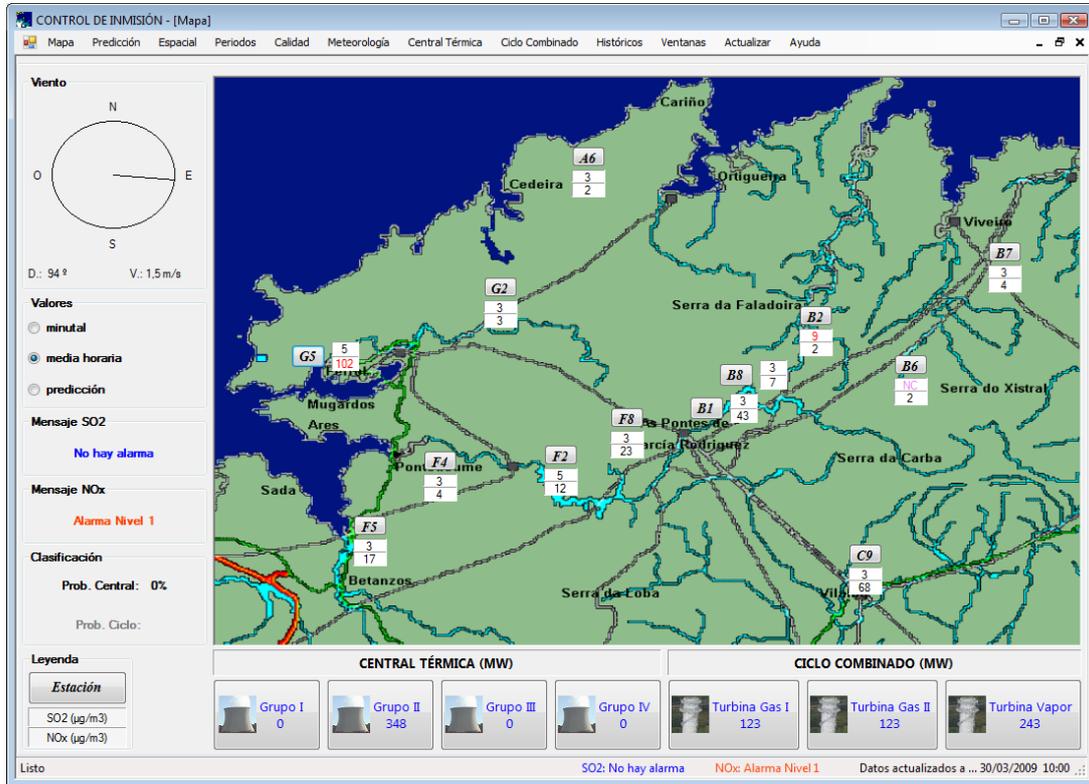


Figura 3.4 Situación registrada a las 10:00 horas del 30/03/2009.

En esta situación, los resultados obtenidos por los nuevos modelos propuestos son los que se recogen en la Tabla 3.2. En ella figuran las probabilidades estimadas de que los niveles de SO<sub>2</sub> y NO<sub>x</sub> observados en cada estación tengan su origen en la Central Térmica o en el Ciclo Combinado.

		B1	B2	B6	B7	B8	C9	G5	F2	F4	F5	F8	G2	A6
CT	SO <sub>2</sub>	--	--	--	--	--	--	--	--	--	--	--	--	--
	NO <sub>x</sub>	--	--	--	--	--	--	--	--	--	--	--	--	--
CC	SO <sub>2</sub>	--	--	--	--	--	--	--	--	--	--	--	--	--
	NO <sub>x</sub>	0	--	--	--	--	0	0	7.5	--	0	40.6	--	--

Tabla 3.2 Probabilidades estimadas por los modelos propuestos (en %).

Los resultados obtenidos parecen ajustarse adecuadamente a lo que cabría esperar.

Por una parte, solo se calculan las probabilidades en aquellas estaciones en que se están superando los niveles de referencia fijados ( $s_0 = 10$  y  $n_0 = 10$ ), que en este caso resultan ser seis y todas por  $\text{NO}_x$ .

En G5, que es la estación en la que se dio la alarma y tiene el nivel más elevado de  $\text{NO}_x$  con  $102 \mu\text{g}/\text{m}^3$ , la probabilidad estimada de que su origen se encuentre en el Ciclo Combinado resulta ser cero.

Esto, que podría resultar sorprendente estando además el viento soplando en esa dirección, no lo es tanto si se tiene en cuenta que la estación está situada en las proximidades de Ferrol y que, dada la hora, lo más probable es que los niveles registrados sean debidos al tráfico de vehículos. Además hay que considerar las bajas velocidades a las que soplaban el viento en los instantes previos y la gran distancia a que se encuentra.

En B1, estación urbana situada en As Pontes, se tiene un nivel de  $\text{NO}_x$  de  $43 \mu\text{g}/\text{m}^3$ . Considerando su situación y la dirección del viento en los instantes previos es de esperar que estos niveles no sean debidos a las emisiones del Ciclo Combinado. De hecho, lo más probable es que sean debidos también al tráfico de vehículos.

Algo parecido ocurre en C9, estación urbana próxima a Vilalba, que registra un nivel de  $68 \mu\text{g}/\text{m}^3$ .

En este sentido, las similitudes en el comportamiento observado en B1 (Figura 3.5), C9 (Figura 3.6) y G5 (Figura 3.1) resultan evidentes. Nótese que, aun siendo todas urbanas, el volumen de tráfico esperado en As Pontes, Vilalba y Ferrol no es el mismo.

La otra estación para la que se obtiene una probabilidad cero es F5, en la que se registran  $17 \mu\text{g}/\text{m}^3$  de  $\text{NO}_x$ .

El comportamiento en ésta estación es más complicado de explicar ya que, aunque está calificada como rural, pasa cerca la autopista AP9.

Los niveles observados podrían deberse a actividades agrícolas o industriales propias de la zona, al tráfico o un poco a todas ellas, sin que pueda descartarse incluso un posible fallo en el analizador o en la calibración del mismo. Véase sino el comportamiento errático reflejado en la Figura 3.7.

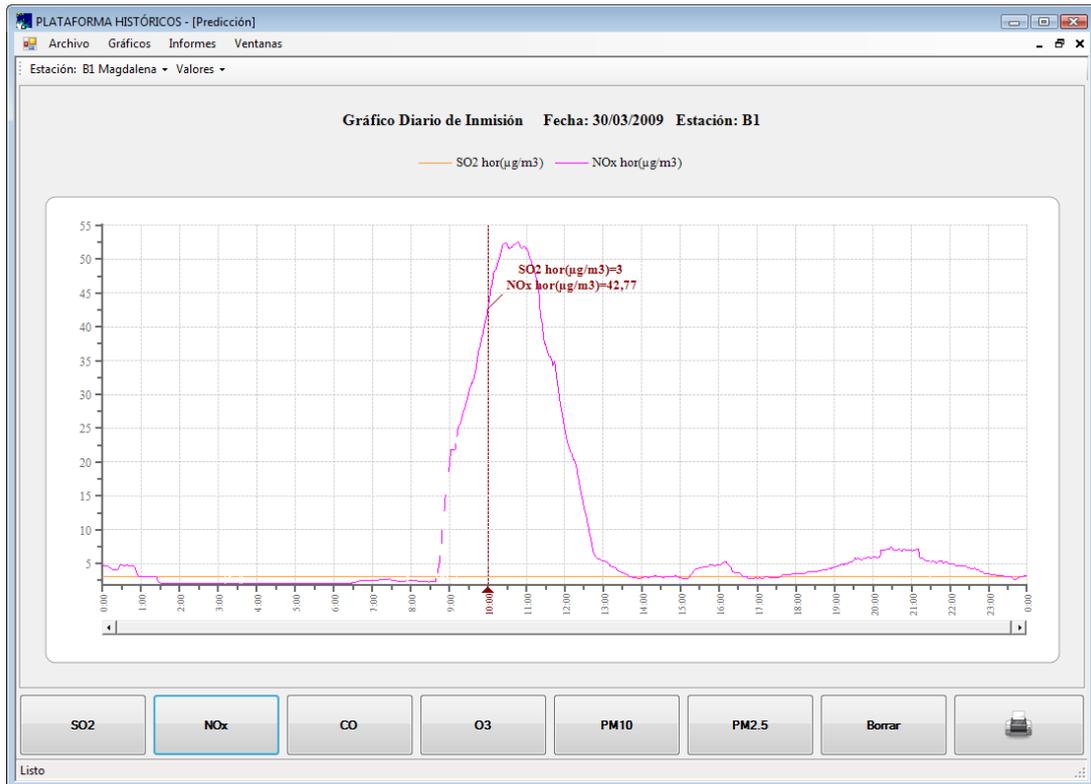


Figura 3.5 Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en B1 (en µg/m<sup>3</sup>).

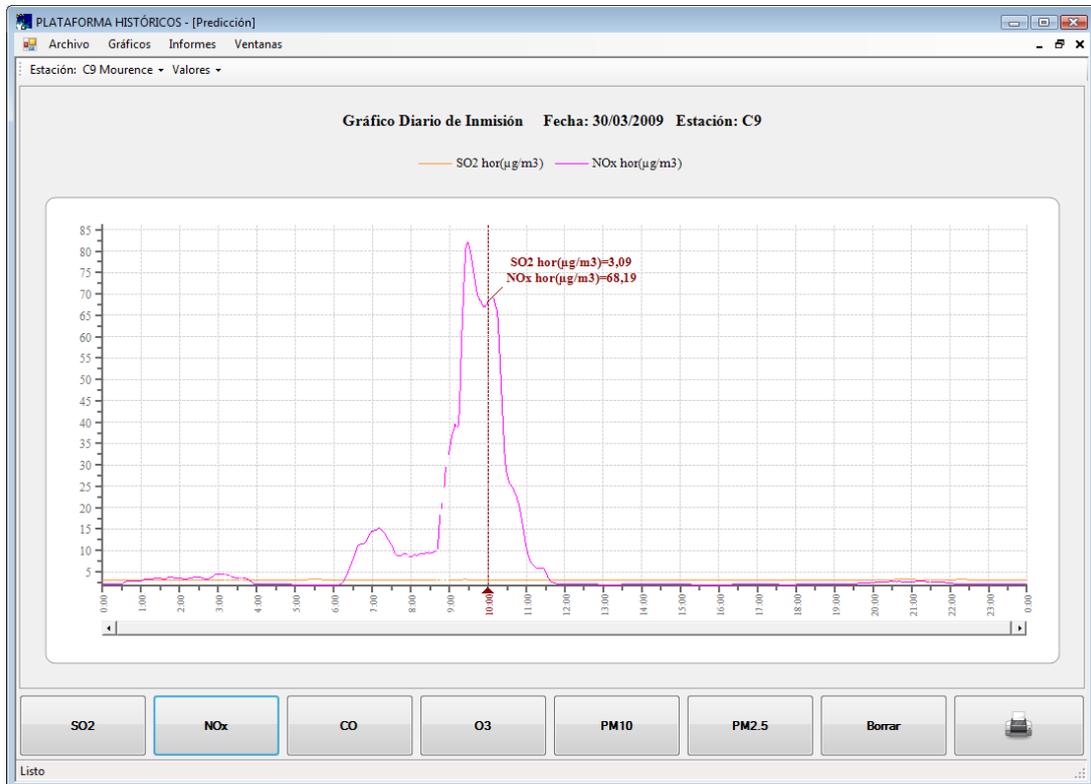
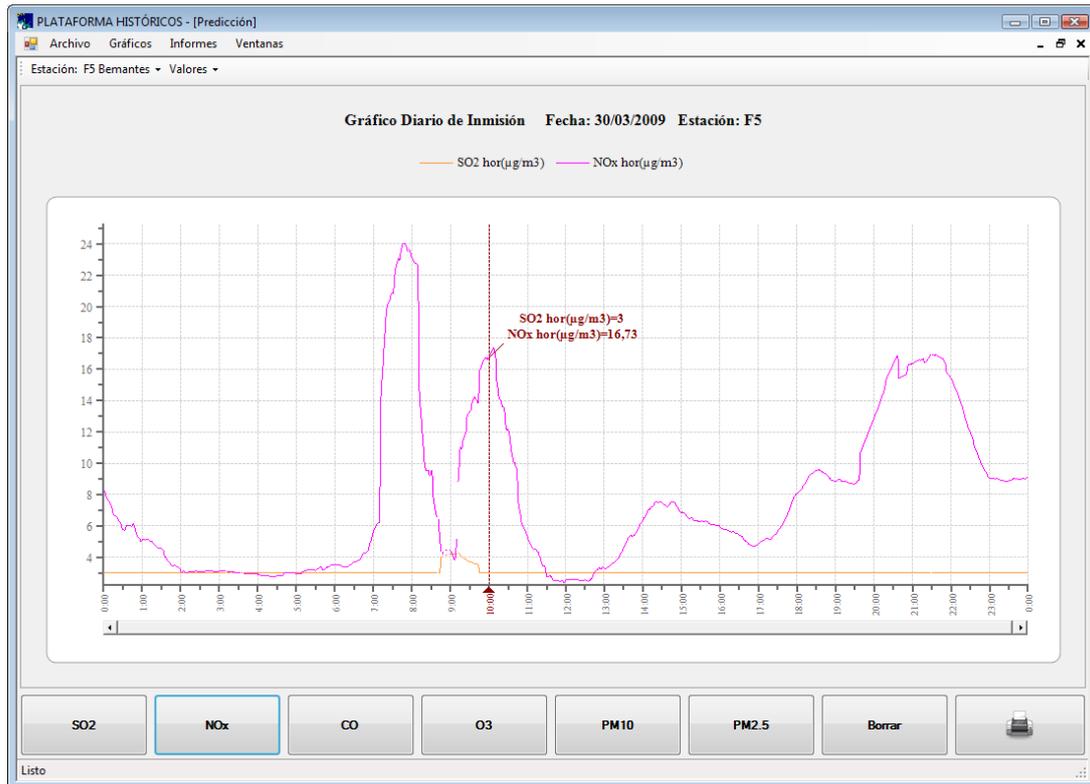


Figura 3.6 Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en C9 (en µg/m<sup>3</sup>).



**Figura 3.7** Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en F5 (en µg/m<sup>3</sup>).

Así, las dos únicas estaciones en las que se obtiene una probabilidad positiva son F8 y F2 con un 40.6% y 7.5%, respectivamente. Ambas situadas en la dirección hacia la que sopla el viento y a una distancia razonablemente próxima.

F8 está calificada como forestal y no tiene industrias ni carreteras próximas, por lo que parece más que razonable que se esté viendo afectada por las emisiones del Ciclo Combinado y, por tanto, al menos una parte de los 23 µg/m<sup>3</sup> de NO<sub>x</sub> observados puedan ser debidos a ésta. Su evolución puede verse en la Figura 3.8.

A F2, que está algo más alejada y calificada como rural/forestal, también parece razonable que esté llegando algo del Ciclo Combinado. Aunque, viendo la evolución en los niveles de SO<sub>2</sub>, incluso podría estar llegando algo de la Central Térmica (véase Figura 3.9). En cualquier caso, tratándose de un nivel de NO<sub>x</sub> de tan solo 12 µg/m<sup>3</sup> no es fácil determinar el origen con gran precisión, al igual que ocurría en F5.

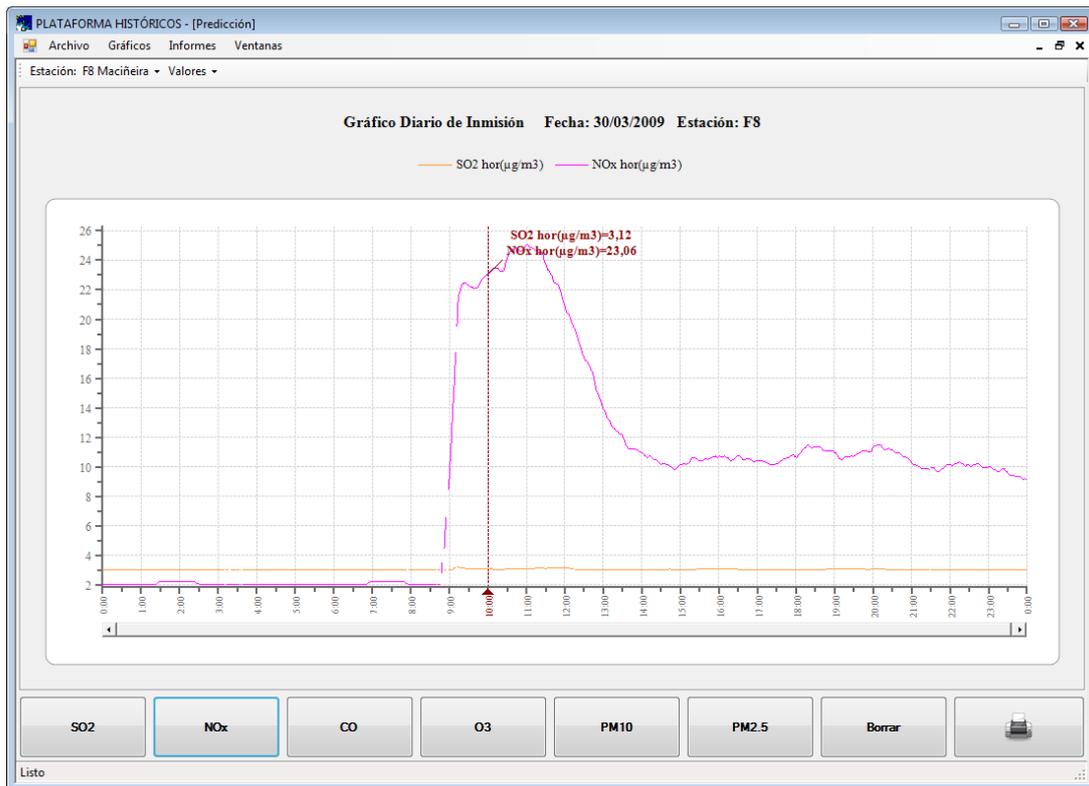


Figura 3.8 Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en F8 (en µg/m<sup>3</sup>).

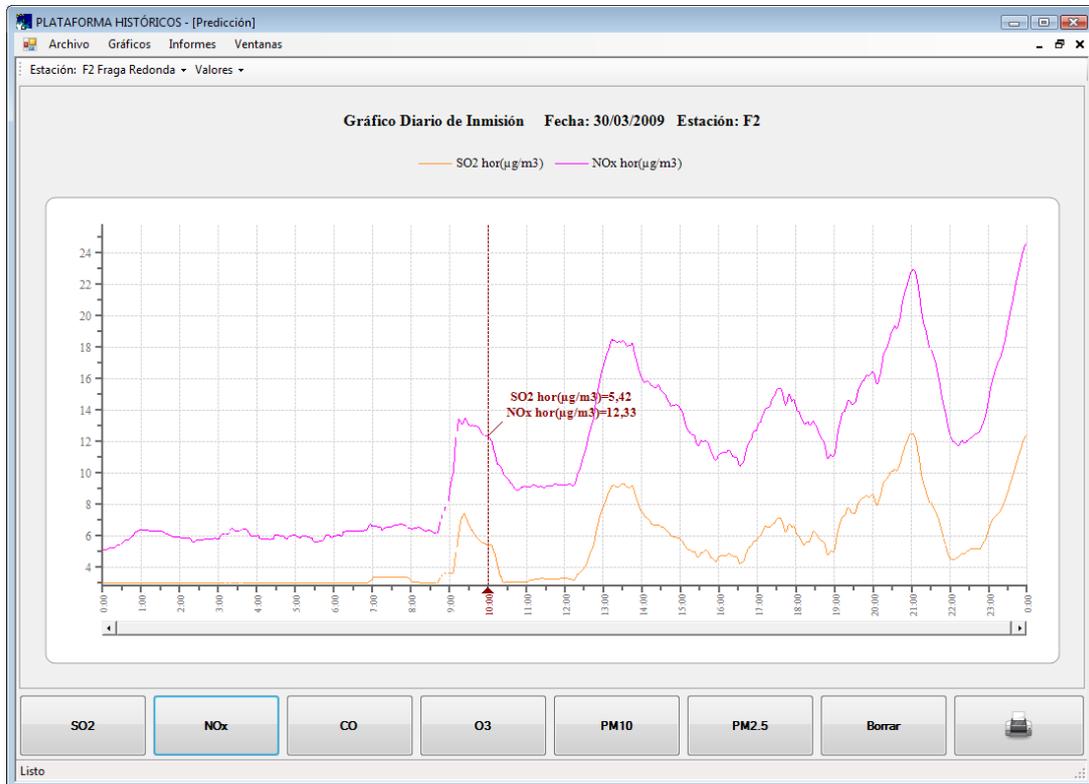


Figura 3.9 Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en F2 (en µg/m<sup>3</sup>).

A la vista de los resultados obtenidos, especialmente para F2, surge inevitablemente la inquietud por saber qué hubiese ocurrido si el nivel de SO<sub>2</sub> fuese un poco mayor, lo suficiente como para que los modelos lo tuviesen en cuenta y se calculase también la probabilidad de que el origen estuviese en la Central Térmica. Para que esto ocurriese tendría que observarse un nivel de SO<sub>2</sub> superior a 10, el fijado para s<sub>0</sub>. Cosa que no ocurre en F2, donde hay tan solo 5 µg/m<sup>3</sup>.

Veamos entonces lo que ocurre si se cambia el parámetro y se fija s<sub>0</sub> = 4. Los resultados tras el cambio son los que se recogen en la Tabla 3.3.

		B1	B2	B6	B7	B8	C9	G5	F2	F4	F5	F8	G2	A6
CT	SO <sub>2</sub>	--	0.8	--	--	--	--	0	8.4	--	--	--	--	--
	NO <sub>x</sub>	--	0	--	--	--	--	0	7.3	--	--	--	--	--
CC	SO <sub>2</sub>	--	--	--	--	--	--	--	--	--	--	--	--	--
	NO <sub>x</sub>	0	0	--	--	--	0	0	2.6	--	0	42.2	--	--

**Tabla 3.3** Probabilidades estimadas por los modelos propuestos (en %) para s<sub>0</sub> = 4.

Se observa como la reducción del criterio sobre la presencia de SO<sub>2</sub> ha hecho entrar en juego una nueva estación, B2, aunque no de manera especialmente significativa.

También ahora se ha calculado la incidencia de la Central Térmica en G5, resultando ser cero la probabilidad estimada en ambos casos.

Más interesante, como se esperaba, son los resultados obtenidos para F2. Ahora se está asignando una probabilidad a la Central Térmica tanto para SO<sub>2</sub> como para NO<sub>x</sub>, al tiempo que se resta incidencia al Ciclo Combinado.

Todo esto se debe a que los ratios de inmisión resultante no se llegan a parecer mucho al de emisión de la Central Térmica. En los niveles de emisión, como se puede ver en la Figura 3.10, el NO<sub>x</sub> se mantiene más bien algo por encima del de SO<sub>2</sub>.

En cualquier caso, tratándose de unos niveles tan sumamente bajos, tanto los resultados obtenidos como las conclusiones que se tomen a partir de ellos deben ir acompañados de una cautela extrema.

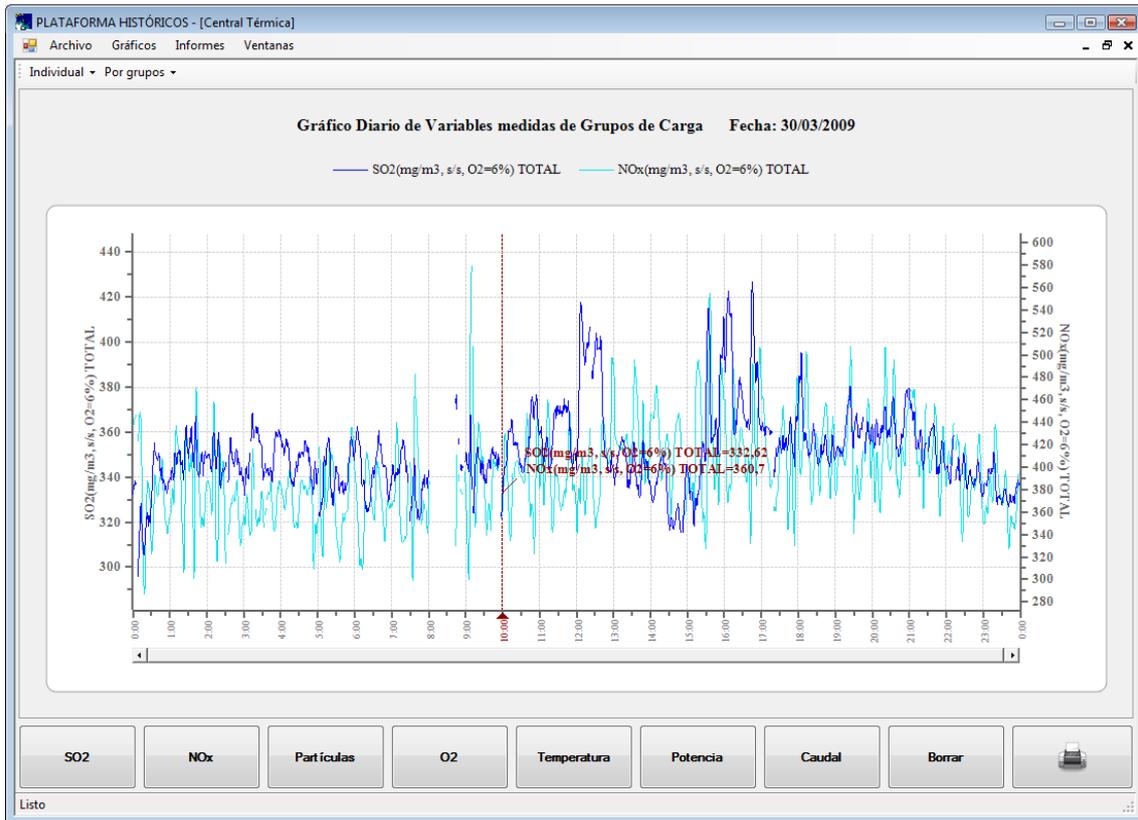


Figura 3.10 Niveles de emisión de SO<sub>2</sub> y NO<sub>x</sub> de la Central Térmica (en mg/m<sup>3</sup>).

### 3.3. UN SEGUNDO EJEMPLO

En este ejemplo se ilustra un instante, correspondiente a las 19:00 horas del día 2 de febrero de 2009, en el que se estaban registrando niveles elevados de  $\text{SO}_2$  en tres estaciones, B2, G2 y G5, siendo extremadamente elevados en la primera de ellas.

Con este ejemplo se pretende comprobar hasta qué punto los modelos son capaces de discriminar episodios atípicos o que nada tienen que ver con la actividad de las centrales.

En las siguientes imágenes (Figuras 3.11, 3.12 y 3.13) se puede apreciar el comportamiento de los niveles de inmisión en estas tres estaciones.

Y, a continuación, se presentan los datos registrados ese día relativos a la dirección (Figura 3.14) y velocidad (Figura 3.15) del viento a 80 metros de altitud.

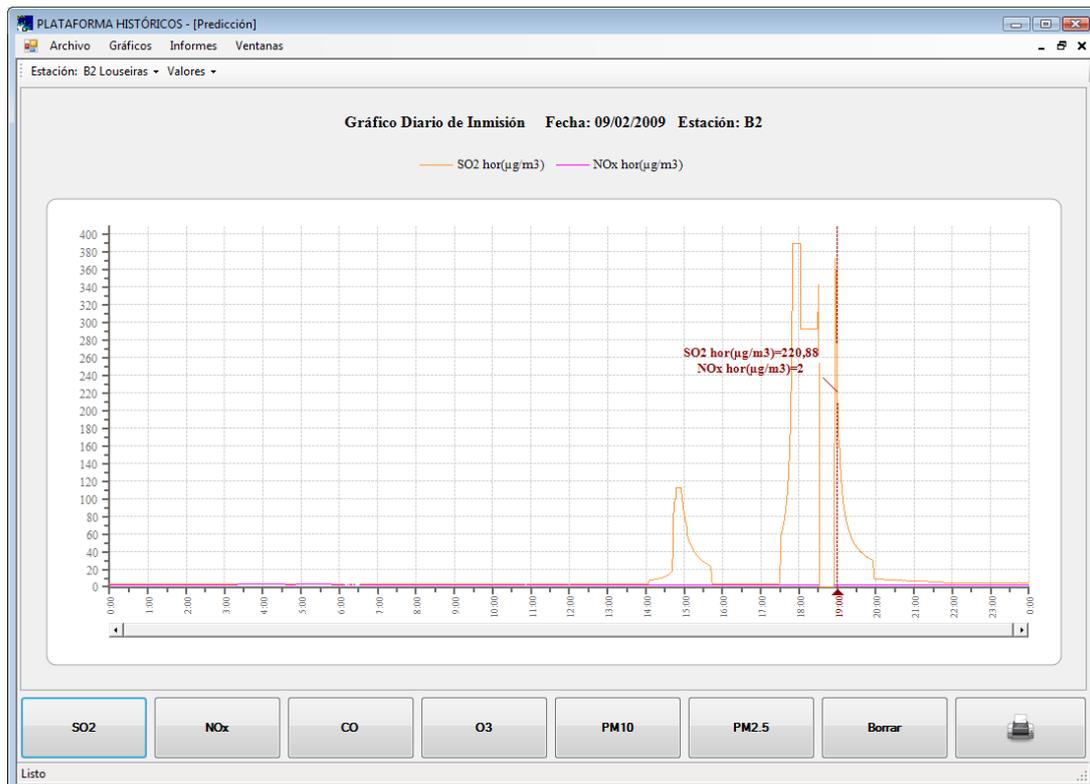


Figura 3.11 Niveles de inmisión de  $\text{SO}_2$  y  $\text{NO}_x$  en B2 (en  $\mu\text{g}/\text{m}^3$ ).



Figura 3.12 Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en G2 (en µg/m<sup>3</sup>).

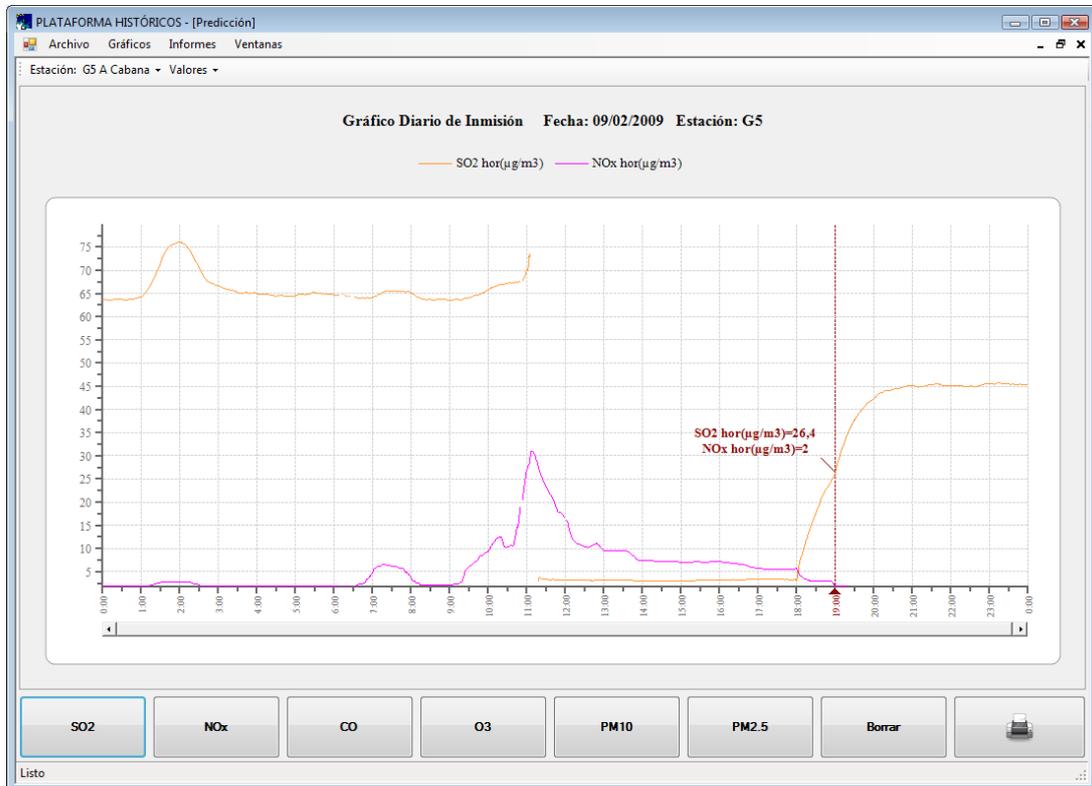


Figura 3.13 Niveles de inmisión de SO<sub>2</sub> y NO<sub>x</sub> en G5 (en µg/m<sup>3</sup>).

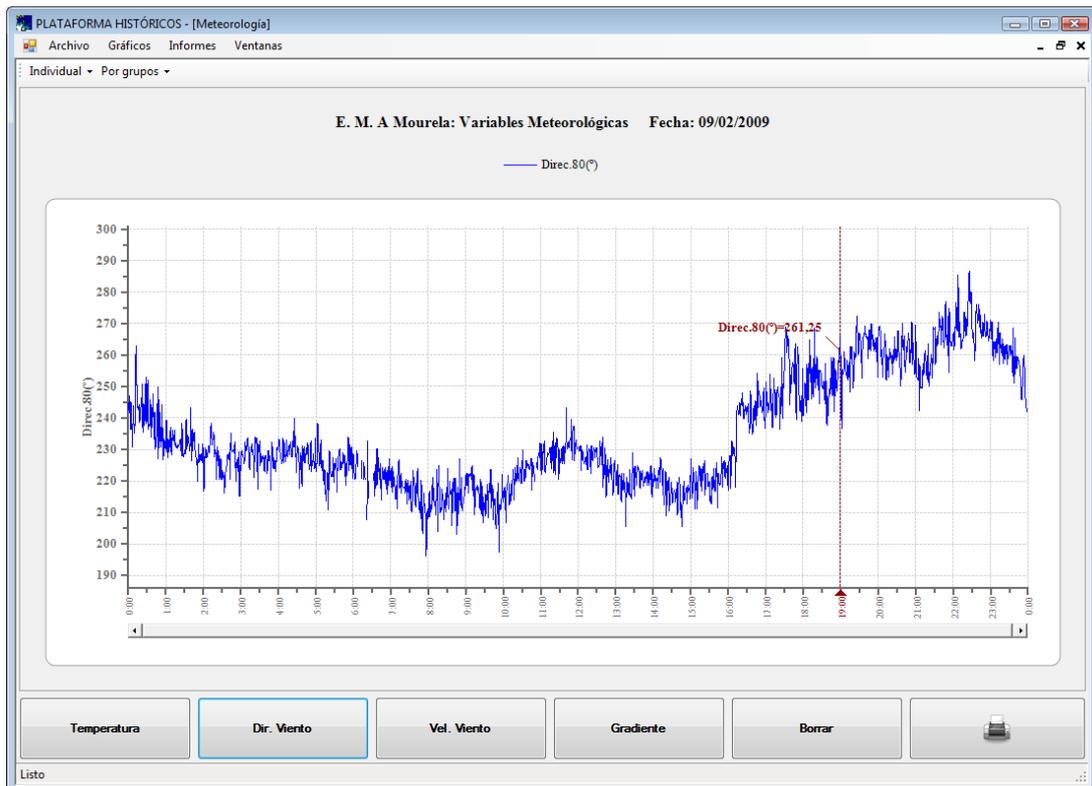


Figura 3.14 Dirección del viento a 80 metros (en grados).

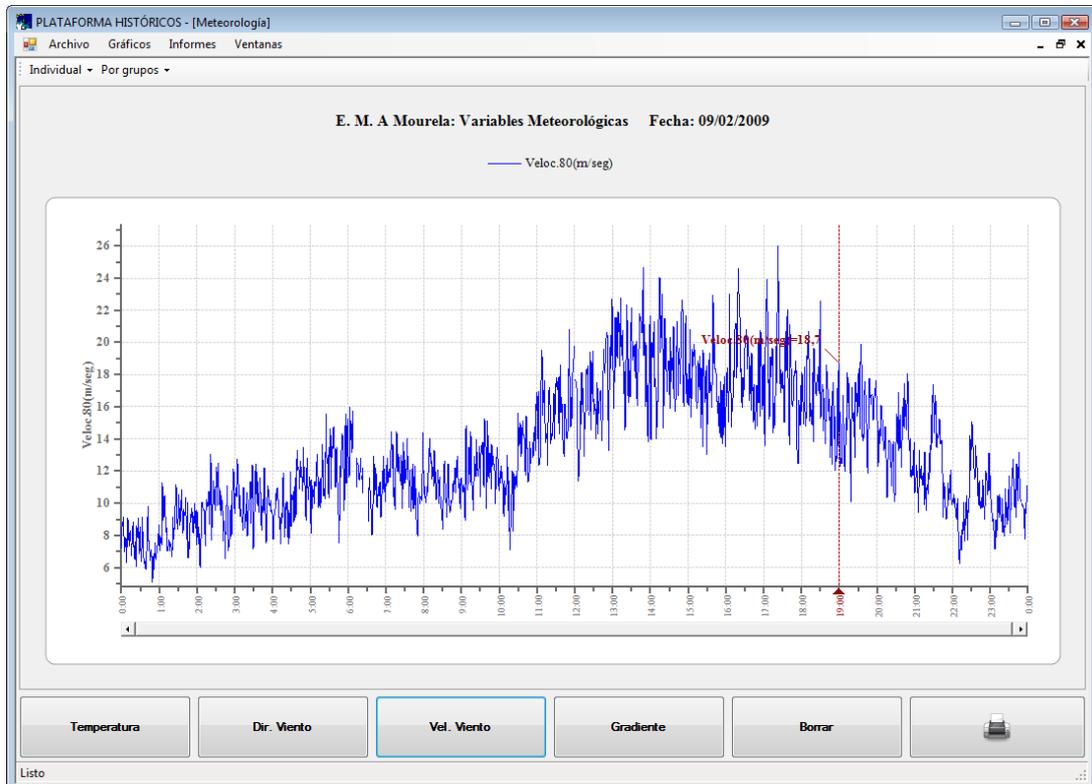


Figura 3.15 Velocidad del viento a 80 metros (en m/s).

La situación registrada por el sistema de predicción para ese instante se puede ver en la Figura 3.16.

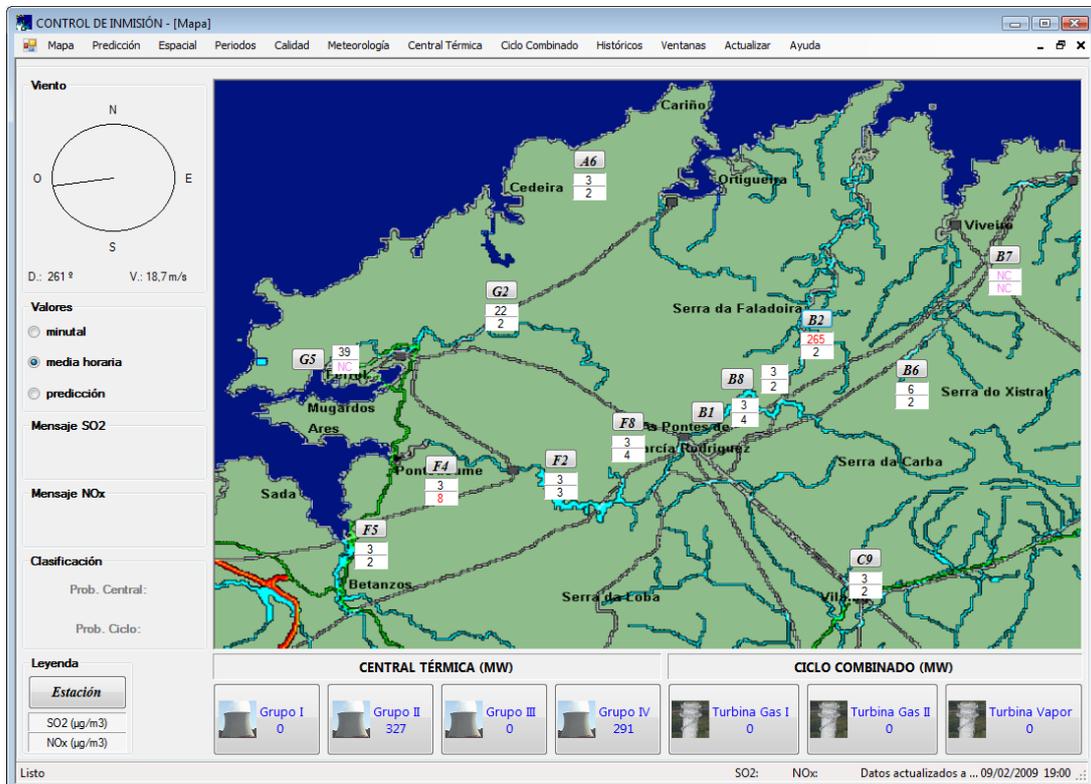


Figura 3.16 Situación registrada a las 19:00 horas del 09/02/2009.

En esta situación, los resultados obtenidos por los nuevos modelos propuestos son los que se recogen en la Tabla 3.4. En ella figuran las probabilidades estimadas de que los niveles de SO<sub>2</sub> y NO<sub>x</sub> observados en cada estación tengan su origen en la Central Térmica o en el Ciclo Combinado.

		B1	B2	B6	B7	B8	C9	G5	F2	F4	F5	F8	G2	A6
CT	SO <sub>2</sub>	--	0	--	--	--	--	0	--	--	--	--	0.1	--
	NO <sub>x</sub>	--	0	--	--	--	--	0	--	--	--	--	0	--
CC	SO <sub>2</sub>	--	--	--	--	--	--	--	--	--	--	--	--	--
	NO <sub>x</sub>	--	--	--	--	--	--	--	--	--	--	--	--	--

Tabla 3.4 Probabilidades estimadas por los modelos propuestos (en %).

Los resultados obtenidos parecen ajustarse adecuadamente a lo que cabría esperar.

Para la Central Térmica, a diferencia de lo que se observa en estas estaciones, los niveles de emisión de NO<sub>x</sub> son claramente superiores a los de SO<sub>2</sub> en todo momento, como se puede ver en la Figura 3.17.

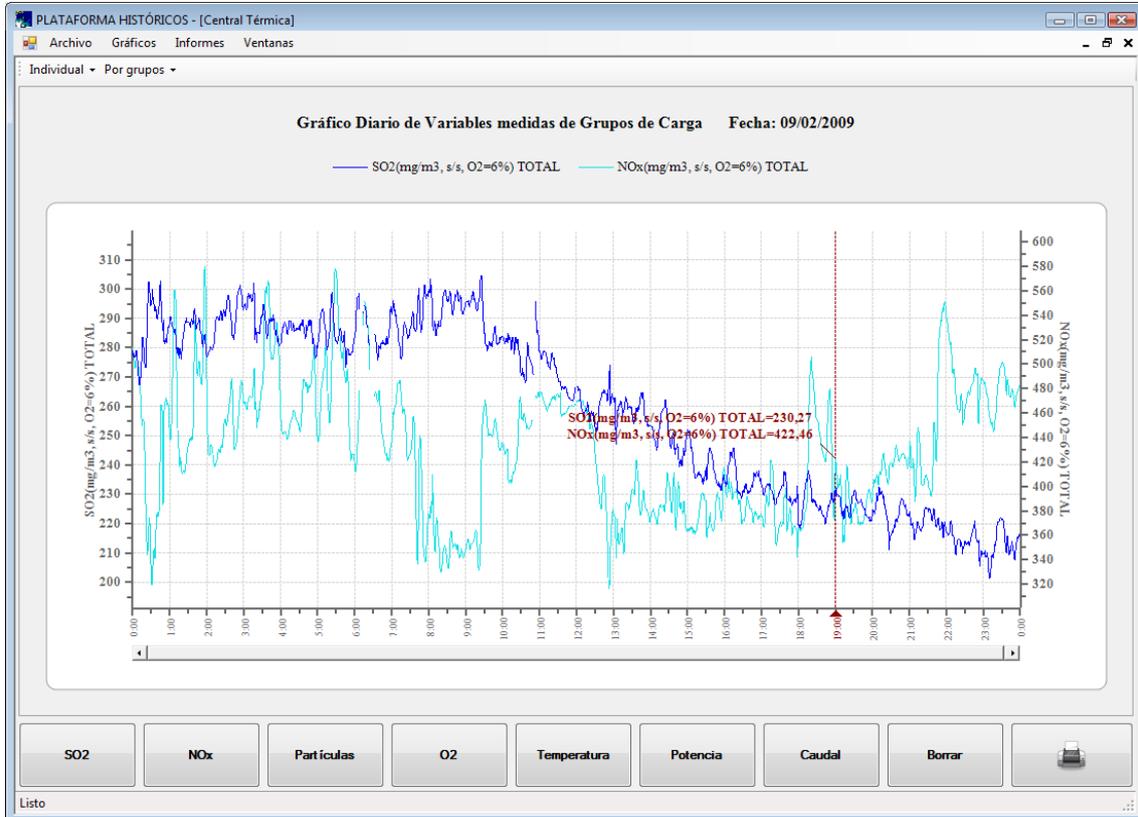


Figura 3.17 Niveles de emisión de SO<sub>2</sub> y NO<sub>x</sub> de la Central Térmica (en mg/m<sup>3</sup>).

## 4. OTROS MODELOS EN ESTUDIO

Para el desarrollo de los nuevos modelos presentados, en particular el de viento, ya se comentó que se habían considerado aspectos relacionados con el estudio de datos direccionales (véase Mardia, Kanti V. & Jupp, Peter E., 2000), siendo ésta una línea de investigación que habrá que explorar con mayor detalle en futuros desarrollos.

Otra línea en la también se ha estado trabajando, y en la que se ha avanzado algo más que en la anterior, es en el uso de datos funcionales (véase Ferraty, F. & Vieu, P., 2006 y Ramsay, J.O. & Silverman, B.W., 2005).

Para éstos se ha llegado incluso a plantear un modelo, pensado para ser utilizado dentro del esquema propuesto (Figura 2.2) en lugar del modelo mixto cuando no hay SO<sub>2</sub>.

Se presenta a continuación.

### Modelo funcional

Es un modelo para la estimación del NO<sub>x</sub> en el que se usan datos funcionales.

Se utilizan para realizar estimaciones para el Ciclo Combinado.

Permite obtener  $\hat{p}_{CC}^n$ , probabilidad estimada de que el origen del NO<sub>x</sub> sea el Ciclo Combinado.

Se asume que no hay presencia de niveles ni de SO<sub>2</sub> ni de NO<sub>x</sub> imputables a la Central Térmica ni de SO<sub>2</sub> imputables al Ciclo Combinado y, por tanto, sus probabilidades son nulas. Es decir

$$s_{CT} = n_{CT} = s_{CC} = 0 \quad \text{y} \quad p_{CT}^s = p_{CT}^n = p_{CC}^s = 0$$

Lo que se hace es lo siguiente:

- Se considera  $\mathbf{n}_F^{fh}$ , datos funcionales históricos del NO<sub>x</sub> de fondo en una estación  $e$ .
- Se considera  $\mathbf{n}^f$ , dato funcional del NO<sub>x</sub> en la estación  $e$  correspondiente a un intervalo de tiempo  $[t_2, t]$ , con  $t_2$  convenientemente seleccionado.

- Se extrae de  $\mathbf{n}_F^{fh}$  una muestra restringida a  $[t_2, t]$  para poder compararla con  $\mathbf{n}^f$ .
- Se asigna la probabilidad de que el origen esté en el Ciclo Combinado en función de una medida de profundidad  $\delta^f$ , convenientemente seleccionada, de  $\mathbf{n}^f$  en la muestra considerada

$$\hat{p}_{CC}^n = \frac{\delta_1^f}{\delta_1^f + \delta_2^f}$$

siendo

$$\delta_1^f = \delta^f(\mathbf{n}^f, 95\% \text{ más cercano de } \mathbf{n}_F^{fh})$$

$$\delta_2^f = \delta^f(95\% \text{ más cercano de } \mathbf{n}_F^{fh}, \text{ más profundo de } \mathbf{n}_F^{fh})$$

## REFERENCIAS

Efron, B. & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Chapman & Hall.

Ferraty, F. & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

Google Earth. Google. <http://www.google.com>.

Mardia, Kanti V. & Jupp, Peter E. (2000). *Directional Statistics*. John Wiley & Sons Ltd.

Piñeiro Lamas, María (2007). *Predicción de los niveles de SO<sub>2</sub> y NO<sub>x</sub> en el entorno de una central térmica*. Trabajo de Investigación Tutelado. Máster de Bioestadística 2005-2007. Departamento de Estadística e Investigación Operativa. Universidad de Santiago de Compostela.

Proymasa-Endesa (2006). *Redefinición de la Red de Inmisión de Endesa en As Pontes*.

R Project. The R Foundation for Statistical Computing. <http://www.r-project.org>.

Ramsay, J.O. & Silverman, B.W. (2005). *Functional Data Analysis*. 2<sup>nd</sup> edn. Springer.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Wand, M.P. & Jones, M.C. (1995). *Kernel Smoothing*. Chapman & Hall.

## ANEXO I: METODOLOGÍA

### I.A ESTIMACIÓN NO PARAMÉTRICA DE LA DENSIDAD

Dada una variable aleatoria  $X$ , si su función de distribución poblacional  $F$  es absolutamente continua con densidad  $f$ , ésta se podría estimar mediante la integral del estimador no paramétrico de Parzen-Rosemblat de  $f$

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t) dt$$

donde

$$\hat{f}_h(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t - x_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(t - x_i)$$

siendo  $K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$  y  $\vec{x} = (x_1, \dots, x_n)^t$  una muestra aleatoria de  $X$ .

Y se tendrá que

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t) dt = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^x K_h(t - x_i) dt = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{x-x_i} K_h(z) dz$$

A  $\hat{f}_h$  así definido se lo conoce como estimador no paramétrico de la densidad tipo núcleo, que depende de la función núcleo  $K$  y el parámetro ventana  $h$ .

En particular, se puede considerar un núcleo gaussiano de la forma

$$K(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

y utilizar, para la selección del parámetro ventana, el “método del pulgar” definido en Silverman, B.W. (1986), que propone tomar

$$\hat{h} = 0.9 \min\left\{\hat{\sigma}, \frac{R}{1.34}\right\} n^{-1/5}$$

donde  $\hat{\sigma}$ ,  $R$  y  $n$  son, respectivamente, la desviación típica, el rango intercuartílico y el tamaño de la muestra.

Aspectos más detallados relativos a la estimación no paramétrica de la densidad pueden verse en Silverman, B.W. (1986) y Wand, M.P. & Jones, M.C. (1995).

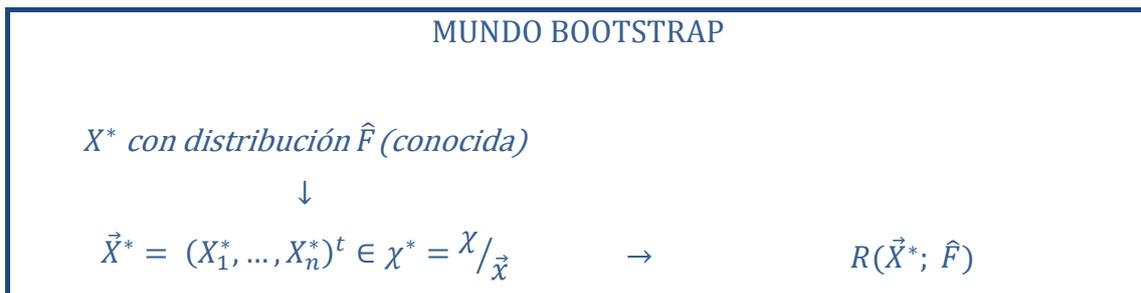
### I.B BOOSTRAP SUAVIZADO

El **Bootstrap** es un procedimiento estadístico que modeliza el muestreo de una población mediante el remuestreo de una muestra.



Como  $F$  es desconocida, la distribución de  $R(\vec{X}; F)$  será, presumiblemente, también desconocida.

Se observa una muestra  $\vec{x} = (x_1, \dots, x_n)^t$  y se estima  $F$  con  $\hat{F}$  a partir de  $\vec{x}$



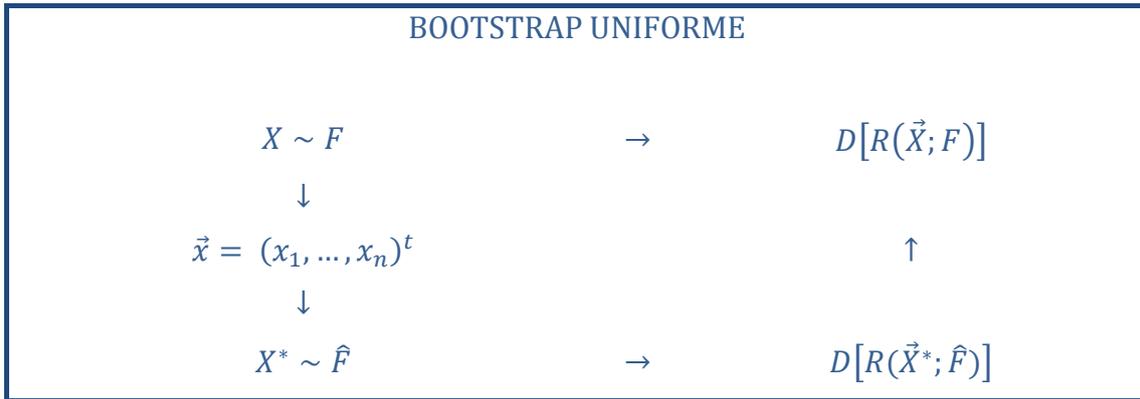
Como  $\hat{F}$  es conocida, la distribución de  $R(\vec{X}^*; \hat{F})$  será conocida o aproximable por Monte Carlo.

Y así se podrá aproximar la distribución de  $R(\vec{X}; F)$  por la de  $R(\vec{X}^*; \hat{F})$ .

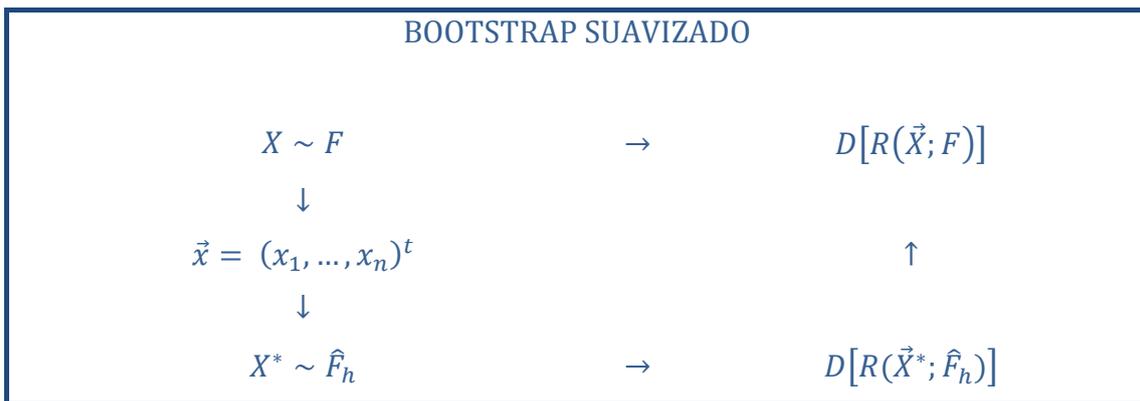
El **Bootstrap uniforme** se basa en la estimación de la función de distribución poblacional utilizando la función de distribución empírica.

Es decir, a partir de una muestra  $\vec{x} = (x_1, \dots, x_n)^t$  se estima  $F$  por:

$$\hat{F}(x) = \frac{n^{\circ} \text{ obs. } x_i \leq x}{n} = \frac{1}{n} \sum_{i=1}^n I_{\{x_i \leq x\}}$$



Se conoce como **Bootstrap suavizado** al planteamiento resultante de reemplazar  $\hat{F}$  por  $\hat{F}_h$ .



Para plantear el procedimiento de remuestreo se considera lo siguiente:

- Sea  $Y$  una variable aleatoria con función de densidad  $K$ .
- Considérese  $T = hY$  y obsérvese que, tal como se ha definido,  $K_h$  es la densidad de  $T$ .
- Sea  $\vec{x} = (x_1, \dots, x_n)^t$  una muestra de la variable de interés y  $X_0$  una variable aleatoria cuya distribución se corresponde con la empírica para esa muestra, es decir,  $X_0 \sim \hat{F}$ .

Teniendo en cuenta esto se tiene que  $\hat{F}_h$  es la función de distribución de  $hY + X_0$ .

De esta forma, las remuestras se construyen como

$$\vec{x}^* = (x_1^*, \dots, x_n^*)^t$$

donde para cada  $i = 1, \dots, n$

$$x_i^* = hy_i + z_i$$

proviendo  $y_i$  de una muestra de  $Y$  (con densidad  $K$ ) y  $z_i$  de una muestra de  $X_0$  (con distribución  $\hat{F}$ ).

Aspectos más detallados relativos al Bootstrap pueden verse en Efron, B. & Tibshirani, R.J. (1993).

## ANEXO II: CÓDIGO R

```

# ----- #
# ----- Rutina para calcular ----- #
# ----- el origen de un episodio ----- #
# ----- #

# -----
# Obtención de una muestra truncada a partir de una estimación no paramétrica de la densidad

# x    Datos de los que se desea extraer la muestra
# n    Número de datos que se van a simular
# a    Valor de truncamiento

rnpt=function(x,n,a){
  val=numeric(n)
  x=x[!is.na(x)]
  lx=length(x)
  h=bw.nrd0(x)

  i=1
  while(i<=n){
    l=sample(1:lx,1)
    aux=x[l]+h*rnorm(1)
    if (aux>a){
      val[i]=aux
      i=i+1
    }
  }

  return(val)
}

# -----
# Cálculo de una muestra de fondo

# MODELO HISTÓRICO

```

---

```

# his_f    Datos históricos de fondo en la estación
# vmin    Valor mínimo prefijado para la medición de la variable
# ndatos  Número de datos que se van a simular

mod_historico=function(his_f,vmin,ndatos){

  prop_e=length(his_f[his_f==vmin])/length(his_f[!is.na(his_f)])
  runi=runif(ndatos)

  mues_f=c(rep(vmin,length(runi[runi<=prop_e])),rnt(x=his_f[his_f!=vmin],n=length(runi[runi>prop_e]),a=vmin))

  return(mues_f)
}

# -----
# Cálculo de la probabilidad utilizando ratios

# MODELO DE RATIOS
# inmi_e   Datos de inmisión de la estación
# emi_ct   Datos de emisión de la Central Térmica
# mues_f   Muestra de fondo
# mues_cc  Muestra del Ciclo Combinado
# nmin     Valor mínimo prefijado para la medición de NOx
# ndatos  Número de datos que se van a simular

mod_ratio=function(inmi_e,emi_ct,mues_f,mues_cc,nmin,ndatos){

  # Ratios de inmisión
  inmi_e[inmi_e<=0]=NA
  mues_f[mues_f<nmin]=nmin
  mues_cc[mues_cc<0]=0
  mues_sin_fcc=inmi_e[2]-mues_f-mues_cc
  mues_sin_fcc[mues_sin_fcc<nmin]=nmin
  ratio_inmi=array(0,dim=c(ndatos,2))
  ratio_inmi[,1]=inmi_e[1]/mues_sin_fcc
  ratio_inmi[,2]=mues_sin_fcc/inmi_e[1]

```

```

# Ratios de emisión
emi_ct[emi_ct<=0]=NA
sum_ct=apply(emi_ct,c(1,2),sum,na.rm=T)
ratio_emi_ct=array(0,dim=c(ndatos,2))
ratio_emi_ct[,1]=sum_ct[,1]/sum_ct[,2]
ratio_emi_ct[,2]=sum_ct[,2]/sum_ct[,1]

# Estimación y comparación de las densidades
prob_ct=rep(-1,2)
auxp=seq(0,1,l=1000)
for (i in 1:2){
  limx=range(c(ratio_inmi[,i],ratio_emi_ct[,i]),na.rm=T)
  # Se necesitan al menos dos datos de inmición y dos de emisión
  if(length(ratio_inmi[!is.na(ratio_inmi[,i]),i])>1 &
length(ratio_emi_ct[!is.na(ratio_emi_ct[,i]),i])>1){
    dens_inmi=density(ratio_inmi[,i],from=limx[1],to=limx[2],na.rm=T)$y
    dens_emi_ct=density(ratio_emi_ct[,i],from=limx[1],to=limx[2],na.rm=T)$y
    aux=dens_inmi
    aux=apply((aux-outer(dens_emi_ct,auxp,""))^2,2,sum)
    prob_ct[i]=auxp[which.min(aux)]*100
  }
}

return(list(prob_ct=prob_ct))
}

# -----
# Cálculo de la incidencia del viento

# MODELO DE VIENTO
# dis_e      Distancia del Ciclo Combinado a la estación
# dir_e      Dirección (ángulo respecto al Norte) de la recta que une el Ciclo Combinado con la
estación
# dir_v      Dirección (ángulo respecto al Norte) del viento en cada instante
# vel_v      Velocidad del viento en cada instante
# cte       Constante para el cálculo de los grados de libertad
# ndatos     Número de datos que se van a simular

```

```

mod_viento=function(dis_e,dir_e,dir_v,vel_v,cte,ndatos){
  ninst=length(dir_v)
  nval=length(dir_v[!is.na(dir_v)])
  prob_v=-1

  if (nval>0){
    proy_v=rep(0,ninst)
    glib_e=rep(0,ninst)
    mues_e=array(0,dim=c(ndatos,ninst))

    for (i in 1:ninst){
      if (!is.na(dir_v[i]) & !is.na(vel_v[i])){
        proy_v[i]=vel_v[i]*cos((dir_v[i]-dir_e)*pi/180)
        if (proy_v[i]>0 | vel_v[i]==0){
          glib_e[i]=cte*log(exp(2/cte)+proy_v[i])
          mues_e[,i]=rchisq(ndatos,df=glib_e[i])
        }
      }
    }

    Fn=ecdf(mues_e)
    prob_v=(1-abs(2*Fn(dis_e)-1))*100
  }

  return(prob_v)
}

# -----
# Cálculo de la probabilidad utilizando ratios y proporciones

# MODELO MIXTO (Etapa 1)
# inmi_e    Datos de inmisión de la estación
# emi_ct    Datos de emisión de la Central Térmica
# mues_f    Muestra de fondo
# nmin      Valor mínimo prefijado para la medición de NOx

mod_mixto_1=function(inmi_e,emi_ct,mues_f,nmin){

```

```

inmi_e[inmi_e<=0]=NA
mues_f[mues_f<nmin]=nmin

# Ratios de emisión
emi_ct[emi_ct<=0]=NA
sum_ct=apply(emi_ct,c(1,2),sum,na.rm=T)
ratio_emi_ct=sum_ct[,1]/sum_ct[,2]

# Muestra de la Central Térmica
mues_ct=inmi_e[1]/ratio_emi_ct

# Muestra sin fondo y Central Térmica
mues_sin_fct=inmi_e[2]-mues_f-mues_ct
mues_sin_fct[mues_sin_fct<0]=0

return(list(mues_ct=mues_ct,mues_sin_fct=mues_sin_fct))
}

# MODELO MIXTO (Etapa 2)
# inmi_e    Datos de inmisión de la estación
# mues_ct   Muestra de la Central Térmica
# mues_ccof Muestra conjunta del Ciclo Combinado y otros posibles focos
# prob_v    Probabilidad de incidencia del viento

mod_mixto_2=function(inmi_e,mues_ct,mues_ccof,prob_v){
  inmi_e[inmi_e<=0]=NA

  prob_ct=rep(-1,2)
  prob_cc=rep(-1,2)
  mues_cc=prob_v*mues_ccof/100
  prob_ct[2]=median(mues_ct/inmi_e[2],na.rm=T)*100
  prob_cc[2]=median(mues_cc/inmi_e[2],na.rm=T)*100

  return(list(mues_cc=mues_cc,prob_ct=prob_ct,prob_cc=prob_cc))
}

# -----

```

---

```
# Cálculo del origen de un episodio
```

```
# RUTINA PRINCIPAL
```

```
# vinmi_e   Datos de inmisión de cada estación
# vemi_ct   Datos de emisión de la Central Térmica
# vemi_cc   Datos de emisión del Ciclo Combinado
# vhis_f    Datos históricos de fondo en cada estación
# vnom_e    Nombres de las estaciones
# vdis_e    Distancia del Ciclo Combinado a las estaciones
# vdir_e    Dirección (ángulo respecto al Norte) de la recta que une cada estación con el
            Ciclo Combinado
# vdir_v    Dirección (ángulo respecto al Norte) del viento en cada instante
# vvel_v    Velocidad del viento en cada instante
# cte       Constante para el cálculo de los grados de libertad
# nmin      Valor mínimo prefijado para la medición de NOx
# ndatos    Número de datos que se van a simular
# s0        Valor mínimo prefijado para el cálculo de SO2
# n0        Valor mínimo prefijado para el cálculo de NOx
```

```
origen=function(vinmi_e,vemi_ct,vemi_cc,vhis_f,vnom_e,vdis_e,vdir_e,vdir_v,vvel_v,cte,nmin,n
datos,s0,n0){
```

```
  nest=ncol(vinmi_e)
  vinmi_e[vinmi_e<=0]=NA
  prob=array(-1,dim=c(4,nest))
  dimnames(prob)=list(c("SO2_CT","NOx_CT","SO2_CC","NOx_CC"),vnom_e)
  prob[3,]=0
```

```
  vemi_cc[vemi_cc<0]=NA
  sum_cc=apply(vemi_cc,c(1,2),sum,na.rm=T)
  funciona_cc=ifelse(max(sum_cc[,2])>0,TRUE,FALSE)
```

```
  for (i in 1:nest){
    if (!is.na(vinmi_e[2,i])){
      vmues_f=mod_historico(vhis_f[,i],nmin,ndatos)

      if (!is.na(vinmi_e[1,i]) & vinmi_e[1,i]>s0){
        if (funciona_cc){
          aux=mod_mixto_1(vinmi_e[,i],vemi_ct,vmues_f,nmin)
          vmues_ct=aux$mues_ct
```

```

    vmues_ccof=aux$mues_sin_fct
    vprob_v=mod_viento(vdis_e[i],vdir_e[i],vdir_v[,i],vvel_v[,i],cte,ndatos)
    if (vprob_v>=0){
      aux=mod_mixto_2(vinmi_e[,i],vmues_ct,vmues_ccof,vprob_v)
      vmues_cc=aux$mues_cc
      prob[4,i]=aux$prob_cc[2]

prob[1:2,i]=mod_ratio(vinmi_e[,i],vemi_ct,vmues_f,vmues_cc,nmin,ndatos)$prob_ct
    }
  }
  else{
    prob[4,i]=0
    vmues_cc=rep(0,ndatos)
    prob[1:2,i]=mod_ratio(vinmi_e[,i],vemi_ct,vmues_f,vmues_cc,nmin,ndatos)$prob_ct
  }
}
else{
  if (funciona_cc){
    if (vinmi_e[2,i]>n0){
      vmues_ct=rep(0,ndatos)
      vmues_ccof=vinmi_e[2,i]-vmues_f
      vmues_ccof[vmues_ccof<0]=0
      vprob_v=mod_viento(vdis_e[i],vdir_e[i],vdir_v[,i],vvel_v[,i],cte,ndatos)
      if (vprob_v>=0){
        prob[4,i]=mod_mixto_2(vinmi_e[,i],vmues_ct,vmues_ccof,vprob_v)$prob_cc[2]
      }
    }
  }
  else{
    prob[4,i]=0
  }
}
}
}

return(prob)
}

```