

## Resumen del trabajo:

### **“Caracterización de patrones meteorológicos a escala regional y local y su relación con los niveles de calidad del aire registrados en la C.A.P.V. Análisis de episodios”**

**Ponente:** Amaia Albizuri, Responsable técnico del trabajo llevado a cabo por Environment and Systems, S.A. para Ihobe, S.A.

Environment and Systems, S.A. <sup>o</sup>/ Luis Briñas 9 1º izda 48013 Bilbao

<http://www.environmentandsystems.com> e-mail: [eyes@environmentandsystems.com](mailto:eyes@environmentandsystems.com)

## Introducción

La caracterización de patrones típicos de dispersión está detrás de muchos diseños de redes de sensores, sobre todo cuando lo que se pretende es vigilar y controlar las emisiones de focos puntuales emitidos a cierta altura.

En este sentido pueden encontrarse 2 líneas de trabajo: la modelización y las campañas experimentales (o la combinación de ambas).

La zona del Bilbao metropolitano fue pionera en España en utilizar campañas experimentales de medida para el diseño y optimización de la red de vigilancia de la calidad del aire a primeros de los años 80. Más adelante, esta metodología se puso en práctica en otras zonas como la de Cartagena y valle de Escombreras, muy castigada por la contaminación industrial, con muy buenos resultados, ya que la optimización de la red fruto de la caracterización de los patrones dispersivos típicos aportó el conocimiento suficiente como para adoptar medidas de control en situaciones episódicas con la credibilidad necesaria para ello.

En la línea de la caracterización de los patrones dispersivos a partir de modelos podemos enumerar la mayoría de los diseños de red para un foco dentro de los estudios de impacto ambiental al que son sometidos. Estos modelos son

necesarios en las fases preoperacionales aunque los resultados de los mismos deben ser validados luego, que no siempre se hace. Muchos focos son objeto de campañas experimentales una vez en marcha para optimizar su diseño de red preliminar adquiriendo un conocimiento real de las condiciones de dispersión y del impacto de sus emisiones.

Las campañas experimentales son muy caras aunque muy efectivas en las primeras fases de implantación de una red de medida de la calidad del aire.

Una vez que la red de sensores está en marcha se le supone capaz de aportar la información suficiente para ‘explicar’ los niveles registrados, es decir, capaz de caracterizar las relaciones causa-efecto y conocer los patrones dispersivos típicos de una zona, que en teoría han tenido que ser tenidos en cuenta en su diseño y posteriores optimizaciones o ampliaciones.

Es difícil encontrar ejemplos en los que las tareas de reconocimiento de patrones sean sistemáticas en una red y sea utilizado como herramienta de gestión en el área de Calidad del Aire. Es lo que se denomina sistema experto (una rama de la inteligencia artificial) y no es otra cosa que una herramienta matemática que incorpora algoritmos que explican las relaciones causa-efecto caracterizadas, capaz de llevar a cabo un diagnóstico e incluso una predicción de los niveles de contaminación.

Se encuentran muchos más ejemplos de utilización de este tipo de sistemas expertos en los temas relacionados con la gestión de los recursos hídricos y la prevención de inundaciones que en el área de la contaminación atmosférica (Chapter 10 “Decision Making for Environmental Management” in “Environmental Data Management”. 1998. Kluwer Academic Publishers)

La dificultad está en el desarrollo y posteriores mejoras de la herramienta, ya que requiere que sea específica y adaptada a un entorno, que sea alimentada por todos aquellos datos necesarios para explicar las relaciones previamente caracterizadas, que se disponga de una base de datos de varios años, que sea

validada y finalmente actualizada periódicamente en un proceso de mejora continua.

En la C.A.P.V. la Red de vigilancia y control de la calidad del aire ha sufrido importantes cambios siempre guiados por estudios específicos, como los relativos a los planes de saneamiento atmosférico en varias cuencas o los estudios de impacto ambiental de varias plantas de generación de energía e incineración de residuos. Se dispone de medidas en tiempo real, tanto de concentraciones de diversos contaminantes como de variables meteorológicas, necesarias para llevar a cabo un análisis diario de la situación encaminado a caracterizar el patrón dispersivo dominante. El objetivo es por tanto establecer, a partir de los datos actualmente adquiridos, las relaciones causa efecto, creando para ello una herramienta de análisis objetivo con una perspectiva que, en lo espacial, cubra desde la escala regional al detalle de las diferentes cuencas aéreas de la C.A.P.V., y en lo temporal describa, a poder ser, la evolución horaria.

Como es sabido muchos de los episodios de contaminación atmosférica en zonas de terreno complejo por su orografía y particularidades, como puede ser la cercanía al mar, son propiciados por unas malas condiciones de dispersión sin que ello signifique que las emisiones hayan aumentado. Incluso puede ocurrir que disminuyendo las mismas no se consiga una reducción suficiente para evitar el problema en la calidad del aire.

La pregunta que desde siempre se ha planteado es si es suficiente la reducción de los niveles de emisión para cumplir con los objetivos de calidad del aire y sobre todo si es lo más adecuado, por el coste y el esfuerzo necesario, si lo que se pretende no es disminuir los niveles de calidad del aire todos los días sino no sobrepasar el límite un número de veces en el que está permitido hacerlo al año.

Esto ha llevado históricamente a la discusión entre plantear estrategias de reducción o control primario de emisiones, reduciéndolas de forma permanente, o

secundario, reduciéndolas en aquellas situaciones de dispersión adversas. La mejor respuesta es que ambas estrategias son posibles y complementarias.

La caracterización de los patrones meteorológicos es necesaria para llevar a cabo las acciones que puedan derivarse de una estrategia de control secundario de emisiones. Los patrones meteorológicos caracterizados a partir de una base de datos con un nº suficiente de años permanecen casi invariables, las situaciones se repiten y la preferencia estacional y frecuencia de aparición pueden establecerse con bastante fiabilidad. Por tanto, los condicionantes de la dispersión de los contaminantes emitidos pueden conocerse si se dispone de la información adecuada.

Muchas de las estaciones de medida de la calidad del aire de la C.A.P.V. dispone entre su equipamiento sensores meteorológicos, para precisamente tratar de explicar los niveles registrados. En ocasiones la información está tan condicionada por el entorno más cercano que es difícil establecer relaciones, otras veces valores semejantes de algunos parámetros no significan situaciones de dispersión parecidas pues hay que tener en cuenta el origen de las masas aéreas a una escala mayor. Además, los procesos de mezcla en la capa límite atmosférica están condicionados por la componente vertical y su parametrización, y las torres meteorológicas de superficie no son suficientes para documentarla.

Aún llegando a la resolución de las cuencas aéreas, no se debe perder una visión más general del problema y de hecho muchas situaciones episódicas en la C.A.P.V. se dan simultáneamente en todo el territorio.

A continuación se detallan los datos de partida empleados, la metodología y las conclusiones, presentando finalmente algunos ejemplos de resultados.

## **Datos de partida**

Los datos de partida para este estudio han sido los valores registrados en la propia Red de vigilancia de la calidad del aire en la CAPV gestionada por la Dirección de Planificación, Evaluación y Control Ambiental del Gobierno Vasco.

El período de años a analizar va desde el año 2002 al 2006 (2002 a 2005 para la muestra de entrenamiento y 2006 para la muestra de comprobación) y los contaminantes elegidos: NO<sub>2</sub>, O<sub>3</sub> y PM<sub>10</sub>. En primer lugar se ha llevado a cabo una selección de estaciones y un control de calidad de datos.

Finalmente, las estaciones de medida de NO<sub>2</sub> utilizadas han sido 63: Muskiz, Abanto, Serantes, Náutica, Barakaldo, Siete Campas, Zorroza, María Díaz, Indautxu, San Adrián, Mazarredo, Txurdinaga, Basauri, Parque Europa, Elorrieta, Banderas, Erandio, Ondiz, Santa Ana, Getxo, Sangroniz, Algorta, Zierbena, Kastrexana, Arraiz, Alonsotegi, Arrigorriaga, Areta, Llodio, Amurrio (sólo en verano y otoño), Zalla, Durango, Lemoa, Larrea, Zelaieta, Larrabetzu, Montorra, Mondragón, Elgoibar, Hernani, Ategorrieta, Avda. Tolosa, Rentería, Lezo, Jaizkibel, Irún, Puio, Tolosa, Beasain, Azpeitia, Avda. Gasteiz, Tres de Marzo, Los Herrán, Fac. Farmacia, Betoño, San Martín, Agurain, Lantarón, Elciego, Valderejo, Izkiz, Mundaka y Pagoeta.

Para ozono las estaciones analizadas han sido 52: Muskiz, Abanto, Serantes, María Díaz, Mazarredo, Txurdinaga, Basauri, Parque Europa, Elorrieta, Banderas, Erandio, Getxo, Sangroniz, Algorta, Zierbena, Kastrexana, Arraiz, Alonsotegi, Arrigorriaga, Areta, Llodio, Zalla, Durango, Lemoa, Larrea, Zelaieta, Larrabetzu, Montorra, Mondragón, Elgoibar, Hernani, Ategorrieta, Avda. Tolosa, Puio, Rentería, Lezo, Jaizkibel, Irún, Tolosa, Beasain, Azpeitia, Avda. Gasteiz, Tres de Marzo, Fac. Farmacia, San Martín, Agurain, Elciego, Valderejo, Izkiz, Mundaka, Pagoeta y Urkiola.

Para el caso de PM<sub>10</sub> se han utilizado únicamente los datos a partir del 2004. El número de estaciones utilizadas es de 46: Abanto, Náutica, Barakaldo,

Indautxu, Mazarredo, Txurdinaga, Basauri, Parque Europa, Erandio, Getxo, Algorta, Zierbena, Kastrexana, Arraiz, Alonsotegi, Arrigorriaga, Areta, Llodio, Zalla, Durango, Lemoa, Larrea, Zelaieta, Mondragón, Hernani, Ategorrieta, Avda. Tolosa (se ha eliminado en el periodo de invierno por tener pocos datos), Puio, Rentería, Lezo, Irún, Tolosa, Beasain, Azpeitia, Avda. Gasteiz, Tres de Marzo, Fac Farmacia, Betoño, San Martín, Agurain, Lantarón, Elciego, Valderejo, Izkiz, Mundaka y Pagoeta.

En cuanto a las variables meteorológicas utilizadas han sido velocidad y dirección de viento y temperatura de 29 estaciones: Muskiz, Abanto, Serantes, Náutica, Barakaldo, Feria, Basauri, Banderas, Sangroniz, Algorta, Zierbena, Kastrexana, Arraiz, Llodio, Durango, Lemoa, Boroa, Hernani, Jaizkibel, Puio, Beasain, Azpeitia, Fac. Farmacia, Lantarón, Elciego, Valderejo, Izkiz, Mundaka y Pagoeta. También se ha utilizado la variable radiación medida en 12 estaciones: Muskiz, Basauri, Durango, Hernani, Beasain, Azpeitia, Fac. Farmacia, Lantarón, Valderejo, Izkiz, Mundaka y Pagoeta. Se han calculado además, las series temporales de la velocidad media diaria del viento y del salto térmico diario.

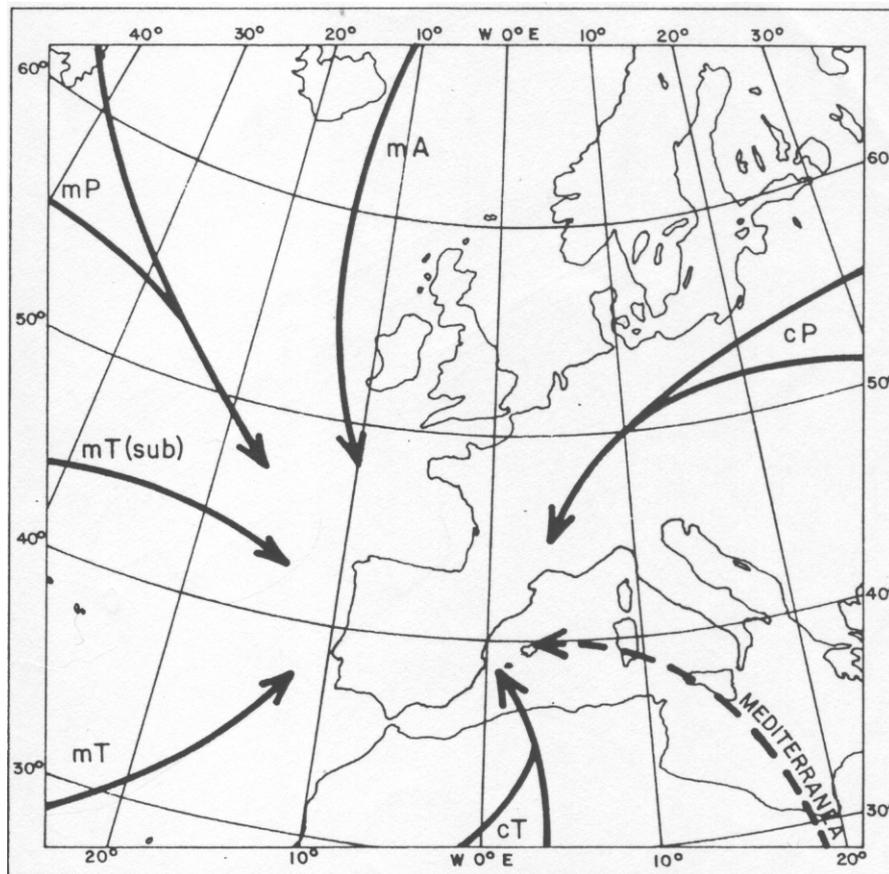
Los patrones característicos del viento en altura se han descrito finalmente a partir únicamente de los datos de 1500 m de altura medidos por el perfilador RADAR en el modo de baja resolución ubicado en Punta Galea, Getxo.

Para incluir información acerca de la precipitación se ha calculado la variable precipitación diaria acumulada en la estación de Mondragón de la Dirección de Meteorología y Climatología del Gobierno Vasco.

Se ha creado una serie temporal de cadencia diaria que nos indica la procedencia de la masa aérea que afecta a la C.A.P.V. estimada a partir de los mapas sinópticos y con la ayuda de los resúmenes mensuales publicados por la Dirección de Meteorología y Climatología del Gobierno Vasco.

Las distintas masas aéreas consideradas son (Figura 1): mA (marítima Ártico), mP (marítima polar, Atlántico norte), mTs (marítima subtropical, Atlántico

subtropical), mT (marítima tropical, Atlántico tropical), mM (marítima Mediterráneo), cP (continental polar, Europa), cT (continental cálida, Norte de África), iC (ibérica cálida) e iF (ibérica fría).



**Figura 1.** Procedencia de las masas aéreas con destino la península ibérica.

Los indicativos que empleamos para las distintas masas de aire, corresponden a la nomenclatura internacional, en la que las letras minúsculas indican su origen (m marítima y c continental) y las mayúsculas su carácter térmico (T cálida o tropical y P y A frías o, respectivamente, polar y ártica).

Otra circunstancia que debe tenerse en cuenta es el tiempo que la masa de aire cT o cP haya podido permanecer sobre el mar Mediterráneo antes de alcanzar la península. Por ello se hace una nueva distinción: mM, si el recorrido sobre el Mediterráneo es grande. Finalmente se han añadido dos categorías más: la iC y la iF, si estamos ante la circunstancia del dominio de las altas presiones sin

apenas gradiente de presión y por tanto bajo la influencia de las circulaciones locales y de mesoescala: iC en verano y iF en invierno.

## **Metodología**

El análisis de los datos se ha efectuado por separado para los 4 períodos meteorológicos típicos del año, conocidas las particularidades de cada uno de ellos en las condiciones de dispersión observadas.

En un principio se han calculado los patrones característicos horarios tanto para las variables meteorológicas: viento en superficie (componentes Vx y Vy), temperatura, radiación, viento a 1500 m de altura (componente Vx y Vy) como de inmisión: NO<sub>2</sub> para días laborables y O<sub>3</sub> (únicamente primavera y verano). Estos patrones son calculados con el método del análisis de conglomerados de K-medias, estableciendo un número 12 para K. Cuando ya se tienen los patrones o categorías para cada una de estas variables se utiliza el método de los árboles de clasificación para establecer las relaciones entre las variables criterio o dependientes (inmisión) y las variables predictoras o independientes (meteorología).

Para el caso de las partículas se ha partido de los valores de concentración media diaria de PM<sub>10</sub>. Las variables meteorológicas utilizadas en este caso han sido las velocidades medias diarias del viento en superficie, el salto térmico diario (diferencia entre la temperatura máxima y mínima), cantidad de lluvia acumulada a lo largo del día y la procedencia de la masa aérea. Tanto para los datos de PM<sub>10</sub> como de velocidad media del viento en superficie y el salto térmico diario se han calculado 7 patrones y la variable lluvia acumulada ha sido categorizada en un número similar de categorías.

Por tanto la metodología utilizada combina técnicas estadísticas como el análisis de conglomerados o CLUSTER utilizando el algoritmo QUICK CLUSTER, también denominado análisis de conglomerados de K-medias, y los árboles de clasificación con la utilización del algoritmo CHAID.

El análisis de conglomerados de K-medias intenta identificar grupos de casos relativamente homogéneos basándose en las características seleccionadas

y utilizando un algoritmo que puede gestionar un gran número de casos. Sin embargo, el algoritmo requiere que el usuario especifique el número de conglomerados.

Las distancias se han calculado utilizando como medida la distancia euclídea simple; el escalamiento de las variables es una consideración importante pero en este caso como se han calculado patrones separados para las diferentes variables no se ha tenido en cuenta.

Se ha utilizado la opción de excluir casos según pareja aplicado a los casos perdidos. Asigna casos a los conglomerados en función de las distancias que se calculan desde todas las variables con valores no perdidos. Las razones para utilizar esta opción son básicamente: reducir el sesgo de las estimaciones, obtener resultados basados en un número suficiente de variables y facilitar procesos posteriores.

Finalmente el método calcula los centros de los conglomerados finales como la media de todos los casos agrupados en cada uno de ellos. Sin embargo, hay que tener en cuenta que los valores que en estos centros se obtienen para las distintas variables dependerán del nº de casos existentes para cada una de ellas dentro del conglomerado.

El procedimiento Árbol de clasificación crea un modelo de clasificación basado en árboles, y clasifica casos en grupos o pronostica valores de una variable (criterio) dependiente basada en valores de variables independientes (predictores). El pronóstico se lleva a cabo asignando a todos los casos del grupo el valor de la categoría más frecuente (con mayor respuesta).

El procedimiento se puede utilizar para:

- ◆ Segmentación. Identifica los casos que pueden ser miembros de un grupo específico.
- ◆ Estratificación. Asigna los casos a una categoría de entre varias.

- ◆ Predicción. Crea reglas y las utiliza para predecir eventos futuros.
- ◆ Reducción de datos y clasificación de variables. Selecciona un subconjunto útil de predictores a partir de un gran conjunto de variables para utilizarlo en la creación de un modelo paramétrico formal.
- ◆ Identificación de interacción. Identifica las relaciones que pertenecen sólo a subgrupos específicos y las especifica en un modelo paramétrico formal.

Entre los métodos de crecimiento del árbol se ha escogido el siguiente:

CHAID. Detección automática de interacciones mediante chi-cuadrado (CHI-square Automatic Interaction Detection). En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente.

Los límites para el crecimiento del árbol vienen determinados por la máxima profundidad y el número de casos mínimo para los nodos parentales y filiales.

Entre los resultados que se obtienen está la Tabla por categoría o patrón. La tabla incluye el porcentaje de ganancia y el porcentaje de respuestas por nodo.

El porcentaje de respuestas es el porcentaje de casos pertenecientes al nodo que pertenecen a la categoría especificada. El porcentaje de ganancia es el porcentaje de los casos totales en la categoría especificada en cada nodo.

Un porcentaje alto de respuesta indica que la categoría o patrón está muy bien caracterizado por los indicadores meteorológicos del nodo, ya que es el porcentaje de casos del nodo que pertenecen al patrón. Un nodo puede presentar una respuesta muy alta para un patrón y sin embargo corresponderle una ganancia baja, ya que la ganancia es el porcentaje de casos de un patrón que pertenecen al nodo.

Que un patrón presente una ganancia alta en un nodo pero que sin embargo no tenga la mayor respuesta en el mismo quiere decir que faltan más indicadores específicos del patrón.

Lo ideal es que tanto la respuesta como la ganancia sean altas.

## **Conclusiones**

En general los resultados obtenidos son satisfactorios.

La clasificación en patrones y su representación gráfica nos da una idea de la situación general en cuanto a los niveles de contaminación registrados, su frecuencia, preferencia estacional y temporal y su distribución espacial en la C.A.P.V.

Los modelos de clasificación han resultado útiles a la hora de seleccionar las variables meteorológicas que influyen en mayor medida en la ocurrencia de valores altos de concentración para cada contaminante.

En la validación de los modelos, por lo general el porcentaje de aciertos ha disminuido ligeramente, como cabría esperar, pero se mantienen las relaciones encontradas y la clasificación de los casos en patrones se ajusta a la frecuencia de los patrones observada en los años precedentes, aunque en ocasiones una mayor aparición ha estado favorecida por unas determinadas condiciones meteorológicas ocurridas con mayor frecuencia en unos años que en otros, y otras veces las causas que provocan los patrones observados han tenido una incidencia limitada a un período de tiempo.

Los modelos de clasificación han resultado ser mejores para la variable  $PM_{10}$ , con un error en la estimación menor en invierno y otoño cuando, por otro lado, los niveles son mayores y/o los niveles altos son más frecuentes.

Para  $PM_{10}$  los modelos se quedan por lo general en un primer nivel dependiendo de la variable contraste térmico, salvo en verano que la variable más influyente es la procedencia de la masa aérea. Además en otoño también entra en un segundo nivel la velocidad de viento media diaria. Contrastes térmicos elevados, procedencia africana de la masa de aire y velocidades de viento bajas son las condiciones más adversas para los patrones de esas variables.

Los modelos para  $O_3$  y  $NO_2$  son más complejos y aunque en algunos casos son muy certeros en la definición de las condiciones meteorológicas con las que se producen valores altos de contaminación, en otros casos no asocian los valores altos medidos con unas únicas condiciones meteorológicas.

El modelo para  $O_3$  en la época de primavera y para los patrones más adversos es más complejo (mayor profundidad) que en verano. Así como en verano la variable más influyente es la temperatura en primavera la clasificación depende de la temperatura, la radiación y puede llegar a incluir el viento en altura. Valores de temperatura altos es una de las causas principales de concentraciones de ozono elevadas.

En un primer nivel del árbol para los modelos de  $NO_2$  el viento en superficie es la variable con mayor influencia y después la variable temperatura es la que sigue dividiendo el árbol. Los patrones cuyas velocidades de viento son más bajas son los que están relacionados con los niveles de concentración mayores. En primavera los modelos presentan una mayor profundidad y son más las variables meteorológicas implicadas.

La decisión de cual es el nivel del árbol suficiente para identificar relaciones está en mantener un equilibrio entre el nivel de complejidad del modelo, el porcentaje de acierto, el porcentaje de equivocación y las consecuencias de esta equivocación.

De los resultados obtenidos en este primer ejercicio de relacionar objetivamente las variables meteorológicas con los niveles de concentración registrados se desprenden también algunas conclusiones sobre la metodología empleada. En el caso de  $PM_{10}$  y  $NO_2$ , los patrones pueden variar más a lo largo de los años e incluso algunos de ellos están más relacionados con causas puntuales en el tiempo. Los patrones se deben ir actualizando con el tiempo y además es conveniente no agrupar muchos años, tal vez 2 ó 3 años serían suficientes. En el caso de  $O_3$  los patrones se mantienen más invariables, aunque

con frecuencias diferentes de un año a otro, en este caso 4 ó 5 años sería el mínimo a agrupar.

El método empleado para clasificar y establecer las relaciones causa-efecto implica una segregación del período anual en las 4 estaciones meteorológicas conocidas. Este hecho favorece mucho la operatividad del método, ya que no hay que esperar todo un año para actualizar y utilizar las reglas resultantes en la gestión diaria de la calidad del aire. Así, al comienzo de cada trimestre sería necesario actualizar los patrones tanto meteorológicos como de contaminación con los datos del trimestre anterior y volver a calcular el modelo de árbol.

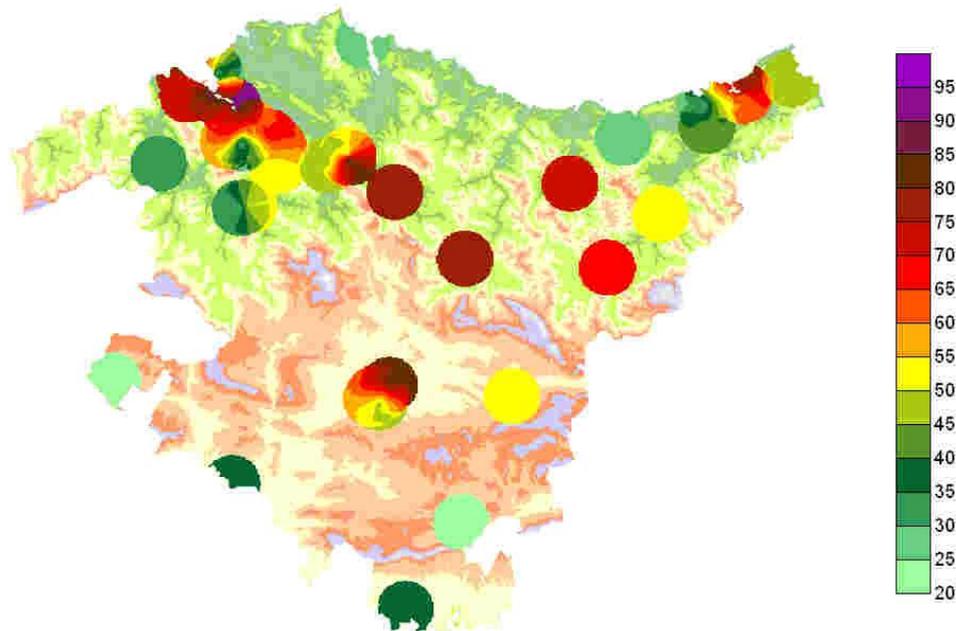
La estrategia de utilización se basaría en la clasificación diaria de los patrones de las variables meteorológicas y de contaminación registrados en las horas previas, estudiar sus características, y en base a los valores previstos (fundamentalmente viento en superficie, temperatura y radiación) decidir con las reglas disponibles, resultantes del modelo de árbol actualizado hasta el año anterior, si en el día en curso y posteriores se puede llegar a alcanzar niveles de contaminación altos, centrando los esfuerzos de vigilancia y control en las medidas de PM<sub>10</sub> y NO<sub>2</sub> en invierno y otoño, y en las medidas de O<sub>3</sub> en primavera y verano.

Por tanto el método desarrollado puede ser eficaz a la hora de efectuar un control secundario de las emisiones y la prevención en el cumplimiento de los límites establecidos.

## Ejemplos

### PM<sub>10</sub> invierno:

El patrón 6 (Figura 2) es uno de los más adversos en esta época del año con una frecuencia de un 10 % de los días en este período.



**Figura 2.** Campo de concentraciones ( $\mu\text{g}/\text{m}^3$ ) calculado por interpolación (método Kriging con radio de búsqueda de 5 Km) para el patrón 6 de PM<sub>10</sub> (invierno).

El nodo 4 es el único nodo terminal en que se pronostica el patrón 6. En el árbol resultante (Figura 3) se aprecia que en el nodo 4 se encuentran todos los días con un salto térmico correspondiente al patrón 1.

El patrón 1 para el salto térmico diario se caracteriza por unos valores del salto térmico cercanos a los 18°C en algunos puntos del interior. En general los valores más bajos se dan en la zona costera y los más altos en el interior de los valles costeros.

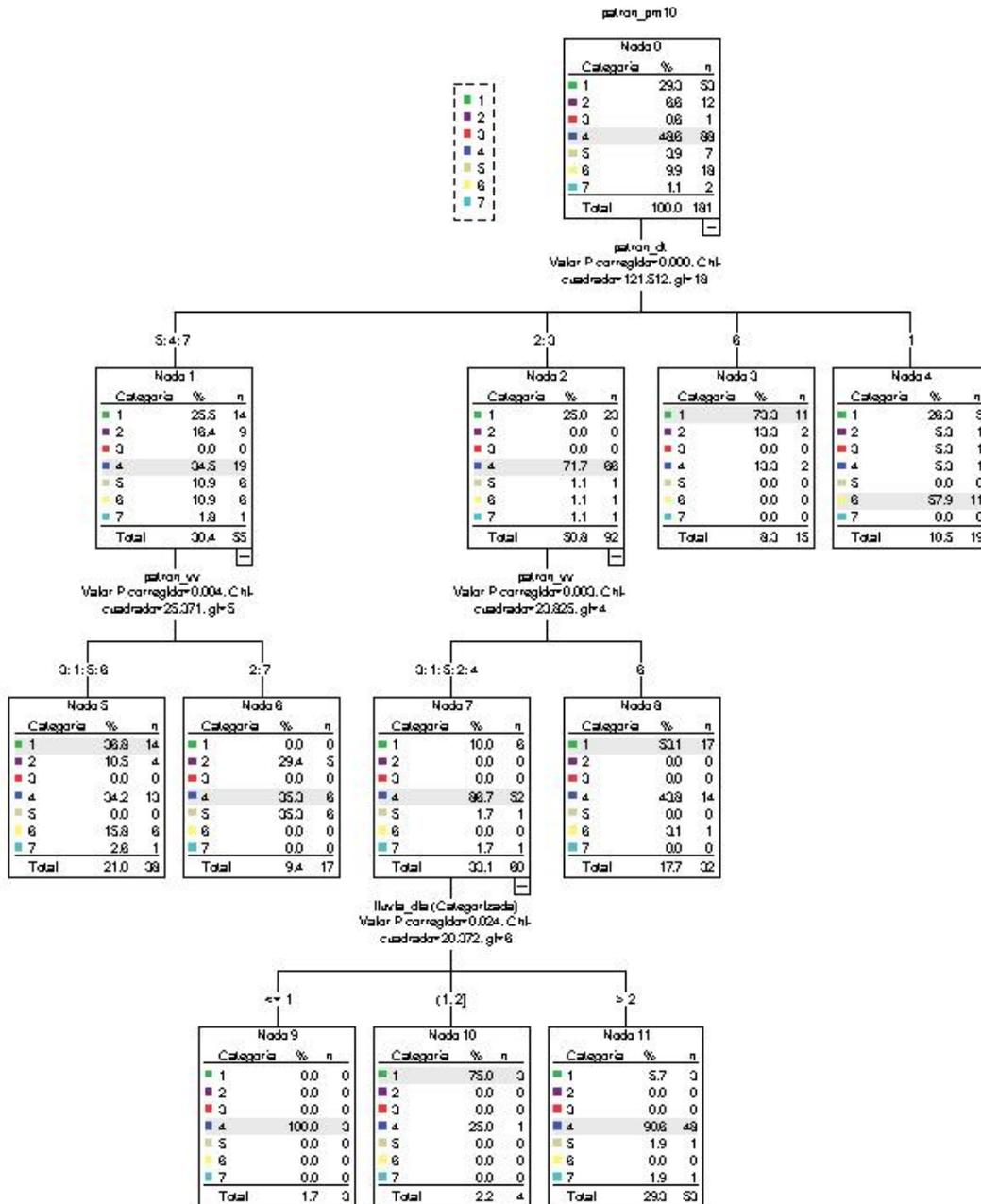


Figura 3. Arbol de clasificación de PM<sub>10</sub> (invierno).

El patrón 6 que en los años 2004 y 2005 presenta la mayor respuesta del nodo 4 del árbol, aumenta dicha respuesta en el 2006 a un 100% pero disminuye su ganancia, aunque el número de casos es muy bajo para que sea muy significativo (Tabla 1).

Muestra	Nodo	Nodo		Ganancia		Respuesta
		N	Porcentaje	N	Porcentaje	
Training	4	19	10.5%	11	61.1%	57.9%
	5	38	21.0%	6	33.3%	15.8%
	8	32	17.7%	1	5.6%	3.1%
	11	53	29.3%	0	.0%	.0%
	6	17	9.4%	0	.0%	.0%
	3	15	8.3%	0	.0%	.0%
	10	4	2.2%	0	.0%	.0%
	9	3	1.7%	0	.0%	.0%
	Test	4	4	4.4%	4	36.4%
5		26	28.9%	4	36.4%	15.4%
8		19	21.1%	2	18.2%	10.5%
11		19	21.1%	0	.0%	.0%
6		10	11.1%	0	.0%	.0%
3		4	4.4%	1	9.1%	25.0%
10		0	.0%	0	.0%	.
9		8	8.9%	0	.0%	.0%

**Tabla 1.** Ganancias del patrón 6 de PM<sub>10</sub> (invierno) en los nodos terminales. Muestra de entrenamiento y de comprobación.

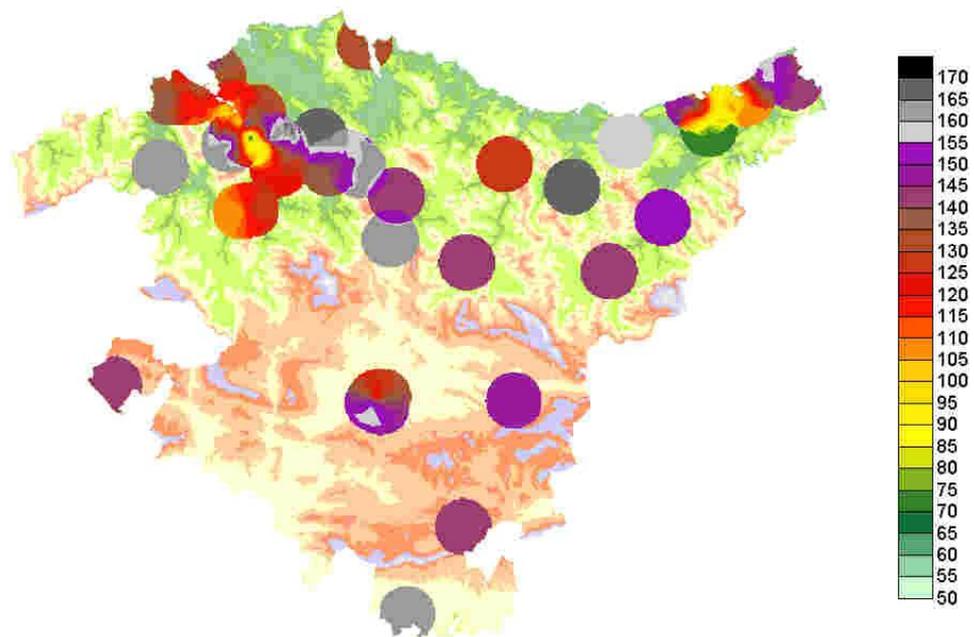
**O<sub>3</sub> verano**

En la clasificación de patrones efectuada para O<sub>3</sub> en la época de verano destacan los patrones 2 (Figura 4) y 5 (Figura 5) como más adversos y con una aparición destacada el año 2003 (Tabla 2).

	Patrones												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
año 2002	397	0	44	176	22	119	205	75	313	410	348	99	2208
2003	211	63	105	249	187	194	171	116	227	233	291	161	2208
2004	373	0	103	270	39	69	143	26	362	383	353	87	2208
2005	329	17	48	206	75	105	189	69	337	358	326	149	2208
Total	1310	80	300	901	323	487	708	286	1239	1384	1318	496	8832

**Tabla 2.** Frecuencias anuales de los patrones de O<sub>3</sub> (verano).

Ambos patrones son típicos del período diurno y con una mayor frecuencia de aparición entre las 12 y las 18 h GMT (Tabla 3).



**Figura 4.** Patrón 2 de O<sub>3</sub> (verano). Valores medios en µg/m<sup>3</sup>.

	Patrones												Total
	1	2	3	4	5	6	7	8	9	10	11	12	
hora 1	72	0	0	14	3	44	55	19	52	91	16	2	368
2	87	0	0	9	2	39	58	17	50	88	15	3	368
3	102	0	0	10	2	34	60	13	39	91	15	2	368
4	121	0	0	6	1	33	69	11	35	75	15	2	368
5	153	0	0	5	0	19	75	6	37	64	7	2	368
6	186	0	0	0	0	18	73	6	21	57	6	1	368
7	195	0	0	0	0	17	67	4	18	57	9	1	368
8	152	0	0	4	0	29	47	3	34	87	12	0	368
9	64	0	4	10	0	23	25	14	68	124	34	2	368
10	17	0	13	21	1	14	9	5	108	84	79	17	368
11	5	0	24	50	13	1	0	4	106	33	111	21	368
12	1	5	23	100	26	0	0	2	54	16	108	33	368
13	0	11	31	105	34	0	0	1	29	11	95	51	368
14	0	16	35	89	46	0	0	1	20	9	93	59	368
15	0	14	32	84	47	0	0	2	22	8	92	67	368
16	0	13	35	80	45	1	0	2	34	7	89	62	368
17	0	10	41	76	38	1	0	2	38	10	100	52	368
18	0	7	32	83	27	2	0	7	58	16	97	39	368
19	1	3	13	55	17	11	2	27	64	38	104	33	368
20	8	1	9	35	7	31	15	21	83	49	84	25	368
21	16	0	6	26	5	37	29	30	85	71	55	8	368
22	29	0	1	17	3	41	32	38	71	97	33	6	368
23	45	0	0	11	3	44	41	27	60	104	28	5	368
24	56	0	1	11	3	48	51	24	53	97	21	3	368
Total	1310	80	300	901	323	487	708	286	1239	1384	1318	496	8832

**Tabla 3.** Frecuencias horarias de los patrones de O<sub>3</sub> (verano).

Si nos limitamos al período del día en que es más frecuente la aparición de estos dos patrones adversos, el árbol de clasificación resultante únicamente los relaciona con los valores del patrón de temperatura en superficie (Figura 6).

El nodo 71 es el único que pronostica correctamente el patrón 2, aunque en el nodo 69 es donde presenta una mayor ganancia (Tabla 4).

En la muestra de comprobación únicamente se han dado 9 casos con un patrón 2 de O<sub>3</sub> en el año 2006 y, siendo la respuesta en el nodo 71 algo inferior, la ganancia ha aumentado considerablemente.

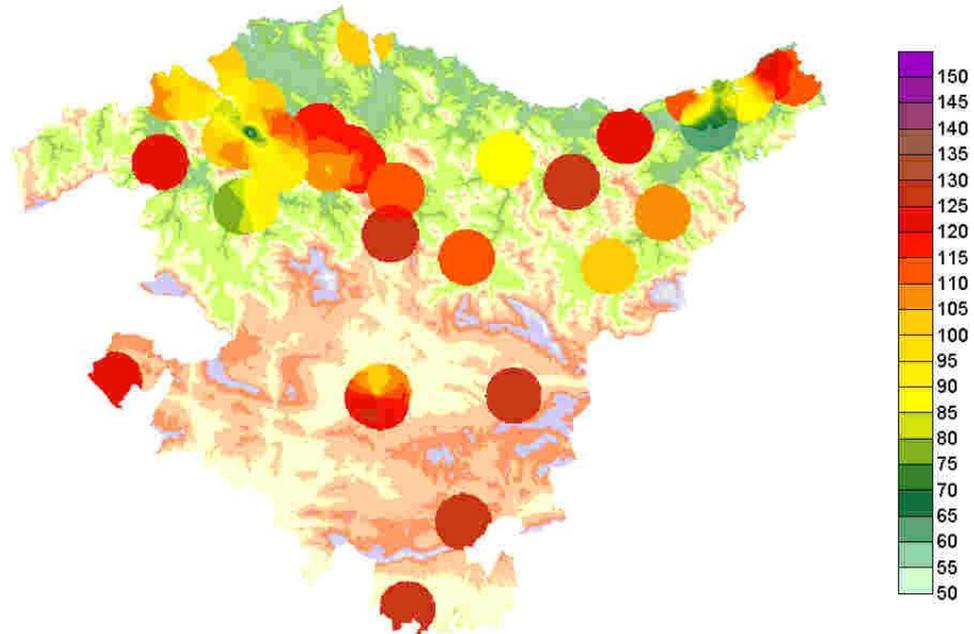
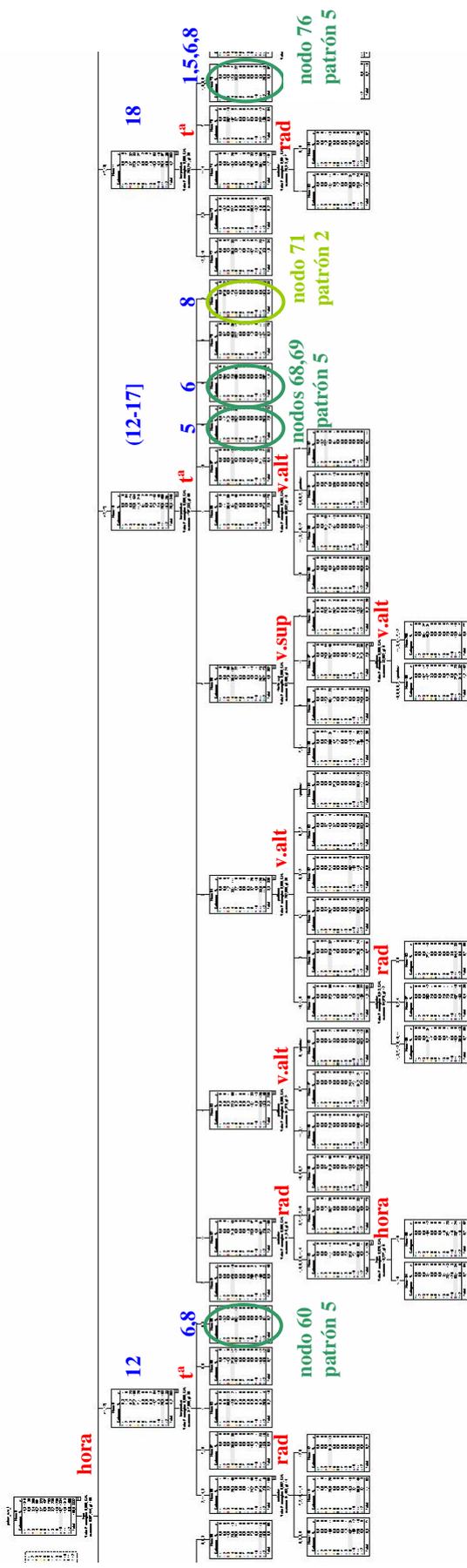


Figura 5. Patrón 5 de O<sub>3</sub> (verano). Valores medios en µg/m<sup>3</sup>.

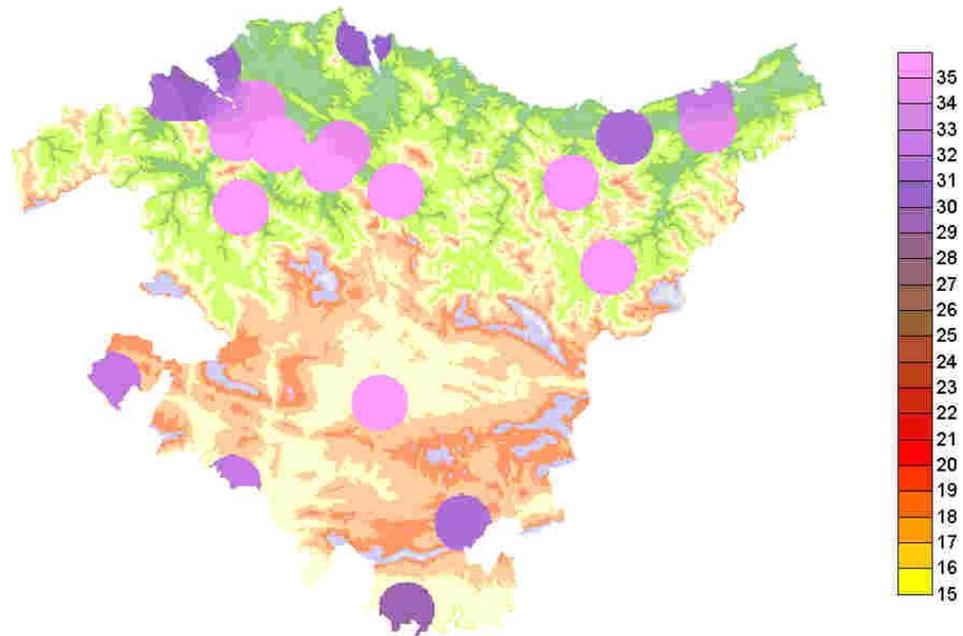
Muestra	Nodo	Nodo		Ganancia		Respuesta
		N	Porcentaje	N	Porcentaje	
Training	71	38	.4%	22	27.5%	57.9%
	69	121	1.4%	32	40.0%	26.4%
	60	22	.2%	4	5.0%	18.2%
	73	22	.2%	4	5.0%	18.2%
	162	11	.1%	2	2.5%	18.2%
	70	75	.8%	5	6.3%	6.7%
	80	36	.4%	2	2.5%	5.6%
	76	57	.6%	3	3.8%	5.3%
	85	20	.2%	1	1.3%	5.0%
	79	48	.5%	1	1.3%	2.1%
	68	178	2.0%	3	3.8%	1.7%
	59	63	.7%	1	1.3%	1.6%
Test	71	14	.6%	6	66.7%	42.9%
	69	57	2.6%	3	33.3%	5.3%

Tabla 4. Ganancias del patrón 2 de O<sub>3</sub> (verano) en los nodos terminales. Muestra de entrenamiento y de comprobación.



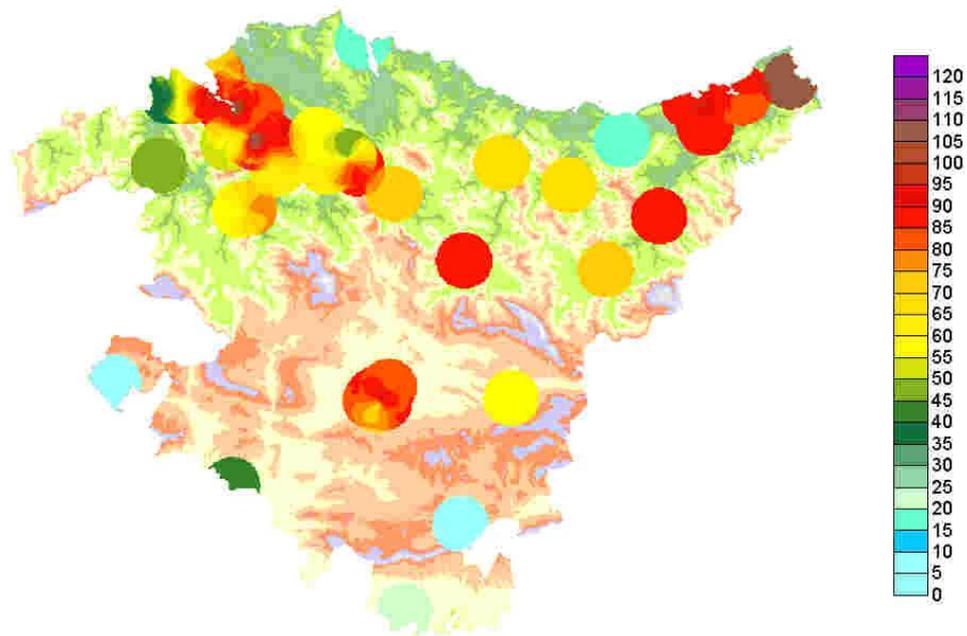
**Figura 6.** Arbol de clasificación de  $O_3$  (verano). Zona del árbol, entre las 12 y las 18 h, en rojo las variables de segmentación, en azul sus valores y en verde los nodos en los que se pronostican correctamente los patrones 2 y 5 de ozono.

En los nodos 71 y 69 están todos los casos con un patrón de temperatura 8 (Figura 7) y 6 respectivamente, en el período entre las 13 y las 17 h.



**Figura 7.** Patrón 8 de  $T^a$  (verano). Valores medios en °C





**Figura 9.** Patrón 12 de NO<sub>2</sub> laborables (invierno). Valores medios en µg/m<sup>3</sup>.

El patrón 10 caracterizado sobre todo por unos valores muy altos en las estaciones de Irún y Rentería ha sido un patrón que se asocia claramente a una situación frecuente en el año 2002 (Tabla 6) pero que prácticamente desaparece a partir de ahí (ningún caso en el año 2006). Los casos se agrupan sobre todo entre las 8 y las 10 de la mañana.

El patrón 9 destaca sobre todo por unos valores más altos en las estaciones de Vitoria respecto a lo registrado en el resto de la Red. La distribución de los casos en los diferentes años es más homogénea que en el patrón 10 y, como en él, es mas frecuente a primeras horas de la mañana.

En la construcción de los árboles se han separado 2 períodos del día. Por un lado el período entre las 8 y las 10 h GMT (Figura 10) y por otro el período entre las 18 y las 22 h GMT (Figura 11).

		Patrones												Total
		1	2	3	4	5	6	7	8	9	10	11	12	
hora	1	76	0	47	75	4	34	0	3	1	0	4	0	244
	2	90	0	28	87	3	31	0	0	1	0	4	0	244
	3	97	0	19	91	1	33	0	0	1	0	2	0	244
	4	99	0	20	85	0	37	0	1	1	0	1	0	244
	5	65	0	24	94	0	56	1	1	1	0	2	0	244
	6	35	0	60	82	5	43	1	7	2	0	9	0	244
	7	10	8	73	18	34	20	18	30	9	2	22	0	244
	8	3	19	23	8	55	15	22	24	30	22	13	10	244
	9	3	16	15	6	47	17	23	29	26	27	20	15	244
	10	6	15	11	12	46	21	25	32	18	22	26	10	244
	11	8	9	23	11	24	32	23	39	18	11	42	4	244
	12	6	6	15	22	15	56	25	35	10	2	51	1	244
	13	6	3	11	11	5	87	21	20	9	0	71	0	244
	14	6	6	6	17	2	94	15	11	7	0	80	0	244
	15	5	6	9	22	2	81	19	15	5	0	79	1	244
	16	1	11	9	24	5	50	33	20	4	1	84	2	244
	17	0	23	10	11	12	29	53	22	3	3	73	5	244
	18	0	32	8	4	28	20	50	26	3	4	55	14	244
	19	0	33	8	3	40	19	38	30	11	6	36	20	244
	20	1	19	12	2	52	25	27	30	12	13	29	22	244
	21	7	8	18	6	58	28	18	32	9	9	29	22	244
	22	13	6	37	9	48	41	11	24	13	5	25	12	244
	23	37	1	58	23	37	44	4	17	7	1	10	5	244
	24	54	1	59	44	22	39	3	10	3	0	8	1	244
Total		628	222	603	767	545	952	430	458	204	128	775	144	5856

**Tabla 5.** Frecuencias horarias de los patrones de NO<sub>2</sub> laborables (invierno).

		Patrones												Total
		1	2	3	4	5	6	7	8	9	10	11	12	
año	2002	163	96	102	172	92	201	138	122	21	106	204	23	1440
	2003	214	38	147	192	134	207	127	83	55	7	247	37	1488
	2004	129	46	165	185	158	271	120	111	84	10	186	23	1488
	2005	122	42	189	218	161	273	45	142	44	5	138	61	1440
Total		628	222	603	767	545	952	430	458	204	128	775	144	5856

**Tabla 6.** Frecuencias anuales de los patrones de NO<sub>2</sub> laborables (invierno).



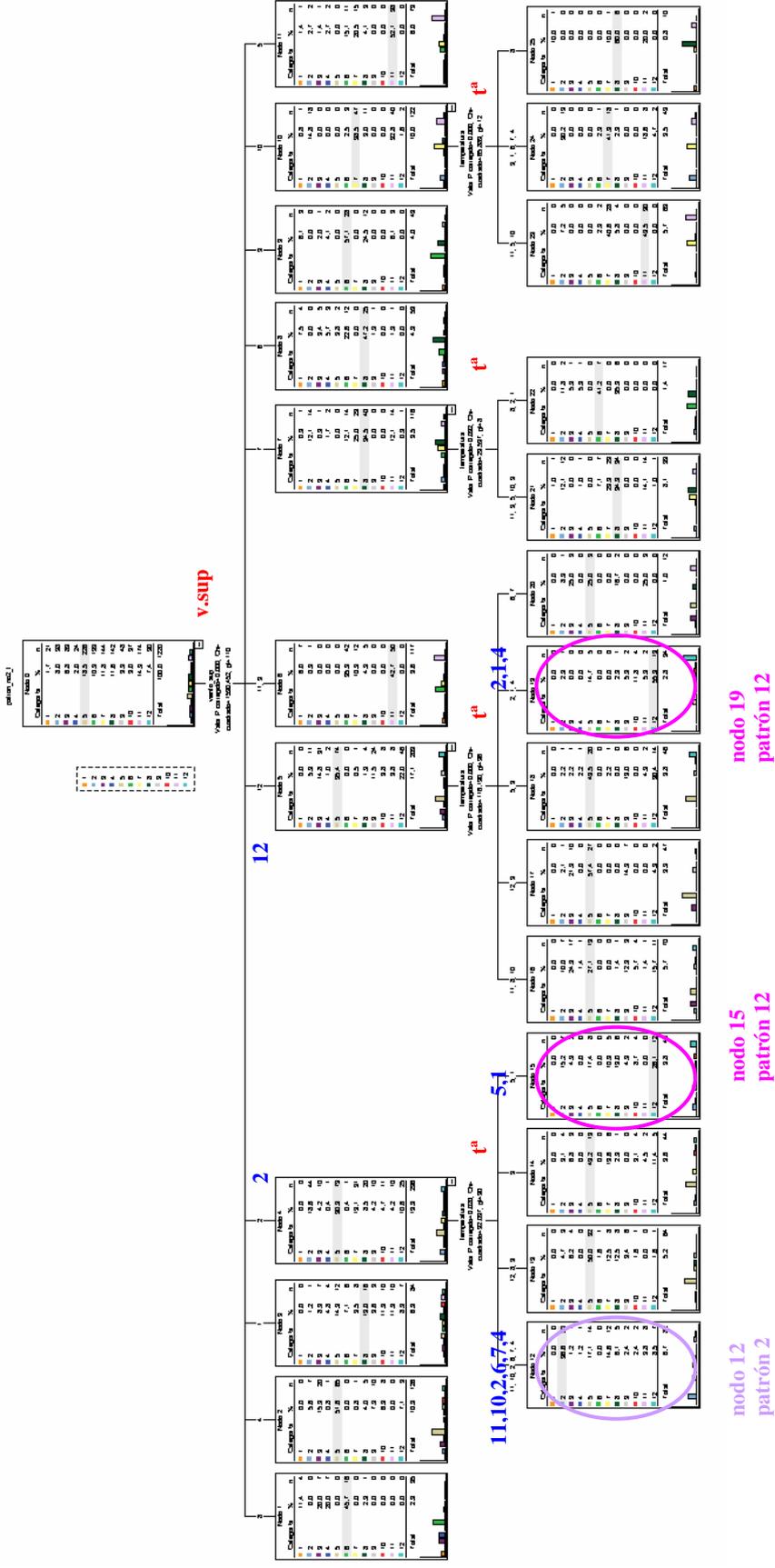


Figura 11. Arbol de clasificación de NO<sub>2</sub> (invierno). Período tarde, en azul las variables de segmentación, en rojo las variables de segmentación, en azul sus valores y en morado claro y rosa los nodos en los que se pronostican correctamente los patrones 2 y 12 respectivamente.

El nodo 16 es el único que pronostica correctamente el patrón 9 de NO<sub>2</sub> para el período de la mañana seleccionado. Valores parecidos en la muestra de comprobación (Tabla 7).

Muestra	Nodo	Nodo		Ganancia		Respuesta
		N	Porcentaje	N	Porcentaje	
Training	16	57	7.8%	20	27.0%	35.1%
	12	51	7.0%	16	21.6%	31.4%
	15	38	5.2%	9	12.2%	23.7%
	14	18	2.5%	3	4.1%	16.7%
	6	97	13.3%	14	18.9%	14.4%
	2	90	12.3%	7	9.5%	7.8%
	11	72	9.8%	5	6.8%	6.9%
	7	86	11.7%	0	.0%	.0%
	10	65	8.9%	0	.0%	.0%
	9	46	6.3%	0	.0%	.0%
	8	39	5.3%	0	.0%	.0%
	5	38	5.2%	0	.0%	.0%
	1	20	2.7%	0	.0%	.0%
	13	15	2.0%	0	.0%	.0%
Test	16	19	9.9%	6	21.4%	31.6%
	12	22	11.5%	5	17.9%	22.7%
	15	8	4.2%	6	21.4%	75.0%
	14	9	4.7%	2	7.1%	22.2%
	6	33	17.2%	4	14.3%	12.1%
	2	12	6.3%	1	3.6%	8.3%
	11	19	9.9%	3	10.7%	15.8%
	7	20	10.4%	0	.0%	.0%
	10	12	6.3%	0	.0%	.0%
	9	14	7.3%	0	.0%	.0%
	8	2	1.0%	0	.0%	.0%
	5	21	10.9%	1	3.6%	4.8%
	1	0	.0%	0	.0%	.
	13	1	.5%	0	.0%	.0%

**Tabla 7.** Ganancias del patrón 9 de NO<sub>2</sub> (invierno mañana) en los nodos terminales. Muestra de entrenamiento y de comprobación.

Los nodos 19 y 15 pronostican correctamente el patrón 12. En la muestra de comprobación (Tabla 8) destaca que el patrón 12 ha disminuido su frecuencia de aparición pero debido fundamentalmente a que no se han dado casos con los valores de los predictores del nodo 19.

Muestra	Nodo	Nodo		Ganancia		Respuesta
		N	Porcentaje	N	Porcentaje	
Training	19	34	2.8%	19	21.1%	55.9%
	18	46	3.8%	14	15.6%	30.4%
	15	46	3.8%	12	13.3%	26.1%
	16	70	5.7%	11	12.2%	15.7%
	14	44	3.6%	5	5.6%	11.4%
	12	82	6.7%	7	7.8%	8.5%
	3	84	6.9%	7	7.8%	8.3%
	2	126	10.3%	9	10.0%	7.1%
	24	43	3.5%	2	2.2%	4.7%
	17	47	3.9%	2	2.2%	4.3%
Test	19	0	.0%	0	.0%	.
	18	7	2.2%	2	9.5%	28.6%
	15	24	7.5%	6	28.6%	25.0%
	16	14	4.4%	1	4.8%	7.1%
	14	17	5.3%	5	23.8%	29.4%
	12	38	11.9%	3	14.3%	7.9%
	3	20	6.3%	1	4.8%	5.0%
	2	23	7.2%	1	4.8%	4.3%
	24	14	4.4%	0	.0%	.0%
	17	11	3.4%	0	.0%	.0%
13	24	7.5%	2	9.5%	8.3%	
21	29	9.1%	0	.0%	.0%	

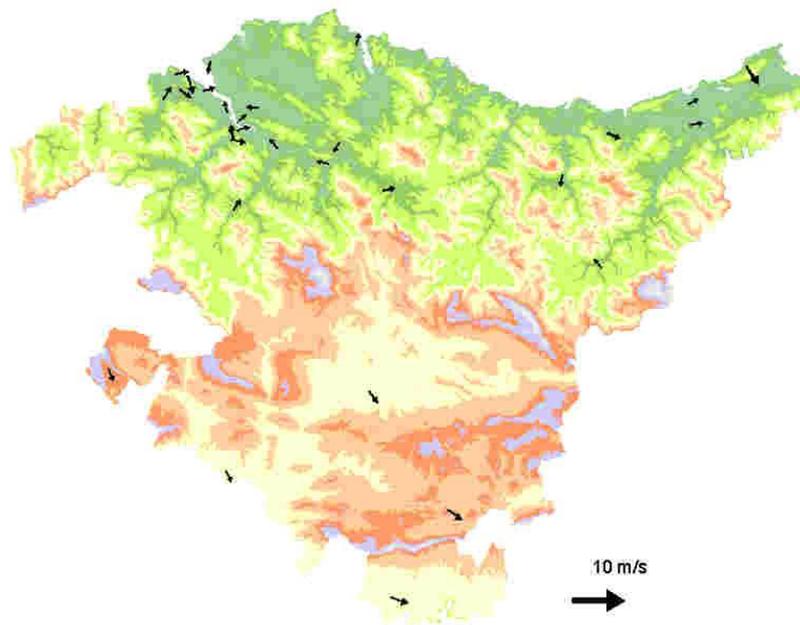
**Tabla 8.** Ganancias del patrón 12 de NO<sub>2</sub> (invierno tarde) en los nodos terminales. Muestra de entrenamiento y de comprobación.

El patrón 2 ha aumentado ligeramente la ganancia y la respuesta en el nodo 12 para la muestra de comprobación (Tabla 9).

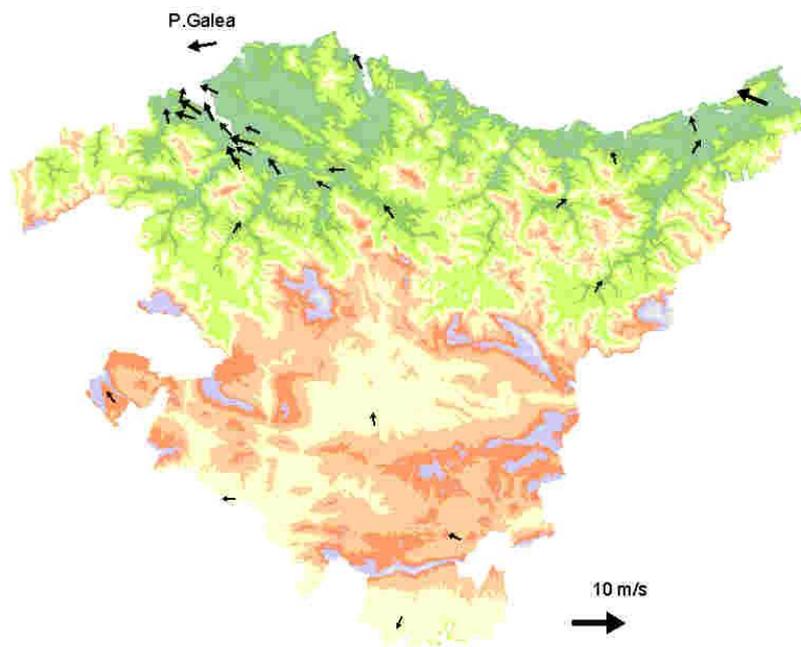
Muestra	Nodo	Nodo		Ganancia		Respuesta	
		N	Porcentaje	N	Porcentaje		
Training	12	82	6.7%	30	30.6%	36.6%	
	24	43	3.5%	13	13.3%	30.2%	
	15	46	3.8%	7	7.1%	15.2%	
	21	99	8.1%	12	12.2%	12.1%	
	22	17	1.4%	2	2.0%	11.8%	
	16	70	5.7%	7	7.1%	10.0%	
	14	44	3.6%	4	4.1%	9.1%	
	20	12	1.0%	1	1.0%	8.3%	
	23	69	5.7%	5	5.1%	7.2%	
	2	126	10.3%	7	7.1%	5.6%	
	13	64	5.2%	3	3.1%	4.7%	
	19	34	2.8%	1	1.0%	2.9%	
	11	73	6.0%	2	2.0%	2.7%	
	18	46	3.8%	1	1.0%	2.2%	
	17	47	3.9%	1	1.0%	2.1%	
	3	84	6.9%	1	1.0%	1.2%	
	6	117	9.6%	1	1.0%	.9%	
	Test	12	38	11.9%	16	40.0%	42.1%
		24	14	4.4%	3	7.5%	21.4%
15		24	7.5%	6	15.0%	25.0%	
21		29	9.1%	2	5.0%	6.9%	
22		3	.9%	0	.0%	.0%	
16		14	4.4%	1	2.5%	7.1%	
14		17	5.3%	3	7.5%	17.6%	
20		2	.6%	0	.0%	.0%	
23		17	5.3%	1	2.5%	5.9%	
2		23	7.2%	1	2.5%	4.3%	
13		24	7.5%	5	12.5%	20.8%	
19		0	.0%	0	.0%	.	
11		10	3.1%	1	2.5%	10.0%	
18		7	2.2%	0	.0%	.0%	
17		11	3.4%	0	.0%	.0%	
3		20	6.3%	1	2.5%	5.0%	
6		23	7.2%	0	.0%	.0%	

**Tabla 9.** Ganancias del patrón 2 de NO<sub>2</sub> (invierno tarde) en los nodos terminales. Muestra de entrenamiento y de comprobación.

La variable viento en superficie es la más influyente en los niveles de  $\text{NO}_2$  registrados. Los patrones de viento en superficie más adversos son el 2 y el 12 (Figuras 12 y 13), asociados a situaciones de calma y desacoplamiento de la capa atmosférica más baja de las condiciones de viento en altura (como el patrón 11). En un segundo término se encuentra para el período de últimas horas de la tarde la temperatura, los patrones con valores más altos son los que se relacionan con los patrones de  $\text{NO}_2$  más adversos.



**Figura 12.** Patrón 2 de viento en superficie (invierno).



**Figura 13.** Patrón 12 de viento en superficie y patrón 11 de viento a 1500 m Punta Galea (invierno).