



**Programa Oficial de Postgrado**  
**Ciencias, Tecnología y Computación**  
**Máster: Matemáticas y Computación**

---

Tesis de Master  
**Calibración Local de Predicciones Numéricas de  
Viento con Técnicas Estadísticas no Lineales  
(Downscaling Estadístico)**

Director: José Manuel Gutiérrez Llorente  
Departamento: Instituto de Física de Cantabria (CSIC-UC)  
Presentado: Jesús García Abad

Santander 8 de Octubre de 2012



1. Objetivos y Estructuración de la Tesis de Máster.
2. Explicación breve de los campos de vientos.
3. Modelos numéricos de circulación atmosférica.
4. Variabilidad de los campos de vientos.
5. Descripción del problema de la predicción del viento.

*Análisis de los datos existentes.*

*Correlación del viento con la presión y la temperatura.*

*Tendencia de los campos de vientos para la rejilla observada.*

6. Simulación con Técnicas de **Downscaling Estadístico**

*Introducción a las técnicas de **Downscaling Estadístico**.*

*Datos de observación.*

*Uso de la Toolbox Meteolab.*

*Técnicas para series temporales.*

*Correlación canónica.*

*Técnica de locales análogos.*

*Curvas ROC.*

*Simulación de las tendencias de viento a tres periodos distintos*

*Validez del método.*

7. Conclusiones
8. Bibliografía

## 1. Objetivos y estructuración de la tesis de Máster.

En esta tesis de Máster, se ha intentado predecir la variabilidad del viento en determinadas zonas de la península ibérica, a partir de datos registrados por el **Centro Europeo de Predicción a Plazo Medio** (ECMWF), por sus siglas en Inglés. Éste, se creó en 1975 como un consorcio de países europeos que acordaron unir sus esfuerzos para crear un centro especializado en la predicción a medio plazo (entre 2 y 10 días). Las primeras predicciones operativas se produjeron en 1979. Es un centro dedicado a la investigación y operaciones y su modelo está considerado como el mejor del mundo, por encima de todos los modelos de predicción operativos de los Estados Unidos.

Parece clarificador que las fuentes de información, no podrían ser mejores, con lo que gran parte de nuestro trabajo está bastante bien ensamblado.

Por otro lado, para predecir la variabilidad del viento, vamos a aplicar técnicas de cálculo estadístico, con el fin de encontrar una respuesta a la pregunta que se ha venido preguntando el hombre desde tiempo inmemorial, acerca de la predicción de las variables atmosféricas.

Para ello la técnica fundamental que vamos a usar es el ***Downscaling Estadístico***, una suerte de técnicas y procedimientos de carácter probabilístico, que nos permitan determinar la variabilidad de una función, en base a valores conocidos de esa variable, en épocas pasadas, y hacer de este modo una extrapolación hacia valores futuros.

Finalmente, y como en cualquier técnica de carácter probabilístico, trataremos de analizar, la correlación de los resultados obtenidos, y la fiabilidad de los mismos en un intervalo de confianza determinado, de gran utilidad, serán las curvas ROC.

Para todo lo anterior, haremos uso de una potente herramienta Matemática, que no es otra que la Toolbox Meteolab, creada y desarrollada en el IFCA por José Manuel Gutiérrez y su equipo investigador de la Universidad de Cantabria <http://www.meteo.unican.es>.

## 2. Explicación de los campos de vientos.

### 2.1 Introducción a la variable viento.

El **viento** es el flujo de gases a gran escala. En la Tierra, el viento es el movimiento en masa del aire en la atmósfera. Günter D. Roth lo define como “**la compensación de las diferencias de presión atmosférica entre dos puntos**”. Es decir cuando entre dos puntos existe una diferencia de presión se produce un movimiento en la masa de aire que se encuentra entre los dos puntos.

En meteorología se suelen denominar los vientos según su fuerza y la dirección desde la que soplan. Los aumentos repentinos de la velocidad del viento durante un tiempo corto reciben el nombre de *ráfagas*. Los vientos fuertes de duración intermedia (aproximadamente un minuto) se llaman *turbonadas*. Los vientos de larga duración tienen diversos nombres según su fuerza media como, por ejemplo, *brisa*, *temporal*, *tormenta*, *huracán* o *tifón*. El viento se puede producir en diversas escalas: desde flujos tormentosos que duran decenas de minutos hasta brisas locales generadas por el distinto calentamiento de la superficie terrestre y que duran varias horas, e incluso globales, que son el fruto de la diferencia de absorción de energía solar entre las distintas zonas de la Tierra.

### 2.2 El viento y la circulación a gran escala.

Las dos causas principales de la circulación atmosférica a gran escala son el calentamiento diferencial de la superficie terrestre según la latitud, y la inercia y fuerza centrífuga producidas por la rotación del planeta. En los trópicos, la circulación de depresiones térmicas por encima del terreno y de las mesetas elevadas puede impulsar la circulación de monzones. En las áreas costeras, el ciclo **brisa marina-brisa terrestre** puede definir los vientos locales, mientras que en las zonas con relieve variado las **brisas de valle y montaña** pueden dominar los vientos locales. Para nuestro estudio, es muy importante, este tipo de vientos, ya que son una etiqueta identificativa de las más importantes en el clima de nuestra tierra.

### 2.3 El viento en la vida diaria y sus aplicaciones.

En la civilización humana, el viento ha inspirado la mitología, ha afectado a los acontecimientos históricos, ha extendido el alcance del transporte y la guerra, y ha proporcionado una fuente de energía para el trabajo mecánico, la electricidad y el ocio. El viento ha impulsado los viajes de los veleros a través de los océanos de la Tierra. Los globos aerostáticos utilizan el viento para viajes cortos, y el vuelo con motor lo utilizan para generar sustentación y reducir el consumo de combustible. Las zonas con cizalladura del viento provocado por varios fenómenos meteorológicos pueden provocar situaciones peligrosas para las aeronaves. Cuando los vientos son fuertes, los árboles y las estructuras creadas por los seres humanos pueden llegar a resultar dañados o destruidos.

Los vientos pueden dar forma al relieve a través de una serie de procesos eólicos como la formación de suelos fértiles (por ejemplo, el loess) o la erosión. El polvo de desiertos grandes puede ser movido a grandes distancias desde su lugar de origen por los vientos dominantes, y

los vientos que son acelerados por una topografía agreste y que están asociados con tormentas de polvo han recibido nombres regionales en diferentes partes del mundo debido a su efecto significativo sobre estas regiones. El viento afecta la extensión de los incendios forestales. También dispersa las semillas de determinadas plantas, y hace posible la supervivencia y dispersión de estas especies vegetales, así como las poblaciones de insectos voladores. En combinación con las temperaturas frías, el viento tiene un efecto negativo sobre el ganado. El viento afecta las reservas de alimento de los animales y sus estrategias de caza y defensa.

#### **2.4 Causas generales del viento**

La gran capa atmosférica es atravesada por las radiaciones solares que calientan el suelo, el cual, a su vez, calienta el aire que lo rodea. Así resulta que éste no es calentado directamente por los rayos solares que lo atraviesan sino, en forma indirecta, por el calentamiento del suelo y de las superficies acuáticas. Cuando el aire se calienta, también se dilata, como cualquier gas, es decir, aumenta de volumen, por lo cual asciende hasta que su temperatura se iguala con la del aire circundante.

A grandes rasgos, las masas de aire van de los trópicos al ecuador (vientos alisios, que son constantes, es decir, que soplan durante todo el año), donde logran ascender tanto por su calentamiento al disminuir la latitud (en la zona intertropical) como por la fuerza centrífuga del propio movimiento de rotación terrestre, que da origen a su vez a que el espesor de la atmósfera en la zona ecuatorial sea el mayor en toda la superficie terrestre.

Al ascender, se enfrían, y por las altas capas vuelven hacia los trópicos, donde descienden por su mayor peso (aire frío y seco) lo cual explica la presencia de los desiertos subtropicales y la amplitud térmica diaria tan elevada de los desiertos (en el Sáhara es frecuente que temperaturas de casi 50º durante el día, por la insolación y la falta de nubes, se vea contrastada con temperaturas muy bajas durante la noche.

Así, en estas zonas desérticas, las temperaturas varían muchísimo del día a la noche por la escasa cantidad de agua y vapor de agua, que contribuirían a una mayor regularidad térmica.

La primera descripción científica conocida del viento se debe a Torricelli : *...los vientos son producidos por diferencias en la temperatura del aire, y por tanto de la densidad, entre dos regiones de la Tierra.*

Otras fuerzas que mueven el viento o lo afectan son la fuerza del gradiente de presión, el efecto Coriolis, las fuerzas de flotabilidad y de fricción y la configuración del relieve.

Cuando entre dos masas de aire adyacentes existe una diferencia de densidad, el aire tiende a fluir desde las regiones de mayor presión a las de menor presión. En un planeta sometido a

rotación, este flujo de aire se verá influenciado, acelerado, elevado o transformado por el efecto de Coriolis en cualquier punto de la superficie terrestre.

La creencia de que el efecto de Coriolis no actúa en el ecuador es errónea: lo que sucede es que los vientos van disminuyendo de velocidad a medida que se acercan a la zona de convergencia intertropical, y esa disminución de velocidad queda automáticamente compensada por una ganancia en altura del aire en toda la zona ecuatorial.

A su vez, esa ganancia en altura da origen a la formación de nubes de gran desarrollo vertical y a lluvias intensas y prolongadas, ampliamente repartidas en la zona de convergencia intertropical, en especial en dicha zona ecuatorial. La fricción superficial con el suelo genera irregularidades en estos principios y afecta al régimen de vientos, como por ejemplo el efecto Föhn.

Globalmente hablando, el factor originador y predominante a gran escala es la diferencia de calentamiento entre unas zonas y otras de acuerdo con determinados factores geográficos y astronómicos, así como por variaciones estacionales o temporales producidas por los movimientos de rotación y traslación del planeta.

Cuando se habla del viento se hace referencia siempre a los vientos en la superficie terrestre hasta cierta altura, que varía según la latitud, el relieve y otros factores. A su vez, este movimiento superficial del aire, denominado *viento*, tiene una compensación en altura que casi siempre sigue una trayectoria opuesta a la de los verdaderos vientos en la superficie.

Así, una depresión, un ciclón o un área de baja presión en la superficie producida por el calentamiento superficial del aire obliga al aire caliente a ascender y da origen a una zona de alta presión en altura donde el aire frío y seco desciende hacia las zonas desde donde procedía el aire en la superficie: de esta manera se forman los cumulonimbos, tornados, huracanes, frentes y otros fenómenos meteorológicos.

Una compensación en altura a la dirección de los vientos son las corrientes en chorro que se producen a gran altura y a gran velocidad. La extraordinaria velocidad de estas corrientes en altura (unos 250 km/h) en sentido aproximado oeste - este se debe a la escasa densidad del aire a la altura donde se producen. En efecto, estos vientos compensan a los vientos alisios que se dirigen superficialmente entre Europa y América del Sur a través del Atlántico y también entre Asia y América del Norte en la misma dirección y con las mismas características.

Como estas corrientes en chorro tienen una altura similar a la que usan los aviones en sus vuelos trasatlánticos, la diferencia entre el vuelo en un sentido o en otro puede ser de un par de horas o más en los trayectos largos. Por otra parte, las grandes velocidades de estas corrientes, que a baja altura podrían ser catastróficas para los aviones, a más de 10 km de altura no resultan tan problemáticas por la escasa densidad del aire.

Los vientos se definen también como un sistema que utiliza la atmósfera para alcanzar el equilibrio mecánico de fuerzas, lo que permite descomponer y analizar las características de éste. Es muy habitual simplificar las ecuaciones de movimiento atmosféricas mediante distintas componentes de vientos que, sumados, dan lugar al viento existente.

La componente del viento geostrófico es *el resultado de realizar el equilibrio entre la fuerza de Coriolis y la fuerza del gradiente de presión*. Este viento fluye paralelo a las isobaras, y se puede decir que los efectos de la fricción en latitudes medias son despreciables para las capas altas de la atmósfera.<sup>5</sup> El viento térmico es un viento que diferencia dos niveles que sólo existen en una atmósfera con gradientes de temperatura horizontales o baroclinia.<sup>6</sup> El viento del gradiente es similar al geostrófico pero también incluye el equilibrio de la fuerza centrífuga.<sup>7</sup>

## **2.5 Características físicas del viento.**

El estudio sistemático de las características del viento es muy importante para:

- dimensionar estructuras de edificaciones elevadas.
- diseñar campos de generación eólica de energía eléctrica
- proyectar y preparar las rutas de vuelo y marítimas..

La medición de la velocidad y dirección del viento se efectúa con instrumentos registradores llamados anemómetros, que disponen de dos sensores: uno para medir la velocidad y otro para medir la dirección del viento. Las mediciones se registran en anemógrafos.

Para que las mediciones sean comparables con las mediciones efectuadas en otros lugares del planeta, las torres con los sensores de velocidad y dirección deben obedecer a normativas estrictas dictadas por la **OMM - Organización Meteorológica Mundial**.

En España, la institución encargada por el Gobierno, de la medida, análisis y pronóstico de las variables meteorológicas y climáticas, es la Agencia **Estatal de Meteorología -AEMET**

### **a) Velocidad de los vientos**

El instrumento más antiguo para conocer la dirección de los vientos es la veleta que, con la ayuda de la rosa de los vientos, define la procedencia de los vientos, es decir, la dirección desde donde soplan. La manga de viento utilizada en los aeropuertos suele ser bastante grande y visible para poder ser observada desde los aviones tanto en el despegue como, en especial, en el aterrizaje.

La velocidad, esto es la rapidez y dirección de los vientos se mide con el anemómetro, que suele registrar dicha dirección y rapidez a lo largo del tiempo. La intensidad del viento se ordena según su rapidez utilizando la escala de Beaufort. Esta escala se divide en varios tramos según sus efectos y/o daños causados, desde el aire en calma hasta los huracanes de categoría 5 y los tornados.

Número de Beaufort	Denominación	Velocidad viento Km/h	Efectos en tierra
0	Calma	0 a 1	Calma, el humo asciende verticalmente
1	Ventolina	2 a 5	El humo indica la dirección del viento
2	Brisa muy débil	6 a 11	El humo indica la dirección del viento
3	Brisa débil	12 a 19	Se agitan las hojas, ondulan las banderas
4	Brisa moderada	20 a 28	Se levantan polvo y papeles, se agitan las copas de los árboles
5	Brisa fuerte	29 a 38	Oscilan árboles pequeños
6	Viento	39 a 49	Se agitan ramas gruesas, dificultad para mantener el paraguas abierto
7	Viento fuerte	50 a 61	Los arboles gruesos oscilan
8	Temporal	62 a 74	Se rompen las ramas de los árboles, dificultad para caminar
9	Temporal fuerte	75 a 88	Daños en árboles, imposible andar contra el viento
10	Temporal duro	89 a 102	Árboles arrancados, daños en construcciones
11	Temporal muy duro	103 a 117	Estragos abundantes en construcciones
12	Temporal huracanado	> 118	Destrucción total

Escala de Beaufort, con sus códigos numéricos y los efectos que genera en la superficie.

El récord de mayor rapidez del viento en la superficie terrestre lo tiene el Monte Washington en New Hampshire (Estados Unidos), con 231 millas por hora, es decir, 372 km/h, registrado en la tarde del 12 de abril de 1934. La causa de esta rapidez tan grande del viento está en la configuración local del relieve, que forma una especie de ensilladura de norte a sur que fuerza al viento del oeste a concentrarse en el paso como si fuera un embudo. Es importante señalar que esta enorme rapidez sólo se alcanza en una especie de tobera poco extendida, siendo mucho menor a una corta distancia de este punto. Todas las cordilleras del planeta tienen puntos similares, donde los vientos soplan con fuerza por la existencia de abras, pasos, collados o ensilladuras donde se concentra y acelera el paso del viento.

### **b) Medida del viento**

La dirección del viento es el punto cardinal desde el que se origina éste y se mide con la veleta. Por ejemplo, el viento del norte viene, obviamente, desde el norte y se dirige hacia el sur.<sup>9</sup> En los aeropuertos se usan las mangas de viento para indicar la dirección del viento y estimar la velocidad a partir del ángulo que forma la manga con el suelo.<sup>10</sup> Las veletas tienen indicadas en la parte inferior las direcciones de los vientos con los puntos cardinales y los puntos intermedios, conformando así lo que se conoce como rosa de los vientos, que se emplean con una brújula en los mecanismos de navegación de las embarcaciones desde hace muchos siglos. La velocidad del viento se mide con anemómetros, de forma directa mediante unas palas rotativas o indirectamente mediante diferencias de presión o de velocidad de transmisión de ultrasonidos. Otro tipo de **anemómetro** es el tubo pitot que determina la velocidad del viento a partir de la diferencia de presión de un tubo sometido a presión dinámica y otro a la presión atmosférica.

### **c) Circulación general de los vientos**

El movimiento del aire en la troposfera, que es el que mayor importancia tiene para los seres humanos, siempre tiene dos componentes: la horizontal, que es la más importante (cientos y hasta miles de km) y la vertical (10 km o más) que siempre compensa, con el ascenso o el descenso del aire, el movimiento horizontal del mismo.

El ejemplo de los tornados sirve para identificar el proceso de compensación *entre el avance horizontal del aire en movimiento y el ascenso del mismo*: el remolino inicial de un tornado gira a gran velocidad levantando y destruyendo casas y otros objetos, pero en la medida en que asciende el viento, el cono giratorio del tornado se hace más ancho, por lo cual disminuye su velocidad de giro.

Dicho ejemplo de los tornados es muy útil porque se ha logrado obtener una información estupenda, de primera mano y estudiar bien todos los procesos generales que ocurren en cualquier tipo de viento.

Pero en especial, la transformación del movimiento lineal del viento superficial en un movimiento giratorio de ascenso vertical del mismo puede verse en cualquier remolino o tornado fácilmente y hasta en cualquier nube de desarrollo vertical como un cumulonimbo o un huracán: varía el tamaño o extensión pero el proceso es el mismo.

Y en tipos de vientos que recorren grandes distancias ocurre el mismo proceso. Así tenemos que los vientos alisios, que circulan entre los trópicos y el ecuador, recorren grandes distancias en sentido noreste-suroeste en el hemisferio norte y en sentido sureste-noroeste en el hemisferio sur.

Pero estos vientos cuando llegan cerca del ecuador ascienden forzosamente, no tanto por la convergencia intertropical, sino por el abultamiento ecuatorial, que es mucho más notorio por razones de densidad en los océanos que en los continentes, y aún más notorio en la atmósfera que en los océanos y al ascender por la fuerza centrífuga del movimiento de rotación terrestre, producen nubes de desarrollo vertical y lluvias intensas, con lo que su velocidad de traslación disminuye rápidamente. Al enfriarse el aire ascendente y perder la humedad que traían con la condensación y posterior precipitación tenemos un aire frío y seco.

Como el aire muy frío es más pesado, tenderá a bajar hacia la superficie formando una especie de plano inclinado que va desde el ecuador hasta los trópicos, siendo su dirección la opuesta a la de los alisios. Esta corriente de aire o viento en la zona superior y media de la troposfera va bajando y desviándose hacia la derecha hasta completar el ciclo de los alisios. Vemos así que el principio de conservación de la materia se cumple perfectamente aquí y los alisios se ven compensados casi perfectamente por los vientos en altura que fueron denominados contralisios, aunque este nombre no haya tenido mucho éxito.

Numerosos trabajos que se refieren al tema de los contralisios niegan su existencia, tal vez porque ese retorno de aire seco y frío se hace sin nubes, con lo que no se puede ver la trayectoria de los mismos. Pero la comprobación experimental de los mismos puede verse en la carencia de nubes en el mar de las Antillas: la alta presión originada por los vientos de retorno denominados contralisios da origen al descenso de un aire frío y seco y los climas de las islas donde este proceso ocurre da origen a un clima inusualmente seco.

El mismo proceso puede verse en los grandes desiertos, donde las noches son sumamente frías y los días sumamente cálidos, en los que pueden darse enormes amplitudes térmicas diarias de 30 y hasta 40 °C.

#### **d) Tipos de vientos**

De acuerdo con la escala o dimensión del recorrido de los vientos tenemos tres tipos de vientos: los vientos **planetarios**, los vientos **regionales** y los **locales**, aunque hay algunos tipos, como los monzones, que son más difíciles de determinar y que ocupan variantes dentro de esta simple clasificación.

Los *vientos globales*, constantes o planetarios, se generan principalmente como consecuencia del movimiento de rotación terrestre, que origina un desigual calentamiento de la atmósfera por la insolación y proceden de centros de acción dispuestos en franjas latitudinales de altas y bajas presiones, es decir, de anticiclones y depresiones.

Estos cinturones se disponen aproximadamente en las latitudes ecuatoriales, subtropicales y polares (círculos polares) y se encargan de transportar una cantidad de energía realmente enorme, ante la cual, la posibilidad de un calentamiento global de carácter antropogénico parecería no tener ningún valor.

Estos vientos son conocidos como alisios en las latitudes intertropicales y vientos del oeste en las zonas templadas.

Otro tipo de viento planetario es el monzón que afecta a Asia y el océano Índico y se genera por las diferencias estacionales de temperatura entre los continentes y el mar. Existen algunos autores que incluyen a los monzones como vientos estacionales ya que se producen, en sentido inverso, en el verano y el invierno.

Durante el verano, el continente (en este caso, Asia) se calienta más que el Océano Índico, por lo que se produce una zona de baja presión continental, que atrae los vientos cálidos y húmedos del océano Índico, que dan origen a precipitaciones muy intensas porque la cordillera del Himalaya y otras constituye una barrera a dichos vientos y obliga al aire a ascender, produciéndose lluvias orográficas.

Durante el invierno, por el contrario, el océano se encuentra más caliente que el continente, por lo tanto, los monzones se desplazan del continente hacia el Océano Índico adonde llevan cielos sin nubes y aire seco, por la escasa cantidad de humedad de las tierras continentales.

#### ***Zona de convergencia intertropical***

La zona de convergencia intertropical es un cinturón de bajas presiones (Strahler señala que este cinturón tiene una presión ligeramente por debajo de lo normal, por lo común entre 1009 y 1013 mb (milibares) y está determinada por el movimiento de rotación terrestre el cual genera lo que se conoce como abultamiento ecuatorial terrestre, mucho más notorio, por la diferente densidad, en los océanos que en los continentes y aún más notorio en la atmósfera que en los océanos. En el diagrama de la circulación global de los vientos puede verse ese mayor abultamiento de la atmósfera en la zona ecuatorial (a la derecha del mismo). Es por ello que el espesor de la atmósfera es mucho mayor en la zona intertropical (la troposfera alcanza casi los 20 km de altura), mientras que en las zonas polares es mucho más delgada.

### ***Zonas de divergencia subtropical***

Son zonas de subsidencia de aire frío procedente de grandes alturas en la zona de convergencia intertropical, es decir, de la franja ecuatorial, y que dan origen, a su vez, a los vientos alisios, que se regresan hacia el ecuador a baja altura, y a los vientos del oeste, que van incrementando su velocidad a medida que aumentan también de latitud.

### ***Zonas de convergencia polar***

Son zonas de baja presión que atraen a los vientos provenientes de las latitudes subtropicales. Estos vientos traen masas de aire más cálidas y húmedas, humedad que van perdiendo por condensación (lluvias, rocío y escarcha) a medida que van encontrando aire más frío con el aumento de la latitud. Esta humedad relativa es la que abastece de hielo por escarcha los casquetes polares de Groenlandia y la Antártida

### ***Vientos regionales***

Son determinados por la distribución de tierras y mares, así como por los grandes relieves continentales. Los monzones también podrían considerarse como vientos regionales, aunque su duración en el tiempo y su alternabilidad estacional los convierten más bien en vientos planetarios.

### ***Vientos locales***

Como los demás tipos de vientos, los vientos locales presentan un desplazamiento del aire desde zonas de baja presión a zonas de alta presión, determinando los vientos dominantes y los vientos reinantes<sup>15</sup> de un área más o menos amplia. Aun así hay que tener en cuenta numerosos factores locales que influyen o determinan los caracteres de intensidad y periodicidad de los movimientos del aire. Estos factores, difíciles de simplificar por su multiplicidad, son los que permiten hablar de vientos locales, los cuales son en muchos lugares más importantes que los de carácter general. Estos tipos de vientos son los siguientes:

- **Brisas** marina y terrestre
- Brisa de valle
- Brisa de montaña

**Viento catabático.** Vientos que descienden desde las alturas hasta el fondo de los valles producido por el deslizamiento al ras de suelo del aire frío y denso desde los elementos del relieve más altos. Aparecen de forma continuada en los grandes glaciares, adquiriendo enormes proporciones en la capa de hielo de Groenlandia y de la Antártida, donde soplan a velocidades continuas que superan los 200 km/h motivado por la ausencia de obstáculos que frenan su aceleración.

**Viento anabático.** Vientos que ascienden desde las zonas más bajas hacia las más altas a medida que el sol calienta el relieve.

### 3. Modelos numéricos de circulación Atmosférica

#### 3.1 Ecuaciones físicas que rigen el comportamiento atmosférico.

El movimiento del aire en las proximidades de la superficie terrestre está controlado por cuatro factores:

- La fuerza del gradiente de presión o bórica.
- La fuerza de Coriolis
- La aceleración centrípeta
- La fuerza de rozamiento

De éstos, la fuerza de Coriolis y la aceleración centrípeta son, en rigor, ficticios, pero conviene atribuir a dichas fuerzas ciertos efectos sobre el movimiento del aire en las proximidades de la superficie terrestre.

Fuerza de la gravedad  $\frac{\vec{F}_g}{m} = -g_0 \vec{k}$

Fuerza bórica  $\frac{\vec{F}_B}{m} = \frac{-1}{\rho} \vec{\nabla} p$

Fuerza de Coriolis  $\frac{\vec{F}_C}{m} = -2 \vec{\Omega} \times \vec{V}$

Fuerza de rozamiento

$$\frac{\vec{F}_R}{m} = \frac{-\mu}{\rho} \nabla^2 \vec{V} + \frac{\partial}{\partial z} \left( K \frac{\partial \vec{V}}{\partial z} \right)$$

Veamos brevemente, cada una de ellas, y la importancia que tienen en la evolución del viento:

#### **Fuerza del gradiente de presión:**

El gradiente de presión, ocasionado por la diferencia de presión entre dos puntos, genera una fuerza que ocasiona el movimiento del aire desde la zona de altas presiones hacia las de bajas presiones. El aire, al calentarse, varía su densidad y por lo tanto su presión atmosférica. Las diferencias barométricas ponen en movimiento las masas de aire, las que entran en una constante pero inalcanzable "búsqueda de equilibrio bórico".

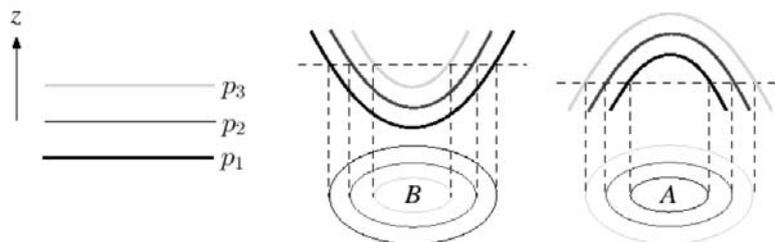
La transformación de energía calorífica en energía cinética puede implicar un ascenso o descenso del aire, pero los movimientos verticales son generalmente mucho menos evidentes que los horizontales,

que pueden abarcar amplias zonas y persistir durante períodos de tiempo que oscilan entre algunos días y varios meses.

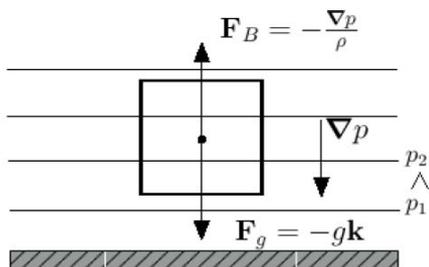
Cabría esperar que la diferencia de presión existente entre la superficie de la tierra y los niveles superiores de la atmósfera ocasionase el escape de ésta, hecho que no se produce a causa de la existencia del campo gravitatorio terrestre. El descenso de la presión del aire al aumentar la altura está compensado por la fuerza de la gravedad dirigida hacia abajo; esto es lo que se conoce como equilibrio hidrostático.

Este equilibrio, conjuntamente con la estabilidad general de la atmósfera y su escaso espesor, limita en gran manera los movimientos verticales del aire. Por término medio, la velocidad de los vientos horizontales es del orden de varios centenares de veces mayor que la de los movimientos verticales, aunque se producen algunas excepciones, particularmente en las tormentas convectivas.

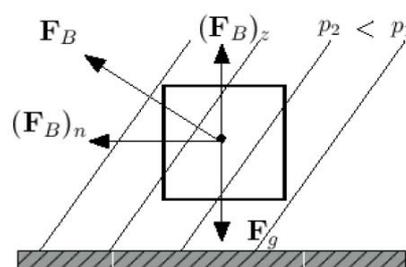
### Dinámica atmosférica. Gravedad y gradiente de presiones



Formación de gradientes horizontales de presión por deformación de las superficies isobáricas planas



Componente vertical del gradiente de presiones: eq. hidrostático



Componente horizontal del gradiente de presiones: viento

### Fuerza de Coriolis

La fuerza de Coriolis una consecuencia del hecho de que el movimiento de las masas de aire sobre la superficie de la tierra está generalmente referido a un sistema de coordenadas móvil ( por ejemplo, la red de meridianos y paralelos que gira con la tierra).

Debido a la rotación del globo, cualquier movimiento en el hemisferio norte es desviado hacia la derecha, si se mira desde nuestra posición en el suelo (en el hemisferio sur es desviado hacia la izquierda). Esta aparente fuerza de curvatura es conocida como fuerza de Coriolis.

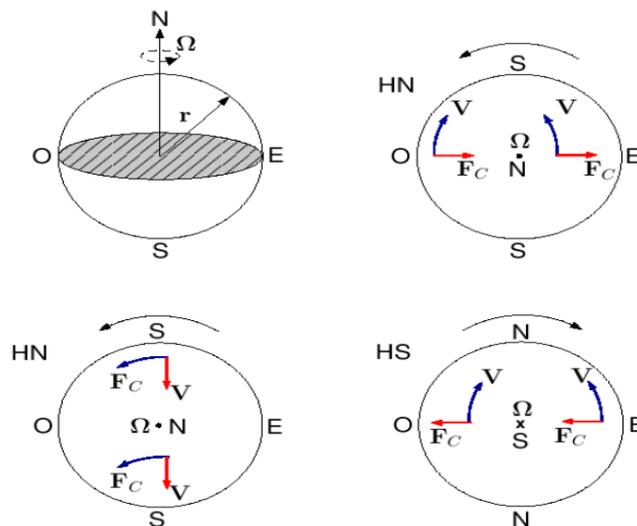
La fuerza de Coriolis es un fenómeno visible. Las vías del ferrocarril se desgastan más rápidamente de un lado que del otro. Las cuencas de los ríos están excavadas más profundamente en una cara que en la otra (de cual se trate depende en que hemisferio nos encontremos : en el hemisferio norte las partículas sueltas son desviadas hacia la derecha). En el hemisferio norte el viento tiende a girar en el sentido contrario al de las agujas del reloj (visto desde arriba) cuando se acerca a un área de bajas presiones. En el hemisferio sur el viento gira en el sentido de las agujas del reloj alrededor de áreas de bajas presiones.

Para que aparezca la fuerza de Coriolis es necesario que el aire se mueva respecto de la Tierra. Esto es la velocidad relativa del aire debe de ser distinta de cero. Se puede ver que la fuerza de Coriolis vale por unidad de masa (módulo)

$$F_C = -2 \Omega \text{sen}(\phi) v$$

**siendo  $v$  la velocidad relativa y  $\Omega$  la velocidad angular de la tierra**

Dinámica atmosférica. Fuerza de Coriolis



Veamos en este punto, algunas definiciones, que surgen de aproximar las ecuaciones anteriormente descritas, que simplificarán bastante los cálculos:

### ***Viento geostrófico .***

El viento geostrófico es una aproximación física al viento real. En él se considera que existe un equilibrio entre la fuerza de Coriolis y la fuerza generada por el gradiente de presión o fuerza bórica (a esto se le llama aproximación geostrófica o equilibrio geostrófico) mientras que, para simplificar el problema, se eliminan de las ecuaciones la aceleración centrípeta y las fuerzas de rozamiento.

Supongamos que tenemos una zona de altas presiones y otra de bajas presiones. Por sí solo eso es suficiente para que se establezca un flujo de aire de las altas a las bajas presiones. Supongamos, además, que las isobaras son rectas, lo cual nos permitirá eliminar la aceleración centrípeta. Así pues, el flujo del viento que se establecería sería perpendicular a las isobaras.

No obstante, como la Tierra rota sobre sí misma, no es un sistema inercial, por lo que aparece una fuerza ficticia como la de Coriolis. La aceleración de Coriolis tiende a desviar los flujos a su derecha en el hemisferio norte y a la izquierda en el hemisferio sur. Si se deja actuar a la fuerza de Coriolis suficiente tiempo y se supone que no existe rozamiento se llega a establecer un flujo paralelo a las isobaras, en donde la fuerza de Coriolis contrarresta exactamente la fuerza generada por el gradiente de presión, por lo que la situación es estable.

### ***Viento de gradiente.***

El viento del gradiente es una aproximación física al viento real. Aunque el nombre se puede aplicar en general a todo viento horizontal que sea paralelo a las isobaras y en el que la aceleración tangencial es nula, es más habitual hacerlo cuando se considera que el viento es producto del equilibrio entre el gradiente de presión, el efecto de Coriolis y la aceleración centrípeta. Es este último término el que permite considerar que la trayectoria del viento es curvada, lo que constituye un gran avance respecto al viento geostrófico, por lo que se considera a veces que el viento del gradiente es un refinamiento del modelo geostrófico. El viento del gradiente realiza (al igual que el geostrófico) la simplificación de suponer que las fuerzas de rozamiento son despreciables.

### ***Viento térmico***

El viento térmico no es un viento real sino más bien un concepto práctico usado en meteorología para calcular la variación del viento entre dos alturas cuando se conoce la estructura térmica de la masa de aire.

Cuando existen diferencias de temperatura en horizontal (atmósfera baroclina) el viento geostrófico varía con la altura. Podemos considerar que el viento geostrófico en altura es entonces la suma del viento geostrófico en la base más un término al que se llama viento térmico. El viento térmico es paralelo a las isotermas. En el hemisferio norte deja el aire frío a su izquierda y el cálido a su derecha, mientras que en el hemisferio sur sucede al revés.

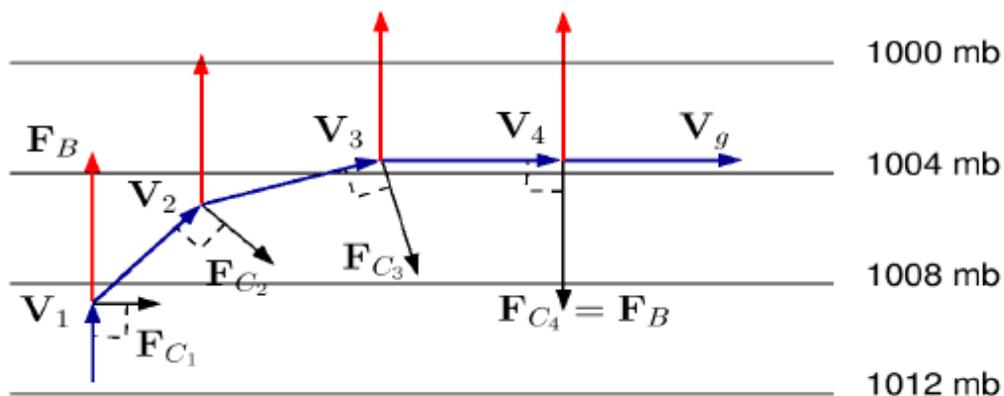
### **Aceleración centrípeta**

Para que un cuerpo siga una trayectoria curva debe existir una aceleración hacia dentro (c), hacia el centro de rotación. La aceleración centrípeta en general es pequeña y sólo adquiere importancia en el caso de vientos que se muevan a gran velocidad siguiendo su trayectoria muy curvados, es decir en las proximidades de las Bajas presiones muy intensas.

### **Fuerza de rozamiento**

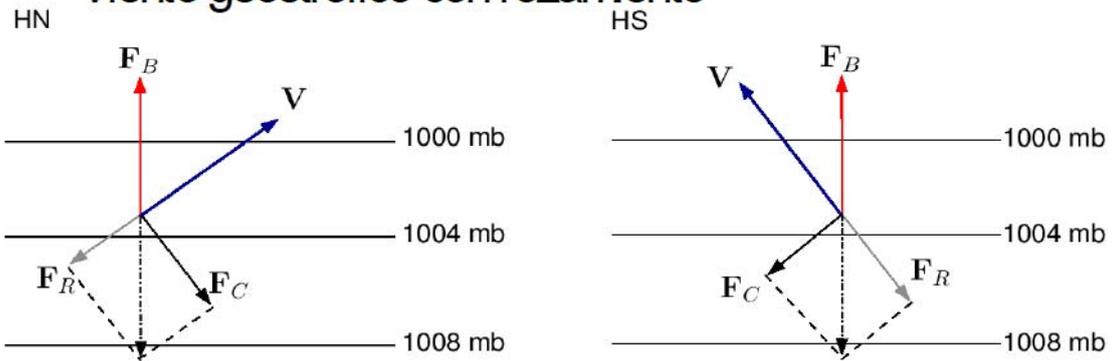
Una fuerza que tiene un efecto importante sobre el movimiento del aire es el que se debe a la fricción con la superficie terrestre. Hacia la superficie (por debajo de los 500 m en terrenos llanos) la fricción empieza a reducir la velocidad del viento por debajo del valor geostrófico. Esta capa de influencia de la fricción se denomina capa límite planetaria. La velocidad del viento disminuye exponencialmente cerca de la superficie terrestre debido a los efectos de rozamiento de la superficie. Éstos implican hacer fricción sobre los obstáculos, edificios, árboles, colinas..., y el esfuerzo ejercido por el aire en la superficie de contacto.

### **Formación viento geostrófico sin rozamiento**

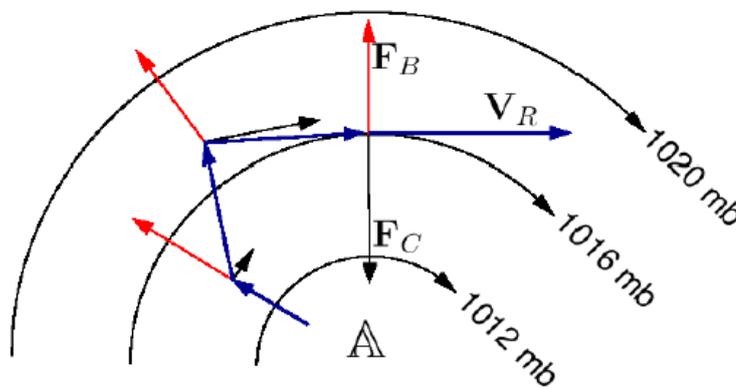
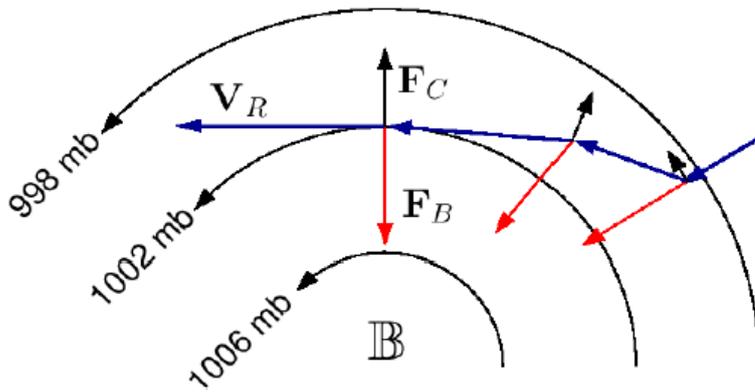


$$F_C = F_B \rightarrow V_g = \frac{1}{2\Omega \rho \sin \phi} \frac{\partial p}{\partial n} = \frac{1}{\rho f} \frac{\partial p}{\partial n} = \frac{g}{f} \frac{\partial z}{\partial n}$$

## Viento geostrófico con rozamiento



## Movimientos horizontales. Viento de gradiente



$$F_B = F_C \pm \frac{V_R^2}{R} \rightarrow V_R = V_g \pm \frac{V_R^2}{fR}$$

### 3.2 Modelos numéricos de resolución.

En primer lugar cabe destacar, que existen diversos modelos de circulación Atmosférica, todos ellos intentan resolver de manera numérica las ecuaciones diferenciales que rigen la evolución de las masas de aire. Cabe aquí preguntarnos el porqué de los modelos numéricos, está bien claro que las ecuaciones anteriormente descritas, son imposibles de resolver exactamente, ya que se trata de un complicado sistema de ecuaciones diferenciales acopladas. Es por ello que es necesario acudir a modelos numéricos para tratar de resolver este tipo de problemas de contorno.

Vamos a ver una pequeña recopilación de estos modelos, clasificados por países y características de evaluación de cálculos.

**GFS España:** Modelo Americano de Aviación, tiene 4 salidas al día (00 06 12 18) horas y la modelización acaba aproximadamente 5 horas después. Usa una resolución espacial de unos 1º de longitud.

**ECMWF:** Modelo del Centro Europeo de Predicción a Medio Plazo, es el modelo de predicción que emite este centro al que pertenecen 23 estados europeos entre ellos España, tiene 4 salidas al día (00 06 12 18) horas, aunque en internet solo aparece una salida.

**HIRLAM:** Modelo regional (de área limitada) de predicción ofrecido por INM, tiene 4 salidas al día (00 06 12 18) horas, con una resolución temporal de 48 horas, y espacial de 0.5º. En internet solo aparece la salida de las 00 y 12 horas.

**GFS:** Modelo americano de Aviación, tiene 4 salidas al día (00 06 12 18) horas y la modelización acaba aproximadamente 5 horas después. Usa una resolución espacial de 1º.

**MRF:** Modelo central americano, los mapas son actualizados entre las 8 y 9 UTC. Tiene una resolución de 1º de longitud.

**MASS:** Modelo regional (de área limitada) de predicción ofrecido por Meteosim.

**NOGAPS:** Modelos de la Marina de los EEUU, tiene 2 salidas al día (00 12) horas, y se actualiza a las (8 20) horas UTC.

**DWD:** Modelo de predicción del Servicio Alemán de Meteorología, estas imágenes son ofrecidas por Meteosort.

**ENS:** Modelo de conjunto resultado del MRF, aparecen dos tipos que son: Diagrama de Espaguetis, aparecen las diferentes predicciones consecuencia de la variación de condiciones iniciales del modelo y Mapa Central media de las diferentes predicciones según variación condiciones iniciales.

**UKMO:** Modelo del Servicio Británico de Meteorología, tiene 2 salidas (00 12) horas, y se actualiza a las (6:45 18:45) horas UTC.

**JMA:** Modelo Japonés de predicción.

**GEM:** Modelo meteorológico global canadiense, tiene 1 salida al día (00) horas, con una resolución temporal de 240 horas.

### **Sistema HIRLAM**

De todos estos modelos el que más juego da es el HIRLAM, ya que como se hemos descrito anteriormente, es el usado por la **AEMET**, con lo que vamos a proceder a una pequeña descripción del mismo.

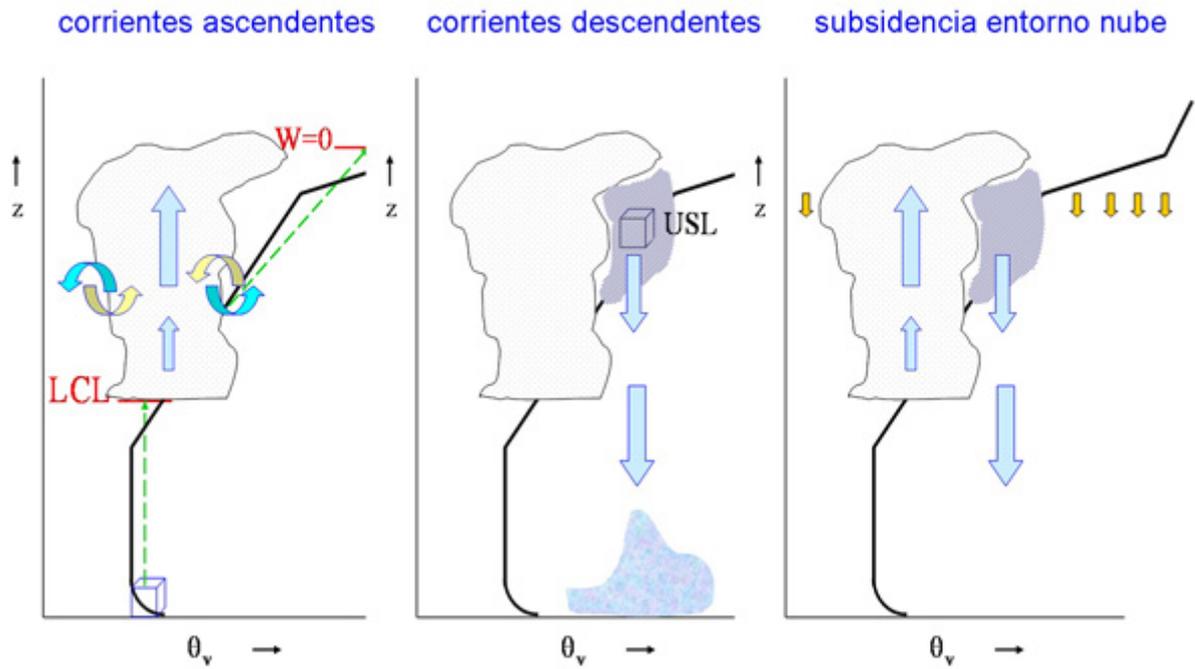
AEMET desarrolla y opera diariamente sistemas de análisis y predicción numérica del tiempo de área limitada desde hace más de 25 años. Desde comienzos de los años 90, colabora con los Servicios Meteorológicos Nacionales de varios países de Europa occidental que conforman el Consorcio HIRLAM en el ámbito de la Predicción Numérica del Tiempo (PNT), para el desarrollo de sistemas operativos de análisis y predicción a corto plazo basados en modelos numéricos de alta resolución. Actualmente se trabaja en una nueva generación de modelos que proporcionarán predicciones a resoluciones horizontales del orden de 1-2km.

La colaboración en el Consorcio HIRLAM se lleva a cabo a través de sucesivos Planes Científicos conjuntos de los países participantes, de varios años de duración, coordinados por un equipo directivo. El grado de consecución obtenido es evaluado por un grupo externo de expertos al término de la vigencia de cada Plan Científico.

Los especialistas de AEMET han contribuido decisivamente al desarrollo de las sucesivas versiones del sistema HIRLAM que se han ido liberando en el transcurso del tiempo en el terreno de la asimilación de datos, las parametrizaciones físicas y la dinámica del modelo. Algunos de estos expertos han liderado distintas áreas de desarrollo desde que nos integramos en el Consorcio.

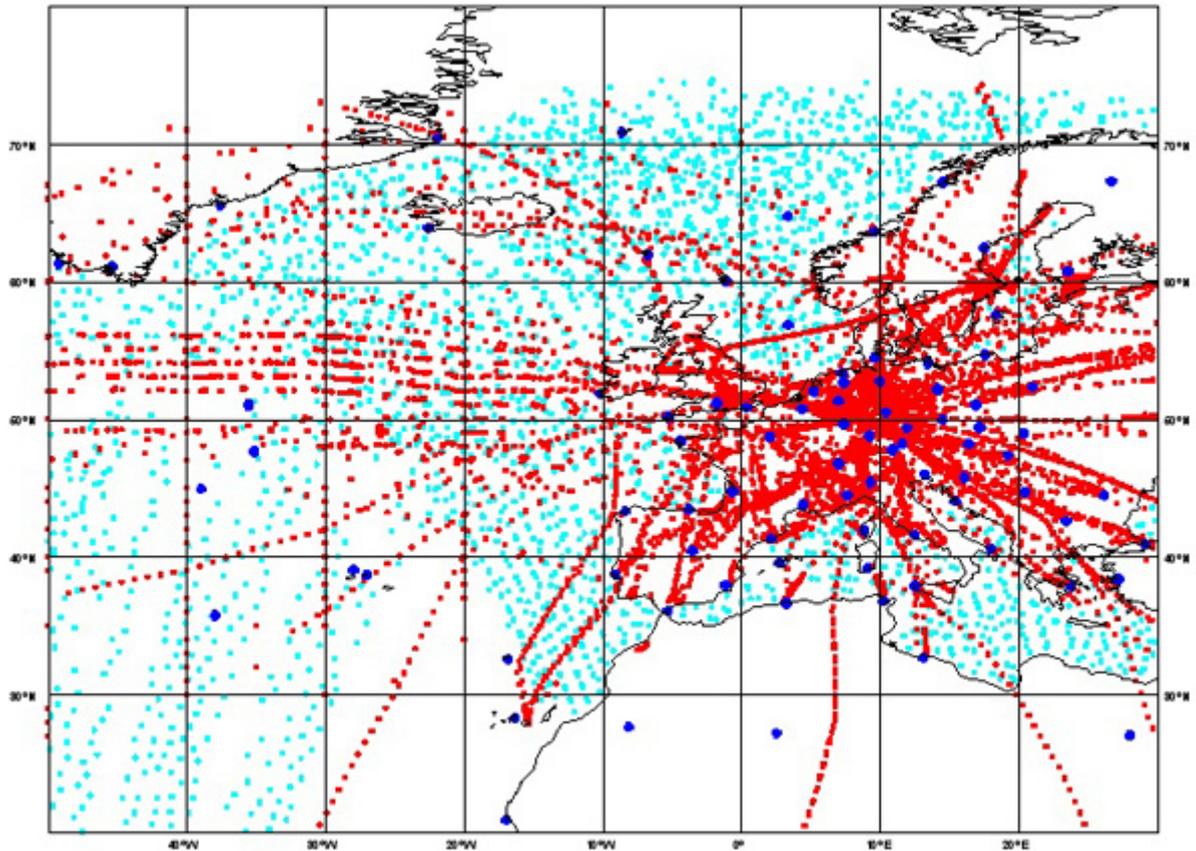
El consorcio HIRLAM está formado por los Servicios Meteorológicos Nacionales de los países escandinavos, Islandia, Estonia, Lituania, Holanda, Irlanda y España

HIRLAM es un modelo hidrostático de puntos de rejilla con una dinámica semilagrangiana, en el que son parametrizados los procesos radiativos y los que suceden a escala sub-rejilla (turbulencia, nubes y condensación, convección, intercambios de agua y energía con la superficie...).



Parametrización de la convección: circulaciones representadas en los esquemas de flujo de masa

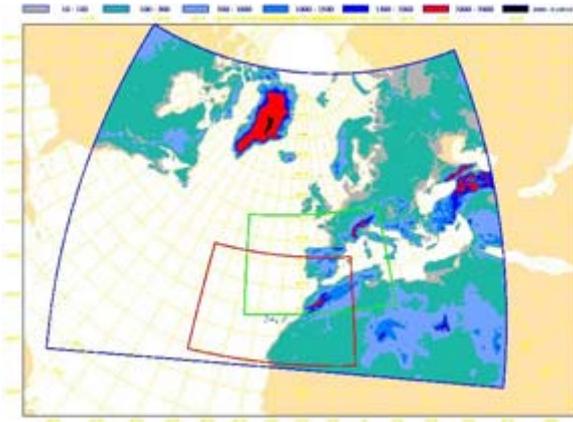
El estado inicial atmosférico, o análisis, se obtiene corrigiendo una primera estimación (basada en una predicción a corto plazo reciente), mediante la asimilación de observaciones convencionales (procedentes de estaciones de superficie en tierra, barcos y boyas, radiosondeos y aviones), así como los datos brutos medidos por los instrumentos a bordo de los satélites meteorológicos, mediante un método variacional tri/tetradimensional (3DVAR o 4DVAR). Los campos iniciales de superficie y suelo (temperatura del agua del mar, espesor y cobertura de nieve, humedad y temperaturas del suelo y subsuelo...) se describen gracias a un sistema de análisis objetivo que utiliza diferentes tipos de observaciones.



Distribución de observaciones meteorológicas asimiladas procedentes de satélite (azul claro), aviones (rojo) y radiosondeos (azul oscuro)

Los desarrollos realizados en la Agencia con el modelo HIRLAM han formado parte de proyectos de los Programas Marco de investigación financiados por la Unión Europea, Acciones COST y experimentos internacionales de intercomparación de modelos sobre situaciones meteorológicas especiales en los que AEMET ha participado. El modelo HIRLAM es uno de los integrantes del sistema pionero de predicción por conjuntos desarrollado en AEMET, SREPS. El sistema HIRLAM también es utilizado para analizar el sistema compuesto europeo de observación meteorológica, que integra todas las redes de los diferentes tipos de medidas. Entre los estudios que se llevan a cabo, está el investigar el impacto en la predicción de fenómenos adversos en el Mediterráneo de la asimilación de observaciones adicionales en zonas especialmente sensibles a las condiciones iniciales (fase final del proyecto MEDEX).

La cadena HIRLAM se ejecuta 4 veces al día en AEMET en 3 dominios distintos: un área euroatlántica con 16km de resolución horizontal y dos centradas en la Península Ibérica y Canarias de 5km de resolución. El número de niveles en la vertical es de 40. Los campos previstos del modelo global del Centro Europeo de Predicción a Plazo Medio (CEPPM) se reciben 4 veces al día y se utilizan como forzamientos en los contornos del dominio de integración y para mejorar la descripción de la componente de larga escala del análisis. Las observaciones utilizadas para determinar el estado inicial atmosférico se reciben regularmente gracias al sistema mundial de telecomunicaciones establecido por la Organización Meteorológica Mundial.

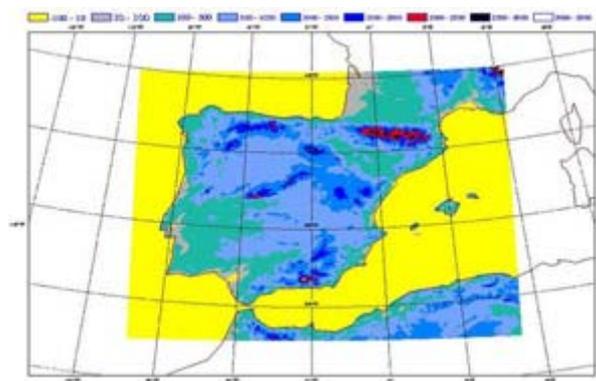


Domínios de integración del modelo Hirlam operativo

### **Sistema HARMONIE**

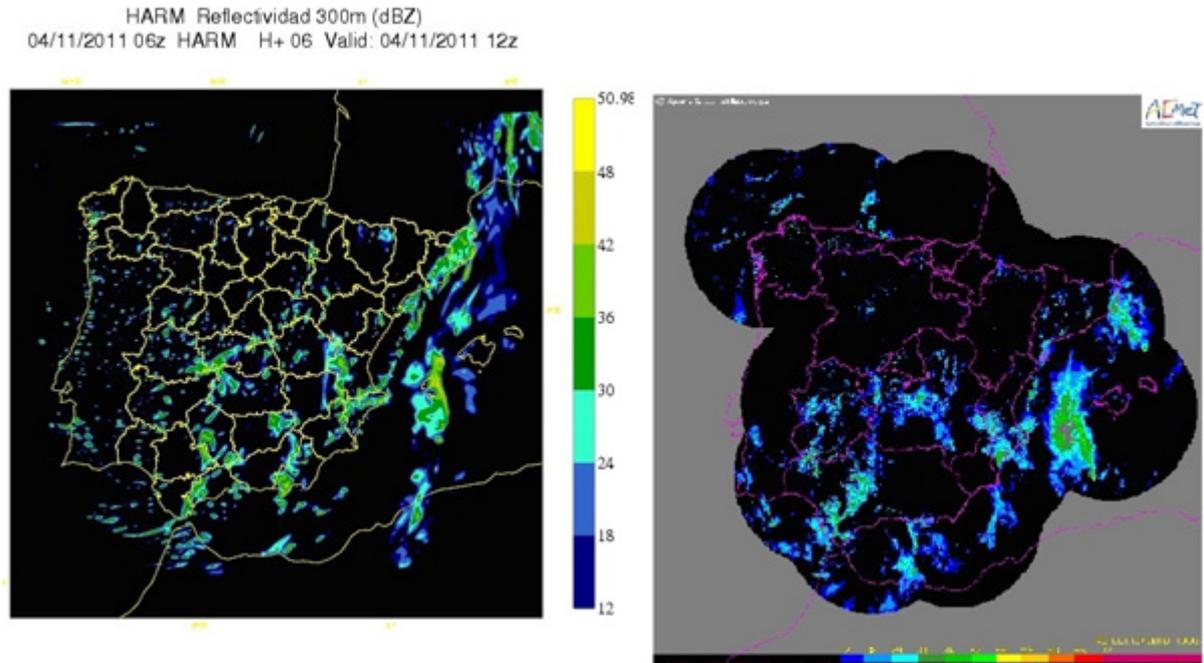
Cuando se avanza hacia resoluciones de unos pocos kilómetros, los efectos no hidrostáticos deben estar representados en los modelos meteorológicos y algunos fenómenos que en resoluciones inferiores deben ser parametrizados comienzan a ser descritos explícitamente. En el año 2006, los Consorcios europeos de PNT HIRLAM y ALADIN acordaron comenzar una cooperación para el desarrollo de sistemas de muy alta resolución basados en modelos no hidrostáticos.

El sistema HARMONIE, que se está desarrollando en HIRLAM, es el resultado de esta colaboración y ha sido diseñado de forma flexible, pudiendo ser utilizado en diferentes escalas espaciales: desde la sinóptica hasta resoluciones inferiores a 1km. Además del modelo de predicción, consta de un sistema de asimilación de datos para la inicialización del estado atmosférico y de un módulo de análisis de superficie. Tiene implementados diferentes conjuntos de parametrizaciones físicas que resultan apropiados para cada escala espacial. El núcleo de la dinámica no hidrostática aprovecha los desarrollos obtenidos por el grupo ALADIN. Las librerías de código se comparten entre el CEPPM, los Consorcios HIRLAM y ALADIN y el Servicio Meteorológico francés.



Orografía del modelo Harmonie a 2,5km de resolución horizontal

El modelo HARMONIE en modo no hidrostático incluye 6 variables de pronóstico para los procesos húmedos: vapor de agua, agua líquida, cristales de hielo, lluvia, nieve y nieve granulada o granizo. Esta versión del modelo se ejecuta diariamente en modo experimental en AEMET sobre un área que cubre la península Ibérica y Baleares a 2,5km de resolución horizontal y 65 niveles en la vertical.

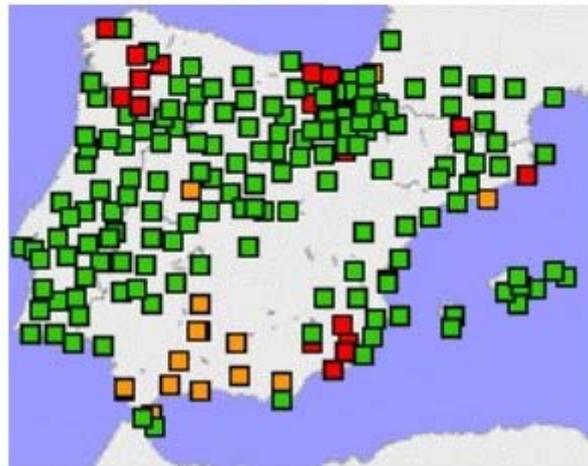
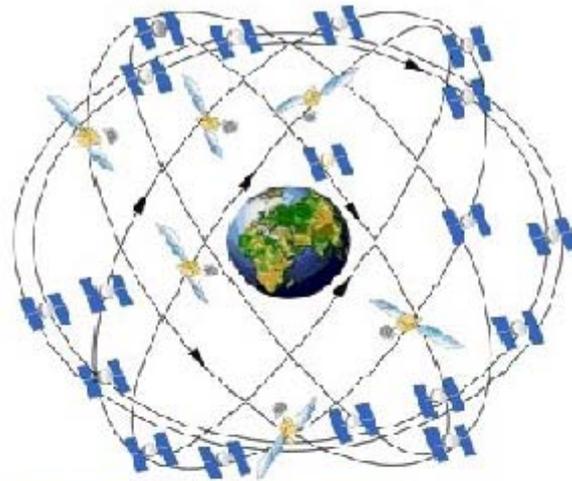


Reflectividades previstas por el modelo HARMONIE (izquierda) y medidas por la red de radares de AEMET (derecha)

La participación de AEMET en el desarrollo de nuevas versiones del modelo HARMONIE se centra tanto en los aspectos de nuevos desarrollos científicos (en asimilación de datos y en el modelo de predicción), como de operación (obtención de nuevas variables a partir de las salidas directas del modelo y mantenimiento de las cadenas operativas).

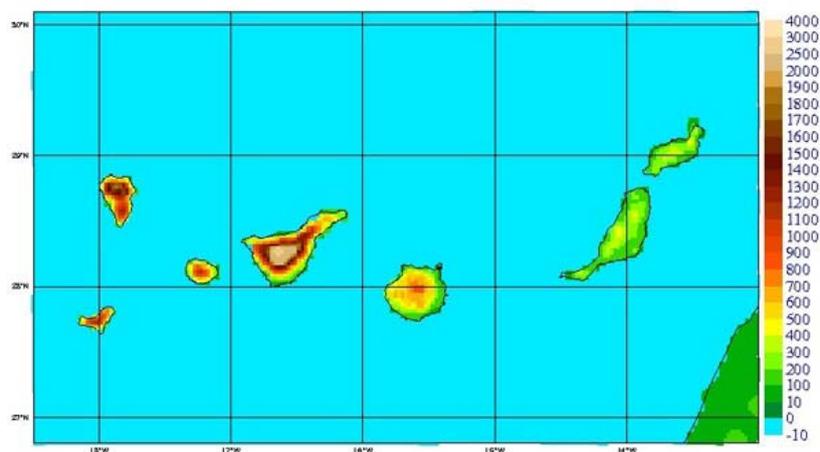
AEMET contribuye activamente a la mejora de la dinámica del modelo, basada en un aumento de la eficiencia y de la precisión (discretización vertical por elementos finitos y conservación de masa), así como a una mejor interacción entre los procesos física-dinámica cuando se aumenta la resolución. En el ámbito de las parametrizaciones físicas, AEMET participa en los desarrollos sobre la representación de los procesos convectivos y los procesos turbulentos y nieblas. En esta última tarea, se colabora con grupos de expertos de la Universidad de las Islas Baleares.

En el campo de la asimilación de datos se trabaja, entre otros, en la asimilación de nuevas fuentes de datos de observación como reflectividades y vientos radiales medidos por los radares de AEMET, y retrasos de la señal de los satélites GPS debidos a la masa y el contenido de humedad atmosféricos. También se trabaja en la inicialización del contenido de humedad del suelo con métodos avanzados.



Distribución de estaciones receptoras de la señal de los satélites GPS

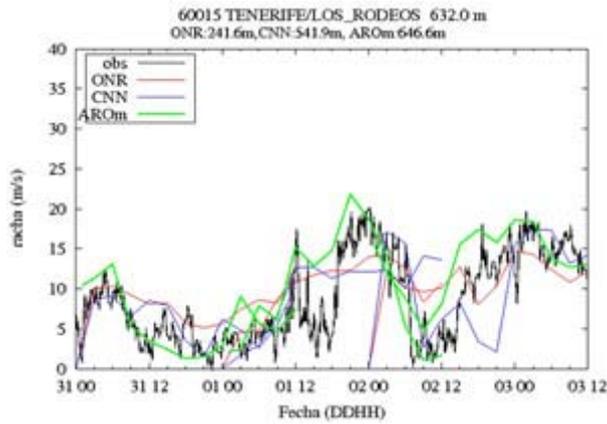
El modelo HARMONIE no hidrostático se prueba también en AEMET, de modo experimental, sobre casos de estudio de fenómenos meteorológicos de tiempo adverso, como los ocurridos en Canarias en el mes de febrero del año 2010.



Orografía del modelo HARMONIE sobre las islas Canarias a 2,5km de resolución horizontal

La complicada orografía de la mayoría de las islas comienza a ser mejor representada por el modelo a 2,5km de resolución, como se puede observar en la figura superior.

Los resultados obtenidos en estos casos de lluvias y vientos fuertes son esperanzadores, a pesar de la poca predecibilidad de los procesos convectivos a escalas pequeñas. La figura contigua muestra el mejor comportamiento del modelo **Harmonie** no hidrostático (en verde), los modelos HIRLAM operativos a 16km (rojo) y 5km de resolución (azul) frente a las observaciones de la racha de viento en Los Rodeos (Tenerife).



Comparación de las predicciones de rachas de viento frente a observaciones

## 4. Variabilidad de los campos de viento

Durante varias décadas se ha registrado el viento en diferentes aeropuertos comerciales cercanos a la costa a lo largo del país, así como en estaciones interiores, estaciones de esquí, y otro tipo de registros. Esta información, junto a registros de presión, temperatura y precipitación pueden obtenerse para intentar predecir en una vía estadística la variabilidad del viento. Diversos trabajos han utilizado estas observaciones de las variables Meteorológicas, para dar predicciones y estimar los valores que tomarían las variables climatológicas, en unos determinados periodos de tiempo, que describiremos a continuación.

Existen factores que conforman la variabilidad del tiempo, en la península ibérica, pasamos a describirlos someramente a continuación:

**Atmosfera:** es obvio, si no hay atmosfera, no hay viento, y ya hemos comentado con anterioridad, la gran cantidad de factores que influyen en la circulación de la atmósfera tanto a gran como a pequeña escala.

**Sol:** es la causa de que en la atmosfera haya cambios de densidad en el aire, lo que provoca vientos: La luz del Sol calienta el suelo, que irradia calor a la atmosfera, por lo que el aire se calienta. Al calentarse el aire cercano al suelo, su densidad disminuye, es decir, se vuelve menos pesado que el aire frío que se encuentra lejos del suelo: esto provoca que el aire caliente tienda a subir y el frío a bajar. El movimiento resultante, se llama viento.

**Rotación Terrestre:** provoca el efecto Coriolis (que no solo afecta a la atmosfera, también a las corrientes marinas): La rotación terrestre provoca turbulencias en la atmosfera, y esto genera movimientos de masas de aire, también llamados vientos. Un ejemplo, sería colocar unas cuantas canicas de cristal (no queremos que estallen) en un plato en el microondas. Al activar el microondas, el plato comienza a girar y puedes comprobar que las canicas se mueven. Así es como funciona el efecto Coriolis en la atmosfera.

**Traslación terrestre:** provoca cambios globales llamados estaciones (primavera, verano, otoño e invierno). Cuando en el hemisferio norte es invierno, en el hemisferio sur es verano... adivina en que hemisferio hace mas frío...

Al igual que antes, el aire caliente tiende a subir y el frío a bajar. De este modo, la estacionalidad también provoca vientos.

**Geografía del lugar:** La cercanía a la costa, presencia de nuestras cadenas de montañas, son factores que generen vientos por variaciones de temperatura (como siempre el aire caliente sube y el frío baja), o de presión (el aire menos denso tiende a subir y el mas denso a bajar), etc.

El objetivo del presente trabajo es explorar, desde un punto de vista estadístico, el comportamiento temporal del viento en la zona de Santander y su bahía, y ver qué relación puede existir con otras partes de la Región, más al sur de la cadena montañosa así como la relación con fenómenos presentes en la meseta Castellana, con énfasis en las fluctuaciones comprendidas entre uno y cuatro días. El análisis se orientó hacia la descripción de estas variaciones y cómo ellas se relacionan entre sí, con el fin de conocer algunos aspectos de la variabilidad temporal del viento en la región.

Vamos a centrarnos, más concretamente en la predicción de la velocidad del viento. Para ello tomaremos una rejilla, de cuatro puntos, en el plano XZ, horizontal, y vamos a extraer medidas del valor de la velocidad del viento, en los cuatro puntos de la rejilla, dados estos a distintas alturas, que llamaremos  $y$ .

Dado que queremos evaluar el valor de  $y$ , la velocidad del viento, esta vendrá dada por:

$$Y = f(X) + \epsilon_m$$

Dónde  $X$ , es el vector de la posición, dónde queremos calcular el valor del viento, y  $\epsilon_m$ , es el valor del error que estamos cometiendo, respecto a su valor real. En este caso, es evidente que tenemos una predicción en cualquier punto, pero si como en nuestro caso, disponemos de un número finito de datos en otras tantas estaciones Meteorológicas, es evidente, que la predicción se realizará para unos puntos determinados, de la rejilla, en este caso serán los extremos.

Con todo esto, la predicción quedará de la siguiente forma:

$$Y = f(X_p) + \epsilon_m + \epsilon_p$$

Dónde  $X_p$ , es el vector de la posición de nuevo en los puntos de la rejilla, dónde queremos calcular el valor del viento, y  $\epsilon_m$ , es el valor del error que estamos cometiendo, respecto a su valor real y  $\epsilon_p$  es el error que se comete, al valorar el valor del viento en estos determinados puntos definidos por la rejilla.

#### **4.1 Datos disponibles para la predicción del viento.**

El estudio se ha realizado con datos de ocurrencia diaria de velocidad del viento de 33 estaciones pertenecientes a la Agencia Estatal de Meteorología (AEMET), seleccionadas de una base de datos de precipitación homogénea del periodo 1957-2002 (ver figura Sólo se permite un 7% de lagunas en el periodo de análisis, definido por la disponibilidad de datos de ERA40 (de septiembre de 1957 a agosto de 2002). Se impone una restricción mayor de un 5% de lagunas permitidas anuales, por lo que puede haber un máximo de 14 días sin datos por año.

Dichos porcentajes se escogieron como un compromiso entre la calidad de las series temporales resultantes y el número de estaciones considerado para el estudio. La región tiene una orografía compleja, con altitudes de las estaciones comprendidas entre 60 y 1350 metros, es decir pasamos desde estaciones en la costa a nivel del mar a otras situadas en la propia cordillera Cantábrica. A continuación extraemos los datos de ERA40:

**ERA40****Rango Geoespacial:**

- **Longitud:** (-180° : 2.5° : 180°)
- **Latitud:** (-90° 2.5° : 90°)

**Rango temporal:****Inicio:** 01/01/1961**Final:** 01/01/2001**Resolución temporal:** 24h

Variable	Niveles	Horas	Unidades	Agregación temporal
U velocidad	300 500 700 850 1000	00	m s <sup>-1</sup>	Instantáneos
Temperatura	300 500 700 850 1000	00	K	Instantáneos
V velocidad	300 500 700 850 1000	00	m s <sup>-1</sup>	Instantáneos
Humedad específica	300 500 700 850 1000	00	kg kg <sup>-1</sup>	Instantáneos
Nivel medio de presión en el día (DM)	0	-	Pa	Valor medio diario
Temperatura media diaria (DM)	0	-	K	Media diaria
Temperatura	0	00	K	Instantáneos
Temperatura Máxima	0	-	K	Valor máximo diario
Precipitación Total (DM)	0	-	m	Valor diario acumulado
Geopotencial	300 500 700 850 1000	00	m <sup>2</sup> s <sup>-2</sup>	Instantáneos

A continuación añadimos la lista con las 33 estaciones de la AEMET, anteriormente seleccionadas:

Estación	Altitud	Lat.	Lon.	$\theta$	HIR	FAR	FAR2
GRADO	60	43.38	-6.06	0.15	0.47	0.00	0.47
SANTESTEBAN	131	43.13	-1.66	0.23	0.48	0.01	0.48
VILLACARRIEDO	212	43.23	-3.80	0.26	0.57	0.00	0.39
MONTAOS-ORDES	306	43.04	-8.42	0.19	0.50	0.00	0.47
AS PONTES	343	43.45	-7.86	0.18	0.61	0.00	0.47
CENICERO	430	42.48	-2.64	0.23	0.53	0.01	0.43
MIRANDA DE EBRO	520	42.68	-2.96	0.23	0.49	0.01	0.45
URRUNAGA	540	42.96	-2.65	0.23	0.47	0.01	0.52
VILLARCAYO	595	42.94	-3.57	0.26	0.63	0.01	0.38
JAVIERREGAY	690	42.59	-0.74	0.26	0.44	0.01	0.52
MONZON DE CAMPOS	754	42.12	-4.49	0.18	0.49	0.01	0.53
OSORNO	809	42.41	-4.36	0.19	0.41	0.02	0.63
SAN MIGUEL DE BERNUY	839	41.40	-3.95	0.26	0.50	0.02	0.44
PANTANO DE SANTA TERESA	840	40.67	-5.60	0.26	0.54	0.02	0.45
ALAR DEL REY	851	42.66	-4.31	0.23	0.59	0.05	0.47
SAN ESTEBAN DE GORMAZ	860	41.57	-3.20	0.23	0.55	0.02	0.41
RETUERTA	900	42.03	-3.51	0.28	0.61	0.04	0.35
LINARES	911	41.53	-3.56	0.26	0.53	0.02	0.47
BESCOS	920	42.63	-0.50	0.26	0.44	0.02	0.53
TORRECILLA	949	42.09	-3.69	0.23	0.34	0.01	0.65
LA MAGDALENA	998	42.78	-5.80	0.30	0.43	0.02	0.50
PANTANO CERVERA	1000	42.87	-4.53	0.30	0.58	0.03	0.44
GARRAY	1010	41.82	-2.45	0.19	0.44	0.02	0.66
PANTANO REQUEJADA	1024	42.91	-4.53	0.30	0.61	0.05	0.41
BOCA DE HUERGANO	1104	42.97	-4.93	0.36	0.60	0.04	0.36
PRIORO	1123	42.89	-4.96	0.31	0.66	0.04	0.34
AVILA	1130	40.65	-4.68	0.30	0.47	0.02	0.42
EMBALSE CUERDA DEL POZO	1150	41.88	-2.70	0.30	0.64	0.03	0.31
RABANAL DE LUNA	1150	42.93	-5.97	0.31	0.67	0.04	0.35
GENESTOSO	1180	43.06	-6.40	0.31	0.76	0.07	0.32
HUERGAS DE BABIA	1222	42.96	-6.09	0.38	0.65	0.08	0.33
PANTANO CAMPORREDONDO	1253	42.90	-4.74	0.34	0.69	0.04	0.39
ZAPARDIEL DE LA RIBERA	1353	40.36	-5.33	0.34	0.67	0.02	0.35
VALOR MEDIO	823	42.42	-4.26	0.26	0.58	0.03	0.48

Aparecen en la tabla anteriormente expuesta los datos, Nombre, altitud (m), localización (latitud y longitud en grados), el umbral de calibración probabilístico  $\Theta$  para cada estación. Las últimas tres columnas muestran los valores de distintos índices de verificación:

- Hit Rate  $HIR \equiv P(p = 1/o = 1)$ .
- False Alarm Rate  $FAR \equiv P(p = 1/o = 0)$ .
- False Alarm ratio o  $FAR2 \equiv P(o = 0/p = 1)$ .

Las estaciones se han ordenado en altitudes.

Los valores de nieve más altos, se corresponden a las estaciones que se encuentran a más altitud, en meses de invierno y verano, sin embargo se aprecian valores altos de la componente U del viento para aquellas estaciones cercanas a la costa, en los meses de invierno, en toda la serie histórica. También se puede apreciar significativamente, la presencia del llamado efecto Foehn, en las zonas cercanas a las costas.

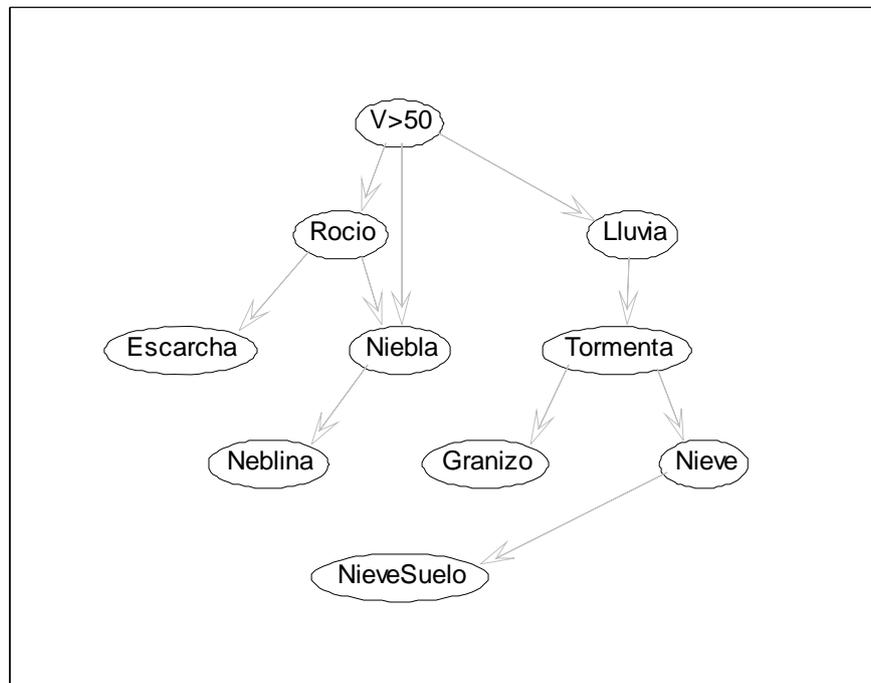
## 5. Descripción del problema de la predicción del viento.

### 5.1 Análisis de los datos existentes.

Los datos existentes, tal y como hemos dicho, con anterioridad, se encuentran en la toolbox Meteolab , en la sección INM, en esta sección podemos crear, una estructura de red, con los datos existentes para cada una de la configuraciones que deseemos. En este caso cargamos en otra red, los datos correspondientes, a la Cuenca Norte, y por último hemos de cargar, en la estructura variable que queremos calcular, la velocidad del viento. Para calcular las variables, usaremos estas estructuras de datos, anteriormente descritas, en los intervalos descritos también.

### 5.2 Correlación del viento con la presión y la temperatura.

Existe, una correlación del viento con la presión y con la temperatura, que no deja de ser bastante importante, para ello podemos usar técnicas de correlación de tipo estadístico, tipo redes Bayesianas, y para una estación de datos cercana a Santander, los resultados obtenidos, son los siguientes:



En este caso estamos observando los valores de la componente U del viento, siempre que esta supere los 50Km/h y la relación que puede existir con otras variables meteorológicas, que se indican en el gráfico. Como podemos observar, las relaciones fuertes en la red Bayesiana, están indicados por las flechas, y van de mayor nivel de relación a menor nivel de relación.

Los datos de probabilidad conjunta, que se esperan son los siguientes:

V>50

0.8558 (Probabilidad de que la velocidad del viento sea mayor que 50Km/h)

0.1442 (Probabilidad de que la velocidad del viento no sea mayor que 50Km/h)

SUCESO CONJUNTO	PROBABILIDAD DE OCURRENCIA	PROBABILIDAD DE NO OCURRENCIA
Rocío	0.748	0.252
Escarcha	0.981	0.019
Nieve	0.931	0.069
Neblina	0.453	0.547
Lluvia	0.444	0.556
Tormenta	0.938	0.062
Granizo	0.975	0.025
Nieve cubre el suelo	0.994	0.006

5.3 Tendencia de los campos de vientos para la rejilla observada.

## 6. Simulación con Técnicas de **Downscaling Estadístico**

### 6.1 Introducción a las técnicas de Downscaling Estadístico.

La palabra *Downscaling Estadístico*, significa, un proceso de troceamiento y análisis discreto de información para su posterior tratamiento estadístico.

El downscaling es en nuestros días un campo maduro y multidisciplinar que involucra diferentes herramientas científicas para acceder y procesar grandes cantidades de datos heterogéneos.

Debido a ello, resultan necesarias herramientas interactivas creadas en entornos sencillos para facilitar el proceso de downscaling para los usuarios finales, de forma que se maximice la explotación de las predicciones disponibles.

Por una parte tenemos las salidas de los GCMs, que no son más que una representación espacial y temporal aproximada de los principales procesos físicos que ocurren en la atmósfera y de sus interacciones con los demás componentes del medio ambiente. De su resolución se obtiene la evolución temporal y espacial (tridimensional) del sistema climático, en función de las condiciones iniciales y de contorno elegidas y de los valores de ciertos parámetros climáticos (por ejemplo, la concentración de CO<sub>2</sub> atmosférico). Esto se denomina experimento numérico, simulación climática o experimento climático y puede orientarse tanto a la descripción del clima contemporáneo (experimentos de control), como a la investigación del clima resultante de uno o más cambios en los parámetros climáticos.

El DP ha sido diseñado siguiendo un enfoque de extremo a extremo conectando de forma transparente a los usuarios con los proveedores de datos. Para conseguirlo, Internet y tecnologías de cómputo distribuido se combinan con herramientas estadísticas para realizar el downscaling: partiendo de las salidas de GCMs hasta llegar a escalas regionales o locales que son las que necesitan las aplicaciones de impacto de cambio climático. De esa forma, los usuarios pueden probar y validar diferentes métodos de downscaling (regresión, redes neurales, análogos, weather typing, etc.) empleando un navegador web, sin preocuparse de los detalles y técnicas empleadas o los datos a los que se ha accedido.

La puesta en marcha de experimentos de downscaling se divide en tres pasos:

- **Predictor:** Definición de la región y los predictores que se usarán en el experimento
- **Predictando:** Definición de uno o varios predictandos de interés en los que realizar el downscaling (con la configuración particular de predictores).
- **Método de Downscaling:** Definición y validación de uno o varios métodos de downscaling para ser aplicados en el experimento (asociando los predictores con los diferentes predictandos).

Una vez que se han definido los elementos de Predictor > Predictando > Método de Downscaling, los métodos de downscaling están preparados para realizar el downscaling de los escenarios (futuros o de control) de los GCMs disponibles.

## 6.2 *Uso de la Toolbox Meteolab.*

El uso de esta herramienta es bastante cómodo para aplicar toda la suerte de métodos estadísticos, que queremos aplicar, para poder resolver, el problema de la predicción para una determinada variable termodinámica. Esta Toolbox, está estructurada en carpetas, con archivos programados en Matlab, de forma, que es sencilla su utilización. Muchos de estos programas están implementados en la web del grupo y en la spin-off Prometeo.

Características principales:

- MeteoLab permite cargar las observaciones y los campos de la cuadrícula de ACM en una forma fácil.
- Gráfica que muestra la serie temporal, patrones espaciales, etc
- Análisis de Componentes Principales (EOF y visualización PCs).
- Clustering (k-means, SOM).
- Análisis de Correlación Canónica.
- Regresión y modelos de series de tiempo.
- Generadores tiempo.
- Downscaling estadístico (pronóstico local con análogos y NN k-).
- Validación probabilístico (puntuaciones de habilidad, curvas ROC, etc.)
- Aplicaciones de las redes neuronales (utilizando NETlab, por Ian Nabney y el obispo Christopher).
- Aplicaciones de las redes bayesianas (utilizando BNT, por Kevin Murphy).

### 6.3 Técnicas para series temporales.

Una **serie temporal** es una secuencia de datos, observaciones o valores, medidos en determinados momentos del tiempo, ordenados cronológicamente y, normalmente, espaciados entre sí de manera uniforme.

El **análisis de series temporales** comprende métodos que ayudan a interpretar este tipo de datos, extrayendo información representativa, tanto referente a los orígenes o relaciones subyacentes como a la posibilidad de extrapolar y predecir su comportamiento futuro.

De hecho, uno de los usos más habituales de las series de datos temporales es su análisis para predicción y pronóstico. Por ejemplo de los datos climáticos, de las acciones de bolsa, o las series pluviométricas. Resulta difícil imaginar una rama de las ciencias en la que no aparezcan datos que puedan ser considerados como series temporales. Son estudiadas en estadística, procesamiento de señales, econometría y muchas otras áreas.

#### Introducción

Las series temporales se usan para estudiar la relación causal entre diversas variables que cambian con el tiempo y se influyen entre sí. Desde el punto de vista probabilístico una serie temporal es una sucesión de variables aleatorias indexadas según parámetro creciente con el tiempo. Cuando la esperanza matemática de dichas variables aleatorias no es constante, ni varía de manera cíclica, se dice que la serie no es estacionaria y presenta una tendencia secular.

#### Componentes

El análisis más clásico de las series temporales se basa en la suposición de que los valores que toma la variable de observación es la consecuencia de cuatro componentes, cuya actuación conjunta da como resultado los valores medidos, estos componentes son:

1. **Tendencia secular** o regular, indica la marcha general y persistente del fenómeno observado, es una componente de la serie que refleja la evolución a largo plazo. Por ejemplo, la tendencia creciente del índice de reciclado de basuras en los países desarrollados, o el uso creciente de Internet en la sociedad, independientemente de que en un mes concreto en un país, por determinadas causas, haya una baja en la utilización de Internet.
2. **Variación estacional.** Es el movimiento periódico de corto plazo. Se trata de una componente causal debida a la influencia de ciertos fenómenos que se repiten de manera periódica en un año (las estaciones), una semana (los fines de semana) o un día (las horas puntas) o cualquier otro periodo. Recoge las oscilaciones que se producen en esos períodos de repetición.
3. **Variación cíclica.** Es el componente de la serie que recoge las oscilaciones periódicas de amplitud superior a un año. movimientos normalmente irregulares alrededor de la tendencia, en las que a diferencia de las variaciones estacionales, tiene un período y amplitud variables, pudiendo clasificarse como cíclicos, cuasicíclicos o recurrentes.
4. **Variación aleatoria,** accidental, de carácter errático, también denominada **residuo**, no muestran ninguna regularidad, debidos a fenómenos de carácter ocasional como pueden ser tormentas, terremotos, inundaciones, huelgas, guerras, avances tecnológicos, etc.

## Notación

Existen diferentes notaciones empleadas para la representación matemática de una serie temporal:

$$X = \{X_1, X_2, \dots\} \text{ o } \{X_t\}_{t \geq 1}$$

Ésta es una de las comunes que representa un Serie de Tiempo  $X$  que es indexada por números naturales. También estamos acostumbrados a ver:

$$Y = \{Y_t : t \in T\}$$

## Herramientas

Herramientas para la investigación de series temporales:

- Consideración de la función de autocorrelación y la función de densidad espectral.
- Efectuando una transformación de Fourier para investigar las series en el dominio de la frecuencia.
- El uso de un filtro para remover ruido no deseado.
- Análisis de los componentes principales (o análisis de la función ortogonal empírica).
- Redes neuronales artificiales.
- Técnicas de análisis de frecuencias de tiempo:
  - Transformación de ondas continuas
  - Transformación de Fourier de corto tiempo
  - Transformación chirplet
  - Transformación de Fourier fraccionaria
- Análisis caótico
  - Dimensión de correlación
  - Gráficos de repetición
  - Análisis de cuantificación repetitiva
  - Exponentes de Lyapunov

En el caso de las ciencias medioambientales, los métodos de análisis de series temporales, son muy apreciados y usados, una vez más la toolbox `MeteoLab`, nos permite hacer los cálculos pertinentes.

#### *6.4 Correlación canónica.*

El **análisis de correlación canónica** es un método de análisis multivariante desarrollado por Harold Hotelling. Su objetivo es buscar las relaciones que pueda haber entre dos grupos de variables y la validez de las mismas. Se diferencia del análisis de correlación múltiple en que éste sólo predice una variable dependiente a partir de múltiples independientes, mientras que la correlación canónica predice múltiples variables dependientes a partir de múltiples independientes. La correlación hipercanónica es una correlación lineal y, por tanto, sólo busca relaciones lineales entre las variables.

Al diseñar el experimento hay que considerar el tamaño de la muestra ya que son necesarias un mínimo de observaciones por variable, para que el análisis pueda representar las correlaciones adecuadamente.

Finalmente, hay que interpretar las cargas canónicas para determinar la importancia de cada variable en la función canónica. Las cargas canónicas reflejan la varianza que la variable observada comparte con el valor teórico canónico.

#### *6.5 Técnica de locales análogos.*

El método de downscaling mediante análogos introducido por Lorenz (1969) es una técnica estadística de downscaling sencilla y potente que asume que circulaciones de flujo similares, o patrones atmosféricos, sobre una determinada región producen resultados similares a nivel local. Siguiendo este esquema, los valores locales de la ODN para un día en particular se pueden estimar de las ocurrencias registradas en un conjunto de días históricos similares, obtenidos a partir de la similitud de los patrones atmosféricos correspondientes. En general, se ha demostrado que el método de análogos funciona igual de bien que otras técnicas de downscaling más complicadas (ver por ejemplo Zorita and von Storch, 1999).

Las distintas aplicaciones que se pueden encontrar en la literatura científica del método de análogos difieren básicamente en el patrón atmosférico, la medida de analogía o similitud, el número de análogos considerado y el método utilizado para deducir un valor local del conjunto resultante de análogos.

En este trabajo, para realizar el downscaling de los valores locales de ODN para las 33 estaciones en un día en particular, representado por el patrón atmosférico se calculó en primer lugar un conjunto de 30 días análogos (aquellos días históricos con un patrón más parecido al del día en cuestión) utilizando la métrica Euclídea estándar. La ocurrencia de nieve en cada estación se estimó mediante la media de las 30 observaciones históricas correspondientes.

Para analizar la capacidad del método de análogos para inferir la ocurrencia diaria de nieve, se realizó una validación cruzada eliminando un elemento de la muestra cada vez para el periodo 1957-2002 (Stone, 1974). Para ello se utilizó un único día/observación del periodo como dato de validación y las observaciones restantes como conjunto de entrenamiento. Por tanto, los patrones de ERA40 (3.1) para este periodo, excluyendo una ventana temporal de un mes en torno al día de validación, se utilizan como base de datos histórica para la búsqueda de análogos y el método de análogos antes descrito se utilizó para obtener una estimación probabilística para la ODN; la ventana de exclusión de un mes impide que la validación pueda presentar problemas de sobreajuste debido a la persistencia o a la correlación serial (también

se probaron ventanas mayores con resultados similares). Este proceso se repitió para cada uno de los días, de manera que cada observación se utilizó una vez como dato de validación.

### 6.6 Curvas ROC.

En la Teoría de detección de señales una **curva ROC** (acrónimo de **Receiver Operating Characteristic**, o Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a  $(1 - \text{especificidad})$  para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo). ROC también puede significar Relative Operating Characteristic (Característica Operativa Relativa) porque es una comparación de dos características operativas (VPR y FPR) según cambiamos el umbral para la decisión (3). No se suele utilizar ROC aislado, debemos decir "curva ROC" o "análisis ROC". Sobre la historia del acrónimo ROC consultar Swets (1996).

El análisis de la curva ROC, o simplemente análisis ROC, proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos subóptimos independientemente de (y antes de especificar) el coste de la distribución de las dos clases sobre las que se decide. La curva ROC es también independiente de la distribución de las clases en la población (en diagnóstico, la prevalencia de una enfermedad en la población). El análisis ROC se relaciona de forma directa y natural con el análisis de coste/beneficio en toma de decisiones diagnósticas.

La curva ROC se desarrolló por ingenieros eléctricos para medir la eficacia en la detección de objetos enemigos en campos de batalla mediante pantallas de radar, a partir de lo cual se desarrolla la Teoría de Detección de Señales (TDS). El análisis ROC se aplicó posteriormente en medicina, radiología, psicología y otras áreas durante varias décadas. Sólo recientemente ha encontrado aplicación en áreas como aprendizaje automático (o machine learning en inglés), y minería de datos (data mining en inglés).

### Conceptos Básicos

Un modelo de clasificación (clasificador o Clasificadores (matemático) o Diagnóstico) es una función que permite decidir cuáles de un conjunto de instancias están relacionadas o no por pertenecer a un mismo tipo o clase. El resultado del clasificador o del diagnóstico puede ser un número real (valor continuo), en cuyo caso el límite del clasificador entre cada clase debe determinarse por un valor umbral (por ejemplo para determinar si una persona tiene hipertensión basándonos en una medida de presión arterial), o puede ser un resultado discreto que indica directamente una de las clases.

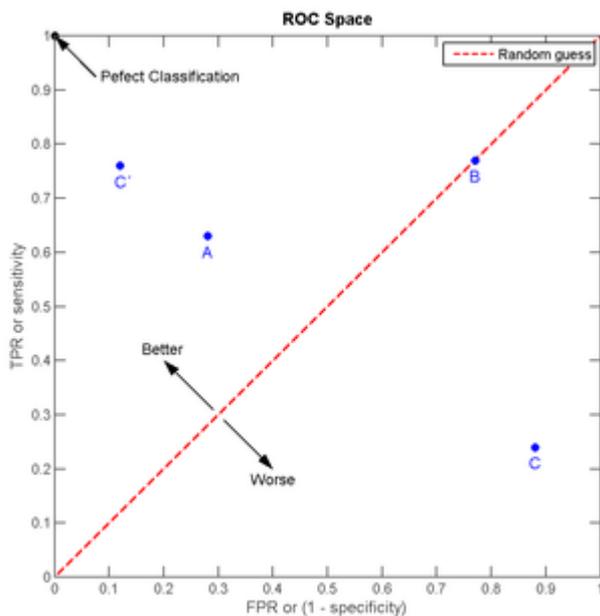
Consideremos un problema de predicción de clases binario, en la que los resultados se etiquetan positivos (p) o negativos (n). Hay cuatro posibles resultados a partir de un clasificador binario como el propuesto. Si el resultado de una exploración es p y el valor dado es también p, entonces se conoce como un Verdadero Positivo (VP); sin embargo si el valor real es n entonces se conoce como un Falso Positivo (FP). De igual modo, tenemos un Verdadero Negativo (VN) cuando tanto la exploración como el valor dado son n, y un Falso Negativo (FN) cuando el resultado de la predicción es n pero el valor real es p. Un ejemplo aproximado de un problema real es el siguiente: consideremos una prueba diagnóstica que persiga determinar si una persona tiene una cierta enfermedad. Un falso positivo en este caso ocurre cuando la prueba predice que el resultado es positivo, cuando la persona no tiene

realmente la enfermedad. Un falso negativo, por el contrario, ocurre cuando el resultado de la prueba es negativo, sugiriendo que no tiene la enfermedad cuando realmente sí la tiene.

Definamos un experimento a partir de  $P$  instancias positivas y  $N$  negativas. Los cuatro posibles resultados se pueden formular en una Tabla de contingencia (o Matriz de confusión)  $2 \times 2$  como sigue:

		Valor en la realidad		total
		$p$	$n$	
Predicción outcome	$p'$	Verdaderos Positivos	Falsos Positivos	$P'$
	$n'$	Falsos Negativos	Verdaderos Negativos	$N'$
total		$P$	$N$	

### El espacio ROC



El espacio ROC y las parcelas de los cuatro ejemplos de predicción A, B, C y C'.

La tabla de contingencia puede proporcionar varias medidas de evaluación (ver caja de terminología). Para dibujar una curva ROC sólo son necesarias las razones de Verdaderos Positivos (VPR) y de falsos positivos (FPR). La VPR mide hasta qué punto un clasificador o prueba diagnóstica es capaz de detectar o clasificar los casos positivos correctamente, de entre

todos los casos positivos disponibles durante la prueba. La FPR define cuántos resultados positivos son incorrectos de entre todos los casos negativos disponibles durante la prueba.

Un espacio ROC se define por FPR y VPR como ejes x e y respectivamente, y representa los intercambios entre verdaderos positivos (en principio, beneficios) y falsos positivos (en principio, costes). Dado que VPR es equivalente a sensibilidad y FPR es igual a 1-especificidad, el gráfico ROC también es conocido como la representación de sensibilidad frente a (1-especificidad). Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC.

El mejor método posible de predicción se situaría en un punto en la esquina superior izquierda, o coordenada (0,1) del espacio ROC, representando un 100% de sensibilidad (ningún falso negativo) y un 100% también de especificidad (ningún falso positivo). A este punto (0,1) también se le llama una *clasificación perfecta*. Por el contrario, una clasificación totalmente aleatoria (o adivinación aleatoria) daría un punto a lo largo de la línea diagonal, que se llama también *línea de no-discriminación*, desde el extremo inferior izquierdo hasta la esquina superior derecha (independientemente de los tipos de base positivas y negativas). Un ejemplo típico de adivinación aleatoria sería decidir a partir de los resultados de lanzar una moneda al aire, a medida que el tamaño de la muestra aumenta, el punto de un clasificador aleatorio de ROC se desplazará hacia la posición (0.5, 0.5).

La diagonal divide el espacio ROC. Los puntos por encima de la diagonal representan los buenos resultados de clasificación (mejor que el azar), puntos por debajo de la línea de los resultados pobres (peor que al azar). Nótese que la salida de un predictor consistentemente pobre simplemente podría ser invertida para obtener un buen predictor.

Cuanto más cerca esté un método de la esquina superior izquierda (clasificación perfecta) mejor será, pero lo que en realidad marca el poder predictivo de un método es la distancia de este a la línea de estimación aleatoria, da igual si por arriba o por abajo.

### **Curvas en el espacio ROC**

Los clasificadores discretos, como los Árbol de decisión o los sistemas de reglas, dan como resultados a valores numéricos una etiqueta binaria. Cuando se usan estos clasificadores con un conjunto concreto de instancias para clasificar o predecir, el rendimiento del clasificador proporciona un único punto en el espacio ROC. Para otros clasificadores, como un Clasificador bayesiano o una Red neuronal artificial, la salida son valores de probabilidad que representan hasta qué punto una instancia pertenece a una de las dos clases.

Para estos métodos se debe fijar un valor umbral que determinará un punto en el espacio ROC. Por ejemplo, si ante una determinada magnitud fijamos ese umbral en 0.8, la probabilidad de las instancias iguales o superiores serán predichas como positivas, y los valores por debajo serán predichos como negativos. Por tanto podremos calcular una tabla de contingencia (o matriz de confusión) para ese umbral de 0.8, y encontrar el punto correspondiente en el espacio ROC. Según vamos variando el umbral (por ejemplo, en pasos de 0.1) tendríamos una tabla de contingencia y un nuevo punto en el espacio ROC. Dibujar la curva ROC consiste en poner juntos todos los puntos correspondientes a todos los umbrales o puntos de corte, de tal modo que ese conjunto de puntos se parecerá más o menos a una curva en el espacio cuadrado entre (0,0) y (1,1). Dependiendo del tipo de modelo la curva se parecerá más a una escalera (métodos no paramétricos) o una verdadera curva (métodos paramétricos).<sup>3</sup> A medida que desplazamos ese valor umbral, en realidad estamos alterando las tasas de verdaderos positivos (VP) y falsos positivos (FP).

## Interpretación de una curva ROC

La curva ROC se puede usar para generar estadísticos que resumen el rendimiento (o la efectividad, en su más amplio sentido) del clasificador. A continuación se proporcionan algunos:

- El punto de inserción de la curva ROC con la línea convexa a la línea de discriminación.
- El área entre la curva ROC y la línea de convexo-paralela discriminación.
- El área bajo la curva ROC, llamada comúnmente AUC (*Área Bajo la Curva*). También se puede encontrar denominada A' ("a-prima"),<sup>4</sup> o el estadístico 'c' (*c-statistic*).<sup>5</sup>
- Índice de sensibilidad o d' (*d-prima*, por cierto siempre minúscula). Es la distancia entre la media de la distribución de actividad en el sistema bajo condiciones de sólo ruido y su distribución bajo condiciones de sólo señal, dividido por su desviación típica, bajo el supuesto de que ambas distribuciones son normales con la misma desviación típica. Bajo estos supuestos, se puede probar que la forma de la curva ROC sólo depende de este parámetro d'.

El indicador más utilizado en muchos contextos es el área bajo la curva ROC o AUC. Este índice se puede interpretar como la probabilidad de que un clasificador ordenará o puntuará una instancia positiva elegida aleatoriamente más alta que una negativa. Se puede demostrar que el área bajo la curva ROC es equivalente a la Prueba de Mann-Whitney, una prueba paramétrica aplicada a dos muestras independientes, cuyos datos han sido medidos al menos en una escala de nivel ordinal. Se trata de una prueba estadística virtualmente idéntica a la realización de una prueba paramétrica ordinaria T de dos muestras en los datos después de haber ordenado las muestras combinadas. Es también equivalente a la Prueba de los signos de Wilcoxon.

También se ha demostrado la relación del área bajo la curva ROC con el Coeficiente de Gini, con la siguiente fórmula  $G_1 + 1 = 2 \times \text{AUC}$ , donde:

$$G_1 = 1 - \sum_{k=1}^n (X_k - X_{k-1})(Y_k + Y_{k-1})$$

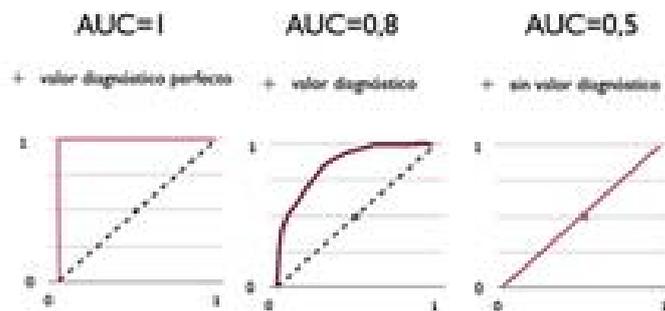
Otra forma básica de calcular AUC es usando un promedio de una serie de aproximaciones trapezoidales.

Sin embargo, se ha comentado que este indicador, en general, reducir la curva ROC en varios metros, hace perder información sobre el patrón de intercambios del algoritmo discriminador en cuestión.

La comunidad de aprendizaje automático utiliza el estadístico AUC para la comparación de modelos. En otras áreas de ingeniería se prefiere la medida del área entre la curva ROC y la línea de no-discriminación.

En ocasiones puede ser más útil mirar a una región específica de la curva ROC más que a toda la curva. Es posible calcular áreas parciales bajo la curva, o AUC parciales. Por ejemplo, nos podríamos concentrar en la región de la curva con razones de falsos positivos más bajas, que es a menudo el interés principal de las pruebas de Detección precoz (o medicine-screening) en la población.

## Curvas ROC para pruebas diagnósticas



### Diferentes curvas ROC

Para la elección entre dos pruebas diagnósticas distintas, se recurre a las curvas ROC, ya que es una medida global e independiente del punto de corte. Por esto, en el ámbito sanitario, las curvas ROC también se denominan **curvas de rendimiento diagnóstico**.

La elección se realiza mediante la comparación del área bajo la curva (AUC) de ambas pruebas. Esta área posee un valor comprendido entre 0,5 y 1, donde 1 representa un valor diagnóstico perfecto y 0,5 es una prueba sin capacidad discriminadora diagnóstica. Es decir, si AUC para una prueba diagnóstica es 0,8 significa que existe un 80% de probabilidad de que el diagnóstico realizado a un enfermo sea más correcto que el de una persona sana escogida al azar. Por esto, siempre se elige la prueba diagnóstica que presente un mayor área bajo la curva.

A modo de guía para interpretar las curvas ROC se han establecido los siguientes intervalos para los valores de AUC:

[0.5, 0.6): Test malo.

[0.6, 0.75): Test regular.

[0.75, 0.9): Test bueno.

[0.9, 0.97): Test muy bueno.

[0.97, 1): Test excelente.

### Historia

La curva ROC se comenzó a utilizar durante la Segunda Guerra Mundial para el análisis de señales de radar, a partir de lo cual se desarrolló la Teoría de Detección de Señales. Después del ataque a Pearl Harbor en 1941, el ejército de los Estados Unidos comenzó un programa de investigación para detectar correctamente los aparatos japoneses a partir de sus señales de radar.

En los años 50, las curvas ROC se utilizaron en Psicofísica para evaluar la capacidad de detección de humanos (y también de no humanos) en señales débiles. En medicina el análisis ROC se ha utilizado de forma muy extensa en epidemiología e investigación médica, de tal modo que se encuentra muy relacionado con la Medicina basada en la evidencia. En

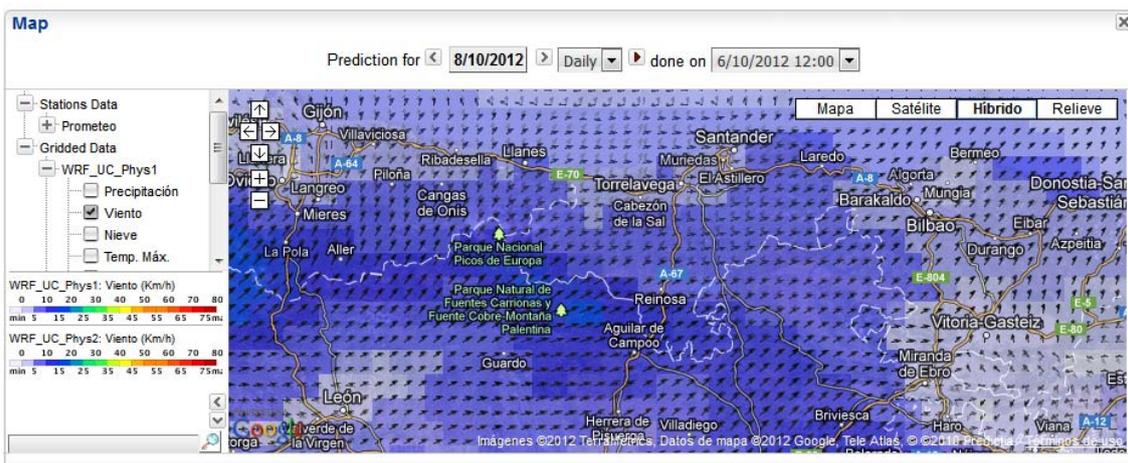
Radiología, el análisis ROC es la técnica de preferencia para evaluar nuevas técnicas de diagnóstico por imagen.

### 6.7 Simulación de las tendencias de viento a tres periodos distintos

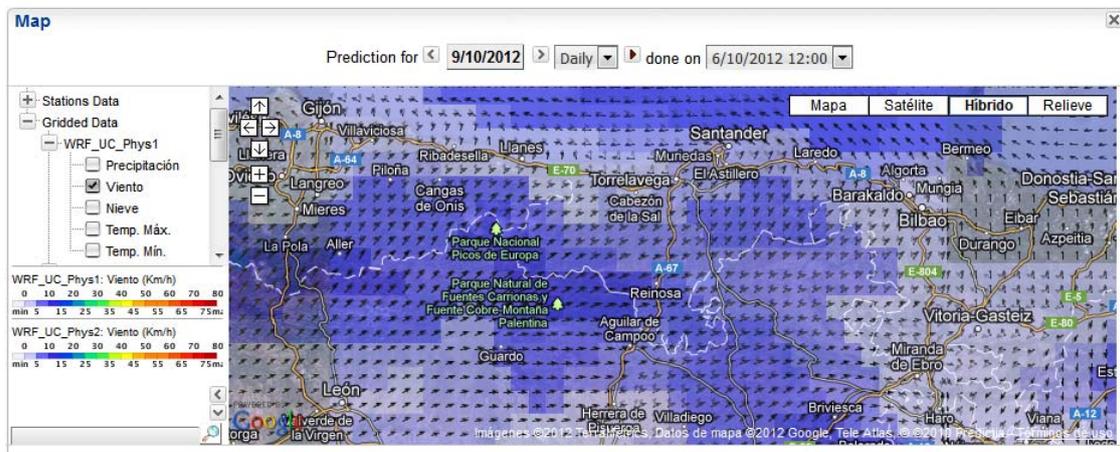
Vamos a mostrar las predicciones en la velocidad del viento, para una determinada región de Cantabria, en este caso es la Centro Sur, de acuerdo con los métodos de downscaling estadístico, que hemos preparado, usando la toolbox, podemos ver cuál será la predicción de viento para la próxima semana, vamos a dar la información:

El sistema de predicción está basado en dos aspectos fundamentales, **WRF-UC multi-física** para la Península Ibérica a **9km** (predicción numérica) y predicciones locales de **PROMETEO** (predicción estadística) basada en Downscaling Estadístico.

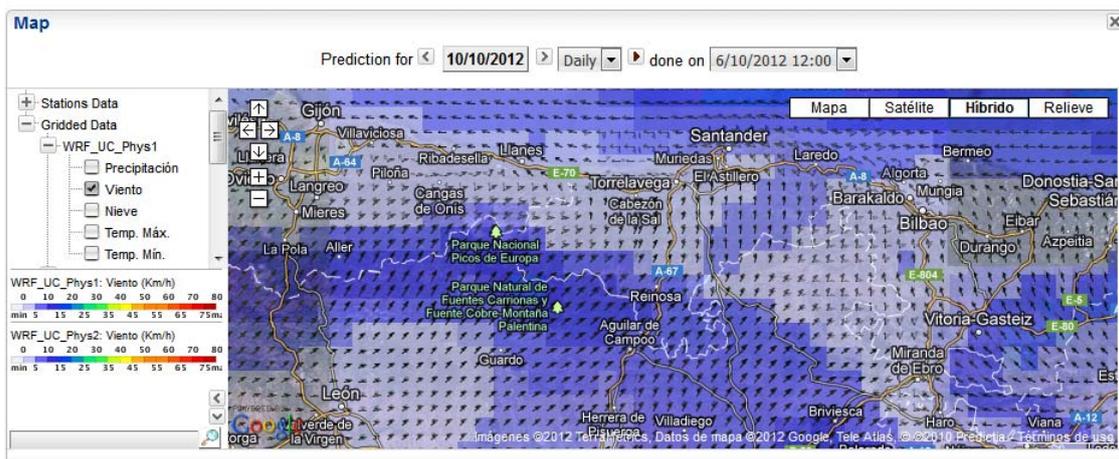
Predicción de viento para el próximo día 8 de Octubre de 2012:



Predicción de viento para el próximo día 9 de Octubre de 2012:

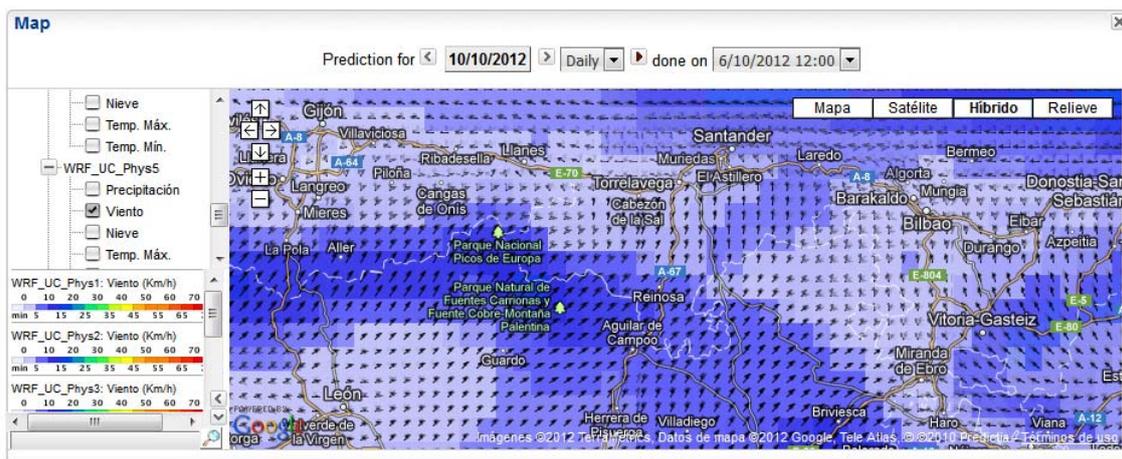


Predicción de viento para el próximo día 10 de Octubre de 2012:



Todas las predicciones se han realizado, en base a cálculos del día 6 de Octubre de 2012, por una parte tenemos la predicción numérica, del grupo de trabajo de Unican, WRF\_UC\_Phys y por otra parte tenemos los datos estadísticos, obtenidos con Meteolab, implementados por Prometeo.

Vamos a poner una visión global, de los datos de cálculo numérico, de WRF\_UC\_Phys1, WRF\_UC\_Phys2, WRF\_UC\_Phys3, y WRF\_UC\_Phys4.

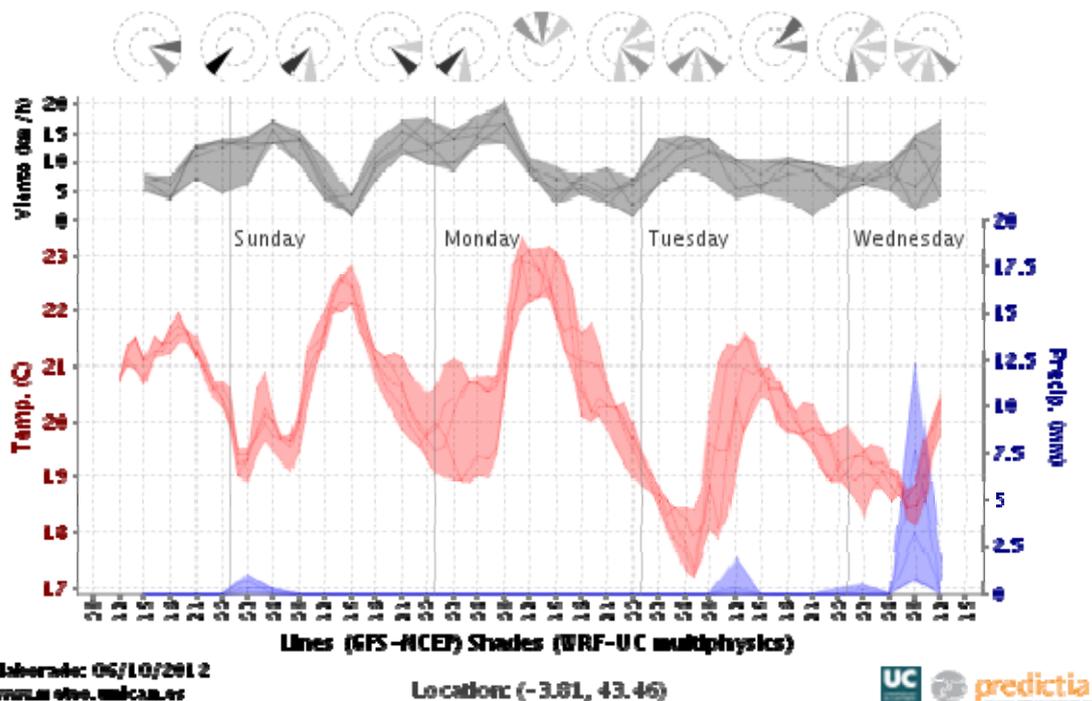


Por otra parte, los datos obtenidos con la toolbox, Meteolab, quedan implementados, de la siguiente forma, para las variables, que a continuación se detallan y la estación de Santander:

Locations			
SANTANDER CENTRO			
Realization date: 06/10/2012			
Variable	07/10/2012	08/10/2012	09/10/2012
Precipitación	0 mm	0.6 mm	1.6 mm
Temp. Máx.	24 °C	24 °C	23 °C
Temp. Mín.	17 °C	17 °C	17 °C
Granizo (%)	0%	0%	0%
Insolación	76 %	80 %	56 %
Niebla (%)	4%	33%	20%
Nieve (%)	0%	0%	0%
Tormenta (%)	0%	0%	0%

Por otra parte, para la estación de Santander, aplicando datos registrados en esta estación hasta 2001, podemos calcular las probabilidades conjuntas, de que se produzcan varios sucesos a la vez, esto ya se indicó en la parte superior, en las redes Bayesianas.

Vamos a añadir, también, un Meteograma, de la evolución prevista, para los días sucesivos:



Es interesante, comparar estos datos, con los que nos proporciona AEMET, dados por el sistema HIRLAM, para los días de observación:

Se muestran los mapas previstos de los modelos numéricos HIRLAM-AEMET Y CEPPM: En la pestaña HIRLAM-AEMET 0.16° según el nivel seleccionado se accede a los siguientes parámetros:

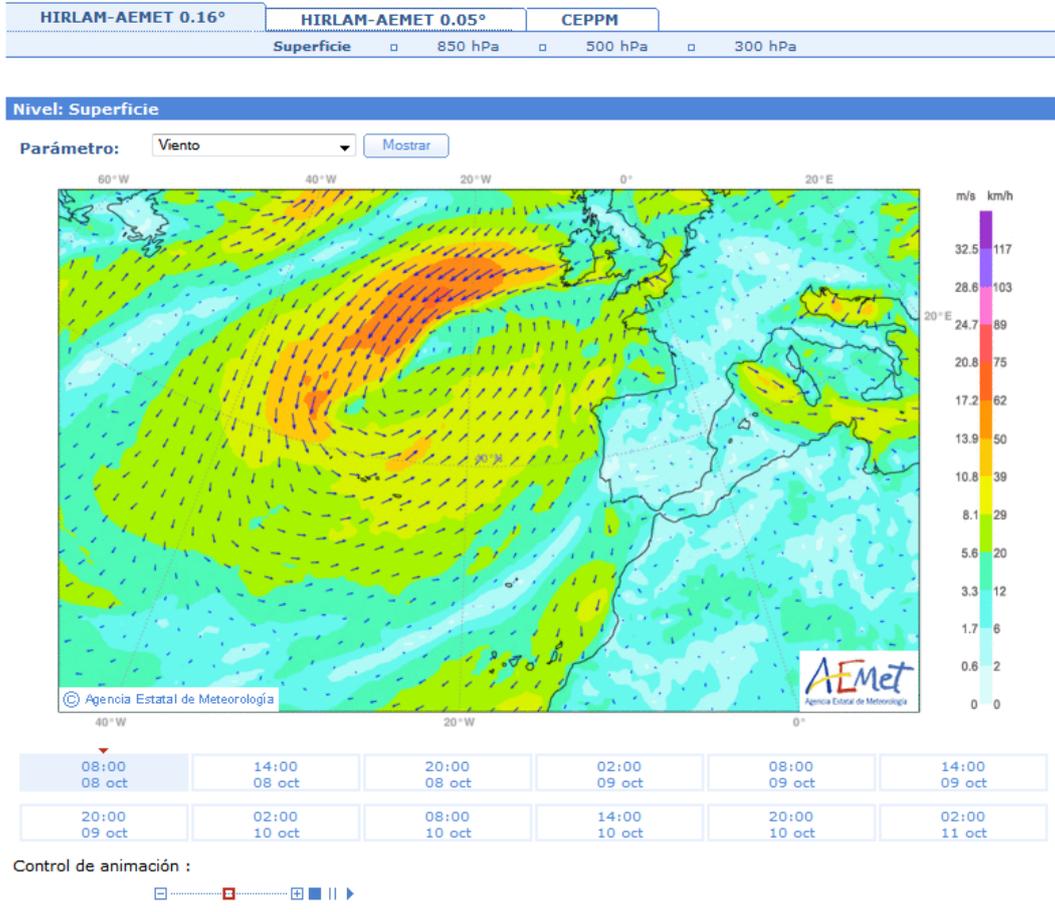
- Superficie: presión, precipitación, viento, nubosidad y temperatura
- 850 hPa y 500 hPa: temperatura y geopotencial
- 300 hPa: viento y geopotencial

En la pestaña HIRLAM-AEMET 0.05° se selecciona por C. Autónomas los parámetros de superficie: temperatura, viento y precipitación.

En la pestaña CEPPM podrá seleccionar entre los parámetros de presión en superficie y geopotencial de 500 hPa para tres zonas del planeta: Atlántico Norte, Hemisferio Norte y Hemisferio Sur.

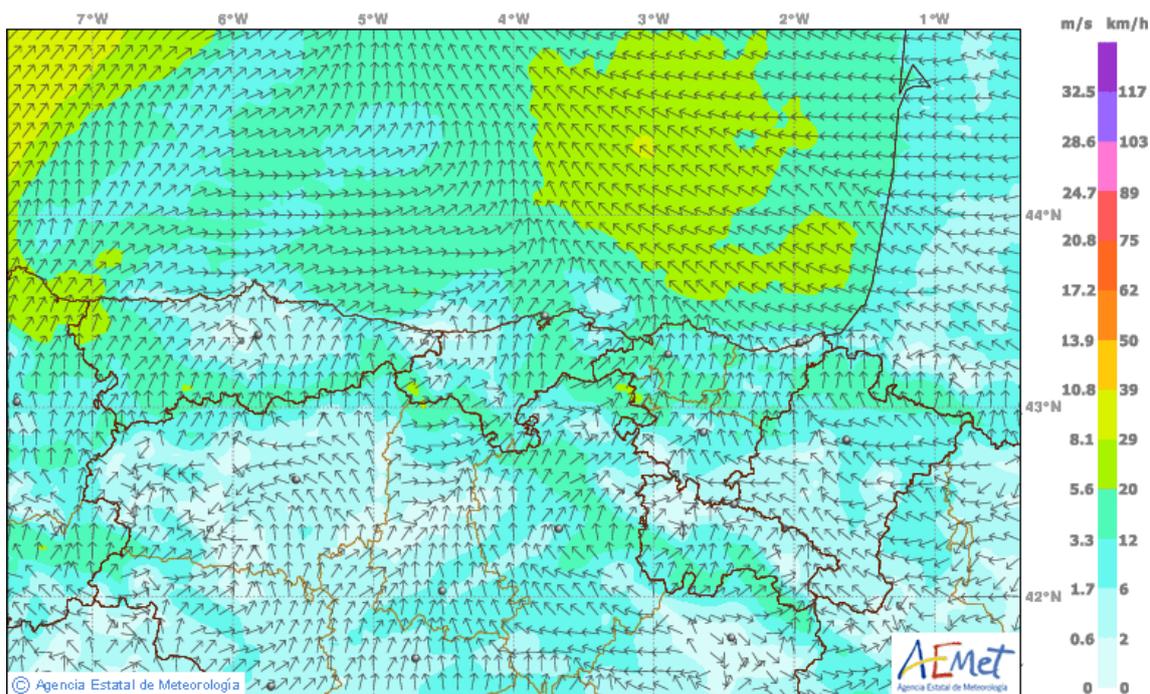
Alcances: los tres primeros días, a intervalos de 6 horas, para HIRLAM-AEMET 0.16°, y los tres días siguientes, a intervalos de 24 horas, del modelo CEPPM. Para HIRLAM-AEMET 0.05° día y medio, a intervalos de 3 horas. Además de a los mapas individuales se puede acceder a una animación de los mismos.

### Modelos numéricos

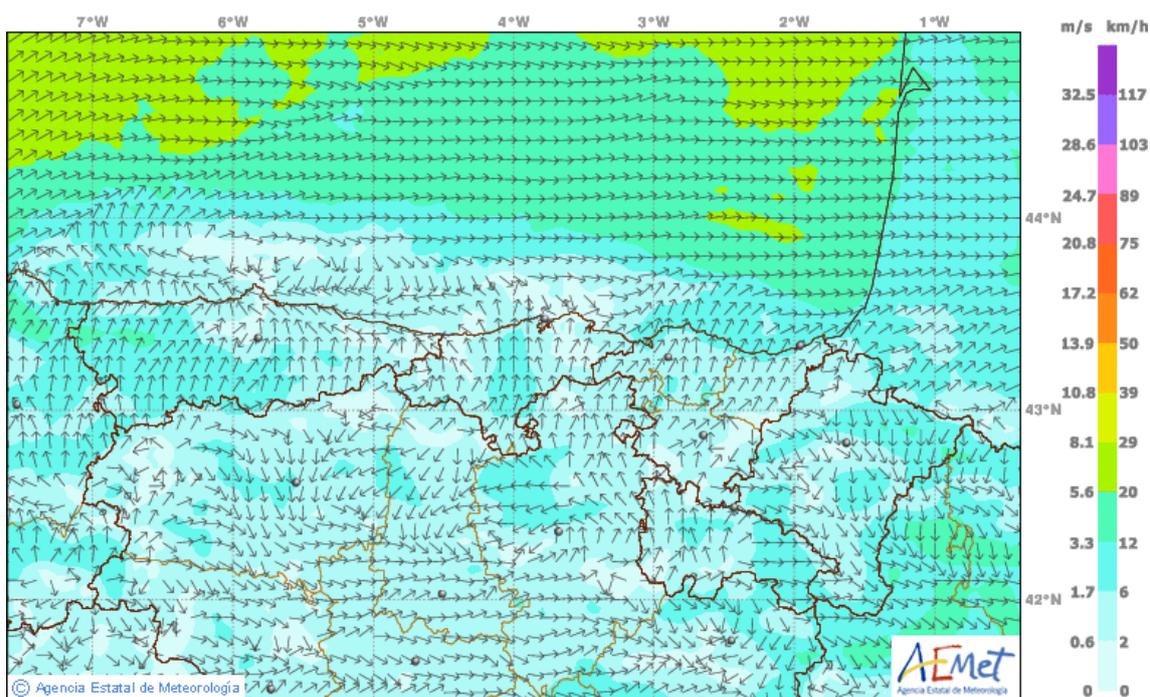


Para estos casos, es decir la predicción de tiempo en Cantabria, solamente tenemos acceso, ala predicción de los dos días posteriores, al la fecha actual de observación.

Predicción de viento para el próximo día 8 de Octubre de 2012:

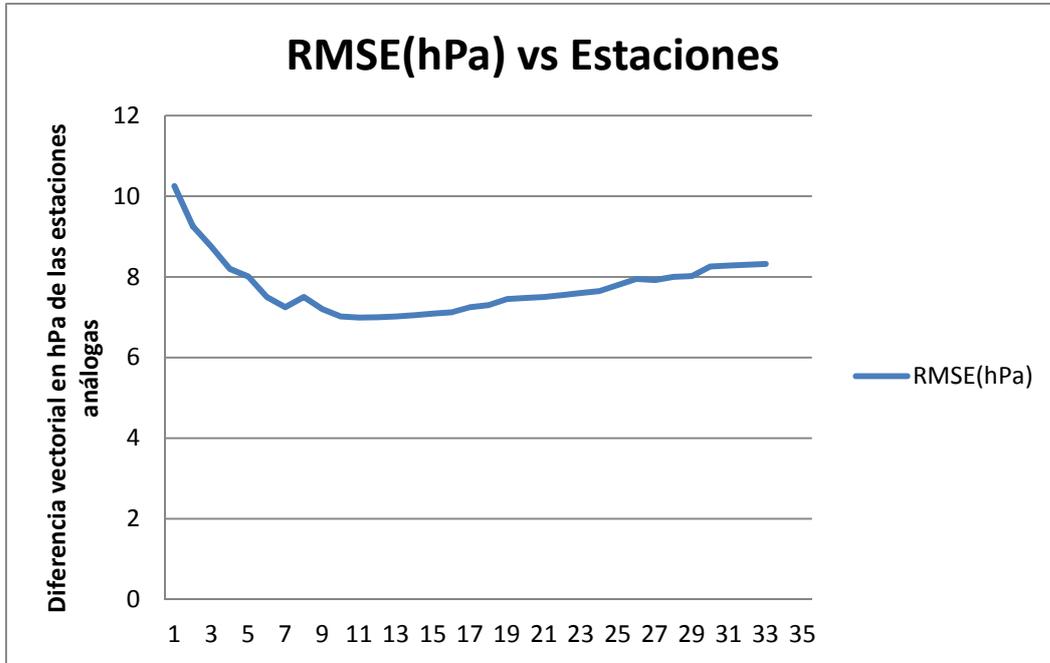


Predicción para el próximo 9 de Octubre de 2012:



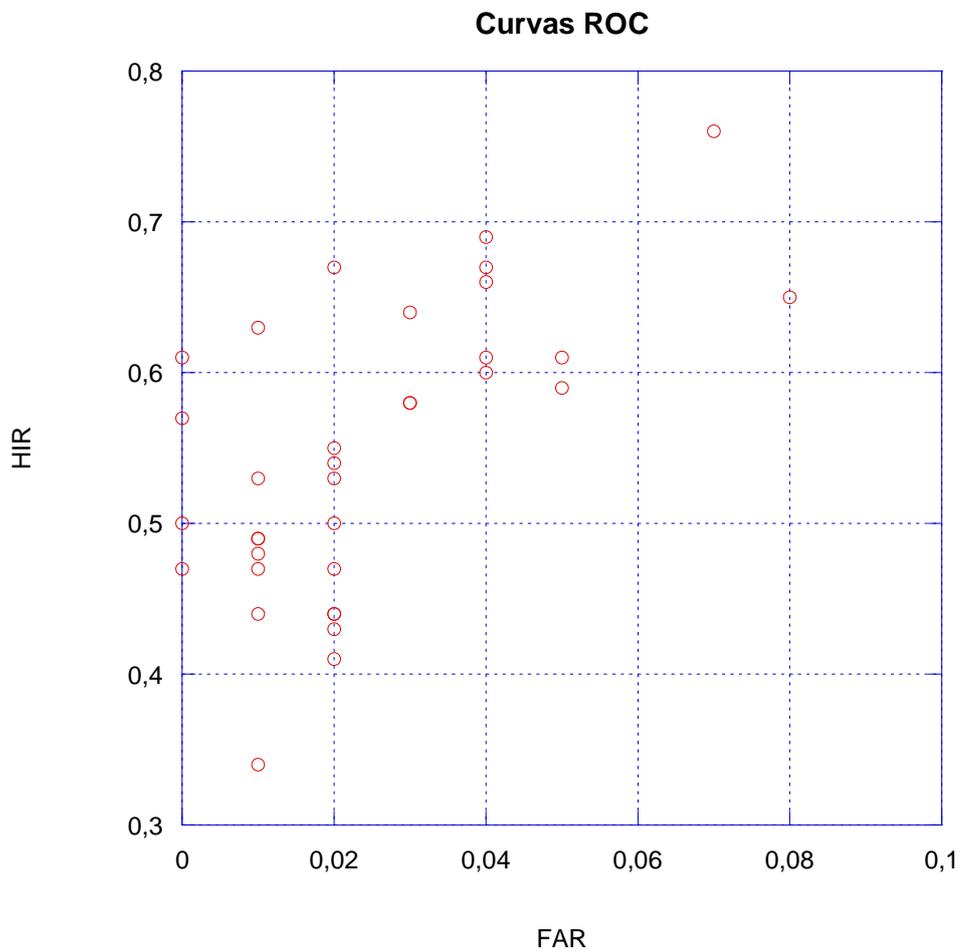
### Resultados de Análogos:

Para los patrones, obtenidos, para cada una de las 33 estaciones en estudio, se ha utilizado la técnica de análogos, para buscar situaciones similares, en los días incluidos en el histórico de datos anteriormente descrito. Se han obtenido los siguientes datos para el valor de la racha media diaria y la velocidad media diaria del viento, para las 33 estaciones:



Se puede observar, como el RMSE, decrece rápidamente con las primeras estaciones, pero a medida que vamos añadiendo más datos a la serie histórica, podemos concluir, que el valor se estabiliza, en el entorno de 8 hPa.

Representación de las curvas ROC:

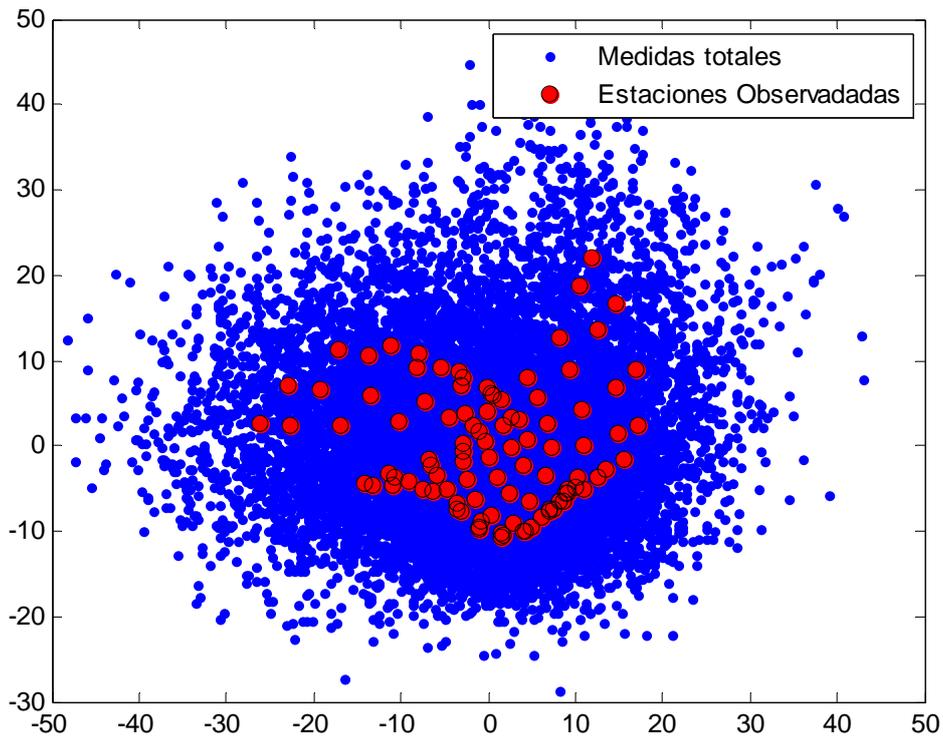


En estas curvas ROC, tal como se explicó en el apartados anterior, se han incluido cada una de las 33 estaciones, anteriormente descritas, vemos que para las primeras estaciones nos encontramos cerca del límite inferior de la caja y a medida, que nos vamos hacia la parte superior, podemos observar que nos vamos hacia la parte superior derecha de la caja, con lo que los resultados, sin ser muy buenos, son aceptables.

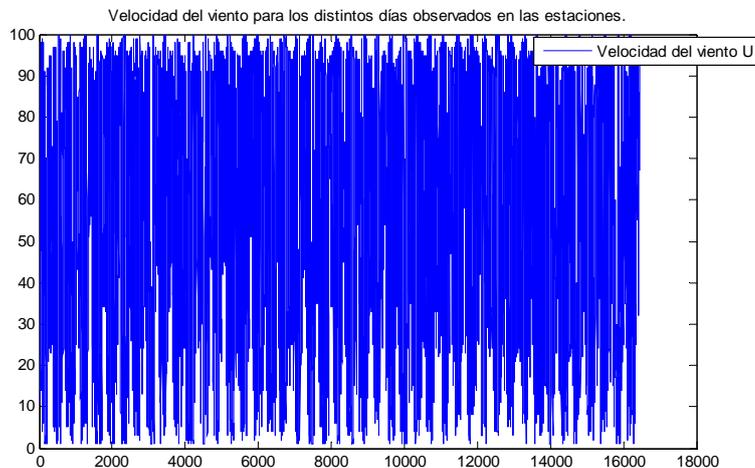
Otros datos analizados:

Dada la versatilidad de la Toolbox, podemos sacar un par de datos más, relacionados con **clustering**, es decir un agrupamiento de las estaciones, de forma, que podamos tener los datos de las estaciones, agrupados por grupos y obtener también una salida de la serie temporal de los datos:

## CLUSTERING



## SERIE TEMPORAL DE DATOS:



### 6.8 Validez del método.

El método es una buena herramienta alternativa, de predicción del tiempo, alternativa a los modelos numéricos utilizados en las distintas agencias Meteorológicas del Mundo, no obstante, creemos que se puede pulir mas, añadiendo más datos a su datasheet, así como la implementación de otras, técnicas de medida, como pueden ser las wavelets, o el análisis de componentes principales, que no se trata en este trabajo.

## 7. Conclusiones

Este trabajo, ha sido bastante interesante, en cuanto a la implementación de Métodos de **Downscaling Estadístico**, que nos han permitido ver una alternativa en el problema de la predicción de variables meteorológicas, alternativo, a la resolución numérica de las ecuaciones.

Por otra parte, y para trabajos futuros se propone, la aplicación de estas técnicas a otro tipo de series temporales, como puede ser , la Bolsa y Mercados Financieros, Sanitaria, etc..

## 8. Bibliografía

Brunet, M., Jones, P. D., Sigró, J., Saladié, O., Aguilar, E., Moberg, A., Della-Marta, P. M., Lister, D., Walther, A., and López, D. (2007). *Temporal and spatial temperature variability and change over Spain during 1850–2005*. *J. Geophys. Res.*, 112:D12117, doi:10.1029/2006JD008249.

**Hsieh, W. W. (2004). Nonlinear multivariate and time series analysis by neural network methods. *Reviews of Geophysics*, 42:1003.**

Gutiérrez, J., Ancell, R., Cofiño, A., and Sordo, C., eds. (2004a). *Redes Neuronales y Probabilísticas en las Ciencias Atmosféricas*. Monografías del Instituto Nacional de Meteorología. Ministerio de Medio Ambiente, Spain.

Gutiérrez, J. M., Cofiño, A. S., Cano, R., and Primo, C. (2005). Analysis and downscaling multi-model seasonal forecasts using self-organizing maps. *Tellus A*, 57:435447.

Gutiérrez, J. M., Cofiño, A. S., Cano, R., and Rodríguez, M. A. (2004b). Clustering methods for statistical downscaling in short-range weather forecast. *Monthly Weather Review*, 132(9):2169–2183.

**Jolliffe, I. T. and Stephenson, D. B., eds. (2003). Forecast Verification: : A Practitioner's Guide in Atmospheric Science. John Wiley and Sons.**

**Lorenz, E. (1969). Atmospheric predictability as revealed by naturally occurring analogues. *Journal of the Atmospheric sciences*, 26:636–646.**

**Martin, E., Timbal, B., and Brun, E. (1997). Downscaling of general circulation model outputs: Simulation of the snow climatology of the French Alps and sensitivity to climate change. *Climate Dynamics*, 13:45–56.**

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., and Mearns, L. O. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting material of the IPCC (Task Group on Data and Scenario Support for Impacts and Climate Analysis, TGICA)*.

Zorita, E. and von Storch, H. (1999). The analog method as a simple statistical downscaling technique: comparison with more complicated methods. *Journal of Climate*, 12:2474–2489.

Webs de interés:

<http://www.aemet.es>

<http://www.meteo.unican.es/>

<http://www.predictia.es/>

<http://www.noaa.gov>

<http://www.ecmwf.int/>

<http://www.ecmwf.int/>