En el municipio de As Pontes de García Rodríguez, al noroeste de la provincia de A Coruña, están situadas las Centrales Térmica y de Ciclo Combinado de As Pontes que constituyen uno de los centros productivos propiedad de Endesa Generación S.A.

Para estas centrales se dispone de un Sistema de Control Suplementario de la Contaminación Atmosférica que incluye la adquisición de datos de calidad de aire en tiempo real, su tratamiento y la realización de operaciones específicas que nos ayuden a la reducción de emisiones. Este sistema, permite adecuar las condiciones de operación al nivel de las emisiones, para tratar así de evitar posibles episodios de contaminación.

Ambas centrales disponen de una Red de Vigilancia de la Calidad atmosférica constituida actualmente por 13 estaciones automáticas y una estación meteorológica central, llamada Estación de A Mourela. Estas estaciones, proporcionan en modo continuo medidas de diversos contaminantes y variables meteorológicas.

 Hace años, nació una fructífera relación entre la Sección de Medio Ambiente de la U.P.T. As Pontes y el Departamento de Estadística e Investigación Operativa de la Universidad de Santiago de Compostela debido al compromiso de Endesa Generación S.A. con el medio ambiente y su interés por mejorar este sistema de predicción.

De esta relación nace el Sistema de Predicción Estadística de Inmisión (SIPEI) que permite obtener predicciones de los valores de dióxido de azufre y de óxidos de nitrógeno, con media hora de antelación, usando modelos aditivos. Este sistema, también predice cuál es el origen del episodio de alteración de calidad de aire, ya que este puede ser causado por la Central Térmica, el Ciclo Combinado u otros posibles focos como por ejemplo el tráfico o las actividades agrícolas de la zona. Esta última metodología se incorporó recientemente.

Años de dedicación, investigación, desarrollo y producción científica han dado fruto a nuevos modelos de predicción, nuevas herramientas de programación y notables mejoras en los sistemas de información.

1. Modelos de Predicción

La puesta en marcha de la nueva Central de Ciclo Combinado y la transformación de todos los grupos de la Central Térmica traen consigo la necesidad de predecir valores futuros de los óxidos de nitrógeno, además de los de dióxido de azufre, como ya se ha comentado la introducción. Los modelos que actualmente están en funcionamiento, las redes neuronales, fueron diseñados para la predicción de los niveles de SO2. Podríamos pensar en adaptar dichos modelos para obtener simultáneamente las predicciones de NOx, ya que se espera que los episodios de alteración de la calidad de aire provocados por este contaminante tengan un comportamiento similar a los causados por el SO2.

Por otro lado, además de predecir también nos interesa diseñar una herramienta que permita decidir cuál es el origen de los episodios de alteración de la calidad de aire: la nueva Central de Ciclo Combinado, la actual Central Térmica transformada, ambos potenciales emisores simultáneamente, o bien, ninguno de ellos (otros posibles focos). Las redes neuronales tienen una gran capacidad predictiva pero no nos permiten visualizar el proceso hasta la predicción, lo que supone un inconveniente ante esta nueva necesidad. Debido a esto vamos a utilizar nuevos modelos de predicción tanto para el NOx como para el SO2.

* 1. La Muestra de Trabajo

Para construir las muestras de trabajo, se ha considerado la idea de matriz histórica, utilizada para entrenar los anteriores modelos de predicción, como por ejemplo las redes neuronales. Dichas matrices son construidas con vectores

 y 

formados con datos reales de medias horarias de SO2 y NOx, respectivamente.

Las series de interés están formadas en su mayor parte por valores próximos a cero, por lo que no resulta apropiado elegir los datos de forma secuencial, ya que las muestras estarían formadas, probablemente, por muchos valores próximos a cero y pocos valores de episodios de inmisión. Puesto que el mayor interés está en predecir, precisamente, los episodios de inmisión, se han construido muestras de trabajo apropiadas para este fin que contienen toda la información relevante sobre los episodios de SO2 y NOx [[1]](#footnote-1).

Para ello se ha realizado un análisis densidades de los datos de inmisión del 2006 y se han construido estratos específicos para cada una de las estaciones de medida y para cada una de las variables de interés, SO2 y NOx. A cada uno de estos estratos se le asigna un rango de valores de  e , de forma que cada vector e  se colocará en el estrato que le corresponda a su valor  e , reemplazando al vector más antiguo, respectivamente.

Además, al construir estas nuevas matrices se han filtrado los datos de forma que no se introduzcan valores atípicos debidos a fallos en los aparatos de medida o a fallos eléctricos y se ha aumentado considerablemente el tamaño de dichas matrices: 8000 registros para el SO2 y de 4000 para el NOx.

Concretamente, los vectores seleccionados para rellenar la matriz histórica provienen de datos reales correspondientes a los años comprendidos entre 2003 y 2006 (ambos incluidos) para el SO2 y a los años 2005 y 2006 para el NOx. Como se puede observar hemos aumentado notablemente el tamaño de las matrices históricas.

* 1. Modelo Aditivo

En los últimos años ha surgido una línea de investigación en el campo de la estadística funcional no paramétrica que permite el desarrollo y aplicación de modelos más generales: los modelos aditivos.

Un campo donde han sido muy utilizados dichos modelos es en el estudio de las series de tiempo medioambientales. Por este motivo se van a utilizar para obtener las predicciones, a media hora, de los niveles de SO2 y NOx en el entorno de la Central Térmica y la Central de Ciclo Combinado.

El modelo planteado es



donde  representa el nivel medio horario del contaminante en el instante i,  es la constante desconocida y,  y  son funciones suaves desconocidas. Estimaremos el modelo de forma no paramétrica utilizando splines con penalizaciones y las matrices históricas correspondientes. El parámetro de suavizado, presente en cualquier metodología no paramétrica, se va a estimar utilizando el método de validación cruzada generalizado.

Hay que destacar que la estimación de los modelos se hace de forma independiente para cada uno de los dos contaminantes, es decir, por un lado utilizamos un modelo de la forma expuesta arriba y las correspondientes matrices históricas para obtener las predicciones de SO2 y por otro, utilizaremos otro modelo similar para obtener las del NOx.

Para poder observar el comportamiento del modelo aditivo seleccionado, hemos evaluado su funcionamiento sobre un episodio de alteración de la calidad de aire, cuya información no ha sido incluida en las matrices históricas. Así veremos si se comporta de forma adecuada ante situaciones reales a la hora de predecir tanto los valores futuros de SO2 como los de NOx.

Las Figuras 1 y 2 muestran las predicciones (a media hora) utilizando el modelo propuesto y la serie real observada para un episodio de alteración de la calidad de aire ocurrido el 12 de Marzo de 2007, para el SO2 y el NOx, respectivamente. En ambas se puede apreciar el buen comportamiento de las predicciones obtenidas por el modelo propuesto.



Figura 1: Episodio de alteración de la calidad del aire ocurrido en B6 el 12 de Marzo de 2007. Predicción dada por modelo aditivo.



Figura 2: Episodio de alteración de la calidad del aire ocurrido en G2 el 12 de Marzo de 2007. Predicción dada por modelo aditivo.

* 1. Comparación con otros Modelos utilizados

Se va a realizar una comparación entre el modelo aditivo, expuesto en el apartado anterior, y la red neuronal que actualmente está en funcionamiento, para la predicción de los valores medios horarios de SO2.

Para hacer la comparación de manera apropiada se ha estimado el modelo aditivo utilizando las matrices históricas con las que se ha entrenado la red neuronal. Dichas matrices están construidas con datos del 2003.

La Figura 3 muestra las predicciones dadas por el modelo aditivo así como las de la red neuronal para el episodio ocurrido en la estación B6 el 12 de Marzo de 2007. Parece que la predicción dada por el modelo aditivo persigue mejor a la serie real y además, no es tan sensible a los cambios como la red neuronal.

Estas consideraciones se confirman en la Tabla 1. Dicha tabla contiene dos medidas de precisión para cada uno de los dos predictores. Si se denota por et el error de predicción observado en el instante t, la raíz cuadrada del error cuadrático medio (MSE) viene dada por



y el error absoluto medio (MAE) por



|  |  |  |
| --- | --- | --- |
| **Modelo****Error** | **Red Neuronal****EC EA** | *Modelo Aditivo* *EC EA* |
| **M** |  3660.72 54.32 | 3024.55 43.17 |
| Md |  2556.82 50.56 | 1016.66 31.89 |

**Tabla 1:** Errores de predicción en el episodio de alteración de la calidad de aire ocurrido en B6 el 12 de Marzo de 2007. Medidas absolutas.

La tabla de errores confirma los resultados de la última figura: los errores, cuadrático y absoluto, cometidos por el modelo aditivo son mejores, en media y en mediana, que los cometidos por la red neuronal.



**Figura 3:** Episodio de alteración de la calidad del aire ocurrido en B6 el 12 de Marzo de 2007. Predicciones dadas por modelo aditivo y por la Red Neuronal.

## predicción

### Situación

En el periodo 2008-2009 la incorporación de la nueva central de ciclo combinado ha causado un importante impacto en los sistemas de predicción.

Esto ha obligado a que los esfuerzos metodológicos se hayan centrado en el desarrollo, implementación y validación de nuevos modelos capaces de predecir no solo niveles de SO2 sino también de NOX.

Además, durante este periodo se han desarrollado nuevas metodologías de clasificación. Pues se tiene interés en predecir si es la Central Térmica o el Ciclo Combinado es el origen de un episodio de alteración de la calidad del aire.

Dado que los modelos desarrollados utilizan para mejorar sus predicciones datos históricos, es necesario disponer de datos lo suficientemente fiables y representativos como para poder validarlos adecuadamente. Por lo que se necesita actualizar las matrices históricas cada cierto tiempo.

Actualmente, el sistema sigue funcionando en continuo aportando predicciones de los niveles de SO2 y NOX, calculando la probabilidad de que se produzca un episodio y, en caso de producirse, permite predecir si es la Central Térmica o el Ciclo Combinado el causante del mismo.

### Modelos de predicción

#### Situación

Actualmente, están implementados los **modelos aditivos** para predecir con media hora de antelación el comportamiento de las variables de inmisión de SO2 y NOX. Estas predicciones, se realizan de manera independiente para cada uno de ellos.

Tanto en las simulaciones realizadas como con los datos reales disponibles en la actualidad se han observado mejores resultados para estos modelos que para las redes neuronales utilizadas anteriormente.

En los aditivos se utilizan las matrices históricas construidas con niveles reales de inmisión en media horaria formando vectores del tipo:

$$\left(x\_{t-5},x\_{t},x\_{t+30}\right)^{T}$$

donde $x\_{i}$ representa el valor real en media horaria en el instante $i$.

Los datos son revisados y analizados para eliminar posibles datos erróneos o anómalos y seleccionar los más representativos. Pues dónde se encuentra la información más relevante es en los episodios, que es lo que queremos predecir, y no en las secuencias de valores establemente bajos.

Con estos datos, se definen estratos particularizados para cada estación y variable de interés asignando a cada uno de ellos un rango de valores de $x\_{t+30}$ que determinará a que estrato pertenece su correspondiente vector:

$$\left(x\_{t}-x\_{t-5},x\_{t},x\_{t+30}\right)^{T}$$

El modelo planteado es de la forma:

$$\hat{X}\_{t+30}=β\_{0}+s\_{1}\left(X\_{t}\right)+s\_{2}\left(X\_{t}-X\_{t-5}\right)$$

donde $X\_{i}$ representa el nivel medio horario del contaminante en el instante $i$, $β\_{0}$ es una constante desconocida, $s\_{1}$ y $s\_{2}$ son funciones suaves también desconocidas y $\hat{X}\_{t+30}$ es la estimación obtenida para 30 minutos después al instante actual.

La estimación del modelo se realiza de forma no paramétrica utilizando Splines con penalizaciones y las matrices históricas correspondientes.

Se usa el método de validación cruzada generalizado para estimar el parámetro de suavizado.

#### En estudio

Se están estudiando modelos semiparamétricos con respuesta bidimensional con el objetivo de seguir mejorando las predicciones. Estos modelos contemplan la existencia de una relación de dependencia en términos de **cointegración** entre las variables de SO2 y NOX. Con ello, se pretende modelizar una posible estructura de correlación que subyace bajo la presunción de existencia de un modelo de corrección de errores.

Siguiendo otra línea de investigación se continúa trabajando con **modelos funcionales**, que permitan obtener mejores predicciones tanto de SO2 como de NOX.

También continúan en estudio los modelos **datos direccionales** que tienen en cuenta variables meteorológicas como pueden ser la dirección y velocidad de viento, que intentarán convertirse en una nueva herramienta de predicción.

### Modelos de clasificación

#### Actuales

Recientemente se ha implantado un modelo de clasificación el cual proporciona, con frecuencia minutal, la probabilidad estimada de que el origen de un episodio de alteración de calidad de aire sea debido a la central térmica o a la de ciclo combinado.

Durante el periodo anterior, se implementó un primer modelo basado en el análisis de la relación existente entre los patrones de emisión e inmisión de SO2 y NOX (modelo inicial de ratios). Pero este modelo ya no es válido puesto que hay que tener en cuenta:

* Diferencias en los niveles de emisión de las centrales.
* Diferencias en las condiciones de emisión de las centrales, especialmente en la altura de las chimeneas.
* El SO2 pierde interés a favor del NOx.
* El modelo inicial no se puede aplicar al Ciclo Combinado.
* Es posible utilizar en el modelo variables meteorológicas (como la dirección o velocidad del viento).

Se quería construir un modelo lo más flexible posible, que utilizase el estudio de datos direccionales y modelos de difusión.

Bajo estas premisas se diseñó una estructura modular compuesta por cuatro modelos, cada uno dotado de identidad propia y orientado a resolver un problema concreto.

* Modelo de históricos

Es un modelo desarrollado exclusivamente para la generación de muestras artificiales del NOx de fondo a partir de los datos históricos de inmisión utilizando una estimación no paramétrica.

* Modelo de ratios

Es una evolución del modelo inicial de ratios en el que además del NOx de fondo se considera el del Ciclo Combinado.

Se utiliza para obtener las probabilidades estimadas de que

1. Hasta el momento los datos disponibles de NOx no son representativos puesto que el Ciclo Combinado aún no está en funcionamiento, por esto los modelos de predicción no podrán ser validados hasta el 2008. [↑](#footnote-ref-1)