

Trabajo Fin de Grado

Grado Ingeniería Aeroespacial

Aplicación del post-procesado estadístico de predicciones meteorológicas por conjuntos a la predicción de trayectoria de aeronaves

Autor: Manuel Gutiérrez Fernández

Tutor: Antonio Franco Espín

Dep. Ingeniería Aeroespacial y Mecánica de Fluidos
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2018



Trabajo Fin de Grado
Grado Ingeniería Aeroespacial

**Aplicación del post-procesado estadístico de
predicciones meteorológicas por conjuntos a la
predicción de trayectoria de aeronaves**

Autor:

Manuel Gutiérrez Fernández

Tutor:

Antonio Franco Espín

Profesor Contratado Doctor Interino

Dep. Ingeniería Aeroespacial y Mecánica de Fluidos
Escuela Técnica Superior de Ingeniería
Universidad de Sevilla

Sevilla, 2018

Trabajo Fin de Grado: Aplicación del post-procesado estadístico de predicciones meteorológicas por conjuntos a la predicción de trayectoria de aeronaves

Autor: Manuel Gutiérrez Fernández
Tutor: Antonio Franco Espín

El tribunal nombrado para juzgar el trabajo arriba indicado, compuesto por los siguientes profesores:

Presidente:

Vocal/es:

Secretario:

acuerdan otorgarle la calificación de:

El Secretario del Tribunal

Fecha:

*A mi familia y amigos
A todos los que me han hecho posible llegar hasta aquí*

Agradecimientos

El fin del Grado se acerca y con él la realización de este trabajo. Ahora, toca agradecer a todos los profesores el esfuerzo que han hecho para formar no sólo a mí, si no a todo un repertorio de grandes profesionales de la ingeniería. La consecución de este trabajo demuestra de cara al público lo que los ingenieros pueden hacer.

Agradecer a la Escuela Superior de Ingeniería de Sevilla el permitirme estudiar en su centro proporcionando todos los medios necesarios para llevar a cabo mi aprendizaje durante estos cuatro años. De igual manera, no puedo dejar atrás a todo el personal que contribuye al funcionamiento de este centro y que hace posible que esta Escuela tenga el prestigio que se merece.

Mencionar a mis padres que mediante su contribución moral y económica, han posibilitado en gran medida que esté aquí en este momento a punto de ser graduado, escribiendo el trabajo fin de grado.

De manera especial, dar las gracias a Antonio Franco Espín, tutor del trabajo, su dedicación, esfuerzo, paciencia y capacidad para transmitirme los conocimientos desarrollados en el trabajo. También a Damián Rivas Rivas, catedrático del departamento, por su implicación en el tema y por darme la oportunidad de empezar el proyecto.

*Manuel Gutiérrez Fernández
Autor del trabajo*

Sevilla, 2018

Resumen

El título del trabajo hace referencia a las trayectorias, y eso es lo que se estudiará en este trabajo. Son muchos los factores que crean incertidumbre en la trayectoria de una aeronave, pero sin duda el viento es el más importante.

Por una parte, el proyecto consiste en analizar técnicas que permiten establecer modelos de predicción para los vientos, que nos ayudan a hacer su tratamiento algo menos complejo. Para ello, se utilizan códigos numéricos en Matlab. Así también, se desarrollan conceptos e indicadores que posibilitan de alguna forma una cuantificación de lo bueno o lo malo que es nuestro modelo de predicción. Algunos de ellos son el Absolute Error, el Ignorance Error, o el CRPS que serán objeto de nuestro estudio.

Una vez que se definen los modelos de predicción, se utilizan como datos de entrada para una trayectoria, la cual puede ser real o no. En este caso, la trayectoria elegida es entre las ciudades de Berlín y Washington D.C. El cálculo de este recorrido con viento permite ver las diferencias entre un vuelo estimado sin viento, calculado por la Mecánica del Vuelo, y uno en el que se utilizan distintos escenarios o predicciones de vientos. Para el cálculo de esas trayectorias, se emplearán modelos de interpolación y ecuaciones para modelar tanto los movimientos de la aeronave como la forma esférica de la Tierra.

Por último, se trata de aplicar a la trayectoria anterior modelos de vientos que han sido calibrados, utilizando los indicadores estudiados en la primera parte del proyecto. De esta manera, se puede comprobar la relación existente entre los tiempos de vuelo dados por los modelos de predicción proporcionados por el ECMWF (European Centre for Medium-Range Weather) y los modelos de predicción calibrados, así como los calculados con los valores reales medidos el día de ejecución de la trayectoria. Los resultados obtenidos permiten realizar ciertas conclusiones en base a la rentabilidad del método.

Abstract

The title of the work refers to trajectories, and that is what will be studied in this work. There are many factors that create uncertainty in the trajectory of an aircraft, but certainly the wind is the most important.

On the one hand, the project consists of analyzing techniques that allow us to establish prediction models for the winds, which help us to make their treatment somewhat less complex. To do this, numeric codes are used in Matlab. Likewise, concepts and indicators are developed that make possible in some way a quantification of the good or the bad that is our prediction model. Some of them are the Absolute Error, the Ignorance Error, or the CRPS that will be the object of our study.

Once the prediction models are defined, they are used as input data for a path that may or may not be real. In this case, the path chosen is between the cities of Berlin and Washington D.C. The calculation of this route with wind allows to see the differences between an estimated flight without wind, calculated by the Mechanics of Flight, and one in which different scenarios or wind predictions are used. For the calculation of these trajectories, interpolation models and equations will be used to model the movements of the aircraft as well as the spherical shape of the Earth.

Finally, the aim is to apply wind models that have been calibrated to the previous trajectory, using the indicators studied in the first part of the project. In this way, you can check the relationship between the flight times given by the prediction models provided by the ECMWF (European Center for Medium-Range Weather) and the predicted models calibrated, as well as those calculated with the actual measured values. the day of execution of the trajectory. The results obtained allow us to make certain conclusions based on the profitability of the method.

Índice Abreviado

<i>Resumen</i>	V
<i>Abstract</i>	VII
<i>Índice Abreviado</i>	IX
<i>Notación</i>	XIII
<i>Índice de Figuras</i>	1
<i>Índice de Tablas</i>	3
1 Introducción	5
2 Estructura del trabajo	7
3 Calibración de los modelos	9
3.1 Introducción	9
3.2 Predicciones meteorológicas por conjuntos	10
3.3 Evaluación de modelos de predicción	11
3.4 Post-procesado estadístico de modelos de predicción	23
4 Predicción de trayectorias	33
4.1 Introducción	33
4.2 Caracterización de la trayectoria	33
4.3 Ecuaciones de la trayectoria	35
4.4 Resolución Numérica de la trayectoria	37
5 Aplicación. Trayectoria sujeta a incertidumbre	41
5.1 Vuelo de estudio	41
5.2 Conjuntos de calibración	43
5.3 Resultados de los vientos tras la calibración	50
5.4 Tiempos de vuelo tras el postprocesado	57
5.5 Conclusiones generales	62
6 Futuras Mejoras. Recomendaciones	65
7 Anexos	67
7.1 B. Cálculo de Scores	68
7.2 C. ECC	69

7.3 D. Trayectoria	71
<i>Referencias</i>	75

Índice

<i>Resumen</i>	V
<i>Abstract</i>	VII
<i>Índice Abreviado</i>	IX
<i>Notación</i>	XIII
<i>Índice de Figuras</i>	1
<i>Índice de Tablas</i>	3
1 Introducción	5
2 Estructura del trabajo	7
3 Calibración de los modelos	9
3.1 Introducción	9
3.2 Predicciones meteorológicas por conjuntos	10
3.3 Evaluación de modelos de predicción	11
3.3.1 PIT histograms	11
3.3.2 Continuous Ranked Probability Score (CRPS)	15
3.3.3 Absolute Error e Ignorance Score (AE/IGN)	17
3.3.4 Resultados de la aplicación de los Scores	18
3.4 Post-procesado estadístico de modelos de predicción	23
3.4.1 Non-homogeneous regression o ensemble model output statistics (NR/EMOS)	23
3.4.2 Calibración de modelos de una variable	24
3.4.3 Calibración de modelos de varias variables	28
Introducción	28
Ensemble Copula Coupling (ECC)	29
4 Predicción de trayectorias	33
4.1 Introducción	33
4.2 Caracterización de la trayectoria	33
4.3 Ecuaciones de la trayectoria	35
4.4 Resolución Numérica de la trayectoria	37
5 Aplicación. Trayectoria sujeta a incertidumbre	41
5.1 Vuelo de estudio	41
5.2 Conjuntos de calibración	43

5.3	Resultados de los vientos tras la calibración	50
5.4	Tiempos de vuelo tras el postprocesado	57
5.5	Conclusiones generales	62
6	Futuras Mejoras. Recomendaciones	65
7	Anexos	67
7.1	B. Cálculo de Scores	68
7.2	C. ECC	69
7.3	D. Trayectoria	71
	<i>Referencias</i>	75

Notación

F	Función de probabilidad acumulada de una variable aleatoria.
f	Función de densidad de probabilidad de una variable aleatoria.
$\mathcal{N}(\mu, \sigma^2)$	Distribución de probabilidad normal de media μ y de varianza σ^2 .
X	Variable aleatoria que se distribuye según una determinada función.
PIT	Probability Integral Transform.
μ	Media de una distribución aleatoria.
σ	Desviación típica de una variable aleatoria.
σ^2	Varianza de una variable aleatoria.
g	Función de densidad de probabilidad de una distribución Gumbel.
$G(\mu, \sigma)$	Función de probabilidad acumulada para una Gumbel, de parámetro de localización μ , y de factor de escala σ .
$E[X]$	Esperanza matemática de una variable aleatoria X .
S^2	Varianza muestral de una variable aleatoria.
\bar{X}	Media muestral de una variable aleatoria.
γ	Constante de Euler-Mascheroni.
$CRPS$	Continuous Ranked Probability Score.
AE	Absolute Error.
IGN	Ignorance Score o Ignorance Error.
$\mathbb{I}\{\}$	Función de Heaviside.
Φ	Función de probabilidad acumulada de una distribución normal de media 0 y desviación típica 1.
ϕ	Función de densidad de probabilidad de una normal de media 0 y desviación típica 1.
CDF	Cumulative Distribution Function.
PDF	Probability Density Function.
$argmin()$	Argumento que minimiza ().
$median(F)$	Mediana de una variable aleatoria con función de probabilidad acumulada, F .
\log	Logaritmo en base 10.
\ln	Logaritmo neperiano.

t_α	Valor de una variable aleatoria con función de densidad de probabilidad t-student correspondiente a una probabilidad α .
NR	Non-Homogeneous Regression.
$EMOS$	Ensemble Model Output Statistics.
$ X $	Valor absoluto de X .
$y X$	Función de probabilidad y condicionada por la variable aleatoria X .
X	Distribución de una variable aleatoria.
$[X]$	Vector X .
ECC	Ensemble Copula Coupling.
F^{-1}	Inversa de una distribución de probabilidad acumulada.
λ	Longitud de un punto de la superficie terrestre.
ϕ	Latitud de un punto de la superficie terrestre.
γ_a	Coefficiente de dilatación adiabática del aire.
R_a	Constante de los gases ideales para el aire.
χ_g	Curso entre dos puntos.
χ	Rumbo entre dos puntos.
ξ	Media muestral de los CRPS.
UT	Tiempo universal.
h	Altura.
M	Número de Mach.

Índice de Figuras

3.1	Forma de los PIT según el modelo. Los modelos se han generado siguiendo una normal de forma $\mathcal{N}(\mu, \sigma^2)$. En este caso, las observaciones se distribuyen como $\mathcal{N}(0, 1)$. Los modelos en bruto tomados para cada caso son: a) $\mathcal{N}(0, 0.5)$, b) $\mathcal{N}(0, 2)$ y c) $\mathcal{N}(0.5, 1)$	11
3.2	Gráficas Ejemplo 1	12
3.3	Gráficas de Normal y Gumbel con 1000 observaciones para Normal Truth	14
3.4	Gráficas de Normal y Gumbel con 1e6 observaciones para Normal Truth	14
3.5	Gráficas de Normal y Gumbel con 1000 observaciones para Gumbel Truth	14
3.6	Gráficas de Normal y Gumbel con 1e6 observaciones para Gumbel Truth	15
3.7	Definición gráfica del CRPS. Dada una observación, por ejemplo de temperatura, el área sombreada coincide con el valor del indicador CRPS	16
3.8	Cálculo de Scores para la Normal Truth siguiendo distintas distribuciones	19
3.9	Cálculo de Scores para la Gumbel Truth siguiendo distintas distribuciones	19
3.10	Media de los Scores con un intervalo de confianza del 95 %. Caso Normal Truth	21
3.11	Media de los Scores con un intervalo de confianza del 95 %. Caso Gumbel Truth	22
3.12	Ejemplo gráfico de lo descrito en el párrafo anterior. En azul, se muestran las diferentes predicciones; en rojo su valor medio; y en verde un valor posible de una observación del día posterior. Se puede ver claramente la necesidad de calibrar el modelo para que la observación no caiga fuera del rango de las predicciones	23
3.13	La figura muestra un modelo de predicción, en este caso continuo, que hay que ajustar para que se corresponda con las observaciones realizadas	24
3.14	Variables de entrada y salida en la calibración de modelos de una variable	26
3.15	La figura muestra los diferentes modelos de temperaturas de 4 zonas cercanas, así como la temperatura para zonas intermedias. En azul vemos un modelo obtenido suponiendo que la interpolación se ha realizado a través de una minimización del CRPS en cada punto de manera independiente. En rojo, se aprecian modelos más suaves de temperatura que tienen más sentido físico. Los modelos se han exagerado mucho para poder entender el error que existe al aplicar el método descrito en esta sección	28
3.16	Variables de entrada y salida en el ECC	29
3.17	Relaciones de presión y temperatura entre Berlín y Hamburgo. Ilustraciones tomadas de [7]	31
3.18	Mapas de temperaturas sobre Alemania. Ilustración tomada de [8]	32
4.1	Ejemplo de segmento de trayectoria	35
4.2	Vientos y temperaturas para cada punto de la malla	35
4.3	Cálculo del viento longitudinal y cruzado	36
4.4	Triángulo de vientos	36

4.5	Interpolación de los valores del viento en la malla	38
5.1	Representación de la trayectoria en proyección cilíndrica equidistante	42
5.2	Representación de los tiempos de vuelo durante la ida para el primer conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales	48
5.3	Representación de los tiempos de vuelo durante la ida para el segundo conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales	48
5.4	Representación de los tiempos de vuelo durante la vuelta para el primer conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales	49
5.5	Representación de los tiempos de vuelo durante la vuelta para el segundo conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales	49
5.6	Representación del ensemble forecast antes de la calibración, tras realizar la calibración individual, EMOS y una vez aplicado el ECC	51
5.7	Representación del ensemble forecast antes de la calibración, tras realizar la calibración individual, EMOS y una vez aplicado el ECC	52
5.8	Representación de las curvas de nivel para el viento zonal. Primer conjunto de entrenamiento	53
5.9	Representación de las curvas de nivel para el viento meridional. Primer conjunto de entrenamiento	54
5.10	Representación de las curvas de nivel para el viento zonal. Segundo conjunto de entrenamiento	55
5.11	Representación de las curvas de nivel para el viento meridional. Segundo conjunto de entrenamiento	56
5.12	Modelo en bruto de los tiempos de vuelo para el recorrido de ida	57
5.13	Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de ida. Primer conjunto de entrenamiento	59
5.14	Modelo en bruto de los tiempos de vuelo para el recorrido de vuelta	59
5.15	Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de vuelta. Primer conjunto de entrenamiento	60
5.16	Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de ida. Segundo conjunto de entrenamiento	61
5.17	Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de vuelta. Segundo conjunto de entrenamiento	62

Índice de Tablas

3.1	Resultados de las medias de los diferentes Scores según las tres distribuciones consideradas. Caso Normal Truth	21
3.2	Resultados de las medias de los diferentes Scores según las tres distribuciones consideradas. Caso Gumbel Truth	22
3.3	Distribuciones según variables. Tabla basada en [8]	27
3.4	Tabla de reordenación de las muestras	30
3.5	Orden global de todos los escenarios	30
5.1	Coordenadas geográficas de los diferentes waypoints	41
5.2	Distancias y cursos entre los diferentes waypoints tanto en ida como en vuelta	42
5.3	Tiempos de vuelos previstos para la ida de los 6 primeros meses. Primer conjunto de entrenamiento	44
5.4	Tiempos de vuelos previstos para la ida de los 6 últimos meses. Primer conjunto de entrenamiento	44
5.5	Tiempos de vuelos correspondientes a cada uno de los días 1 y 15 del 2017 según los vientos observados en cada punto de la trayectoria en el recorrido de ida. Primer conjunto de entrenamiento	45
5.6	Tiempos de vuelos previstos y observados para los primeros 14 días del mes de Diciembre. Recorrido de ida. Segundo conjunto de entrenamiento	45
5.7	Tiempos de vuelos previstos para la vuelta de los 6 primeros meses. Primer conjunto de entrenamiento	46
5.8	Tiempos de vuelos previstos para la vuelta de los 6 últimos meses. Primer conjunto de entrenamiento	46
5.9	Tiempos de vuelos correspondientes a cada uno de los días 1 y 15 del 2017 según los vientos observados en cada punto de la trayectoria en el recorrido de vuelta. Primer conjunto de entrenamiento	47
5.10	Tiempos de vuelos previstos y observados para los primeros 14 días del mes de Diciembre. Recorrido de vuelta. Segundo conjunto de entrenamiento	47
5.11	Resultados de la optimización para el primer conjunto de entrenamiento	50
5.12	Coefficientes de correlación entre ambas ciudades para el primer conjunto de entrenamiento	51
5.13	Resultados de la optimización para el segundo conjunto de entrenamiento	51
5.14	Coefficientes de correlación entre ambas ciudades para el segundo conjunto de entrenamiento	52

1 Introducción

La predicción de trayectorias resulta muy interesante para toda aquella persona implicada en el mundo de la aeronáutica y la navegación aérea. No sólo para ellos y ellas es esencial el conocimiento de esta disciplina si no que también la meteorología juega un papel muy importante, y por tanto, los meteorólogos son imprescindibles para su desarrollo.

El objetivo de este trabajo es dar un paso más en la fusión de la meteorología y la aeronáutica. Como numerosos trabajos anteriores del departamento de Ingeniería Aeroespacial y Mecánica de Fluidos de la Universidad de Sevilla, los vientos son el objeto de estudio de este trabajo. La novedad que incluye este proyecto es que los modelos de viento que proporciona el European Centre for Medium-Range Weather Forecasts (ECMWF) no se consideran tal cual para el cálculo de trayectorias. Lo que se intenta es realizar una serie de modificaciones a esos vientos proporcionados por el ECMWF, de manera que se puedan obtener otros modelos de vientos que predigan mejor la realidad. A partir de ello, mediante una serie de técnicas se resuelven problemas de trayectorias que integran los nuevos modelos de viento calculados, comparándose las características de la trayectoria obtenida, con la trayectoria que se obtendría si se utilizasen los valores en bruto del viento del ECMWF.

Este estudio ha sido siempre bastante complejo, ya que se trabaja esencialmente con variables aleatorias. Por ello, su cálculo siempre se basará en probabilidades y estimaciones, siendo la matemática y la estadística dos ramas necesarias para la obtención de los nuevos modelos.

El campo de la meteorología es un campo muy abierto, por lo que el trabajo es sólo una primera aproximación, que puede completarse y enriquecerse en un futuro, pero que sirve de base para poder entender las técnicas empleadas para resolver el problema, así como las dificultades que éstas conllevan.

Cabe comentar simplemente que muchas de las técnicas recogidas en este trabajo han sido estudiadas antes por numerosos científicos y meteorólogos importantes como Tilmann Gneiting ([2]), Thordis L. Thorarinsdottir ([3]), Michael Scheuerer ([4]) y Roman Schefzik ([8]), entre otros. Las hipótesis realizadas por estos científicos se incluyen en este proyecto, como también algunas nuevas relacionadas con nuestro problema en cuestión. Sin embargo, este trabajo introduce, como novedad, una aplicación de tales técnicas a la predicción de trayectoria de aviones.

Además de las técnicas contempladas en el trabajo, existen muchas más en lo referente al post-procesado de modelos de viento. Algunas de ellas, aunque no son tratadas aquí, se recogen en la bibliografía ([4] y [8]).

2 Estructura del trabajo

En este capítulo se resume la estructura básica del trabajo, con el objetivo de dar a lector una visión rápida de las partes de las que consta el proyecto, pudiendo identificar el mismo la/s partes que más le interese conocer. La estructuración del trabajo será por capítulos divididos a su vez en secciones y subsecciones. En adelante, el contenido propiamente dicho será el que sigue:

- En el Capítulo 3 se estudia todo lo referente a la Calibración de los Modelos, es decir, tanto las herramientas necesarias como las técnicas empleadas para ello. A su vez se considera el caso en el que sólo se trabaja con una variable, y cómo se puede extrapolar su estudio a varias variables.
- En el Capítulo 4 se describen todas las hipótesis y características de la Predicción de Trayectorias, para una trayectoria genérica. Se muestran también tanto la integración del viento en la trayectoria, como la resolución numérica de la misma. Además se contemplan técnicas de resolución numérica en el caso en los que se dispone de varios escenarios de vientos distintos.
- En el Capítulo 5 se analiza una trayectoria concreta. En nuestro caso esa trayectoria tendrá origen en Berlín y destino en Washington D.C. En este capítulo se aplican las técnicas analizadas en los Capítulos 3 y 4 a un caso particular, remarcando cuáles son las peculiaridades más importantes a tener en cuenta cuando se tiene un caso concreto. Por último, se realiza una comparativa de los resultados obtenidos tanto para los vientos como para los tiempos en función de los diferentes escenarios y modelos considerados.
- En el Capítulo 6 se recogen una serie de recomendaciones y mejoras futuras para abordar una posible continuación del proyecto.
- Al final del todo se recopilan en unos Anexos los códigos de Matlab que se han empleado en los capítulos anteriores para resolver las cuestiones planteadas que requerían de cálculo numérico.

3 Calibración de los modelos

En este capítulo se analiza cómo se puede evaluar cuantitativamente si un modelo de predicción está en consonancia con las observaciones de la realidad. Todo ello se realiza de una forma sencilla para que sea asequible su entendimiento.

3.1 Introducción

Los procesos de calibración de variables que suponen una incertidumbre meteorológica como pueden ser el viento, la temperatura o la presión siguen siendo importantes porque condicionan el carácter operativo de la aeronave.

El primer intento para llevar a cabo la creación de modelos de predicción y su calibración, es intentar que los modelos recojan todas las posibilidades de combinación que tienen estas variables. Sin embargo, esto es imposible, ya que se tendría infinitos modelos y por tanto, haría falta ordenadores con capacidad de computación muy elevada. No sólo eso sería un problema, si no que el tiempo de cálculo empleado sería mucho mayor que el tiempo que pasa hasta el día en el que se necesita conocer la predicción. En definitiva, todo un caos que impide conseguir resultados efectivos.

Por ese razonamiento el European Centre for Medium-Range Weather Forecast (ECMWF) utiliza los denominados Numerical Weather Prediction. Esto es un conjunto de predicciones meteorológicas basadas en ligeras variaciones de las condiciones iniciales e incluso de los parámetros que definen los modelos de predicción, que dan lugar a escenarios futuros diferentes.

Aun así, hay veces que el conjunto de predicciones proporcionado por el ECMWF no se adapta bien al comportamiento de una determinada variable en la realidad. Para que un modelo sea adecuado, la forma de la distribución de las observaciones tiene que ser muy parecida a la de las predicciones. Por ello, surge ahora la necesidad de un post-procesado o calibración de los modelos del ECMWF, que combinen realidad y predicción para generar nuevos modelos futuros algo más precisos.

Existen muchos métodos diferentes de calibración. En este trabajo nos centraremos en uno de ellos denominado Non-homogeneous regression o ensemble model output statistics (NR/EMOS). Antes de proceder con el EMOS, se introducirán algunos conceptos que servirán de ayuda para implementar dicho método.

3.2 Predicciones meteorológicas por conjuntos

Una novedad que introduce este trabajo es el hecho de que no se trabaja con valores únicos para las variables meteorológicas, si no que éstas vienen dadas por conjuntos. El *European Centre for Medium-Range Weather Forecast* (ECMWF) proporciona dichos conjuntos que también son llamados *ensembles*. Los *ensembles* no son más que una serie de predicciones meteorológicas reales obtenidas utilizando el método *Ensemble Prediction System* (EPS). Este método, no sólo busca la mejor condición inicial que conlleve una predicción más acertada o realista, si no que proporciona un grupo de predicciones adicionales obtenidas mediante ligeras variaciones de las condiciones iniciales. El resultado es un conjunto de predicciones futuras que tienen como función reducir en la medida de lo posible la incertidumbre existente en la naturaleza, ya sea por los posibles errores en la medición o por el desconocimiento de gran parte de la dinámica terrestre, [13] [14]. Cada una de esas predicciones será también denominada con el nombre de *escenario* con el objetivo de enfatizar que se trata de una predicción entre múltiples predicciones abordadas.

Son muchos los organismos que participan en el proyecto del ECMWF. Básicamente, se diferencian en el método que utilizan para obtener los *ensembles* y en el número de miembros del mismo (escenarios). Los modelos disponibles son los siguientes [15]:

- **POAMA:** desarrollado por *Bureau of Meteorology en Australia*, cuenta con 33 miembros en su ensemble.
- **MOGREPS-15:** implementado por *China Meteorological Administration* tiene 15 miembros.
- **CPTEC EPS:** el ensemble del *Centro de Previsao Tempo e Estudos Climaticos* tiene 15 miembros.
- **ECMWF EPS:** el modelo del ECMWF es el que contiene más miembros, un total de 51.
- **JMA One-week EPS:** *Japan Meteorological Agency* posee un modelo con 27 miembros.
- **KMA EPSG:** el modelo desarrollado por *Korea Meteorological Administration* tiene un total de 24 miembros.
- **PEARP:** el modelo de *Météo-France* está compuesto por 35 miembros.
- **MOGREPS-G:** el modelo de *MetOffice* (Exeter, Reino Unido) tiene 12 miembros.
- **NCEP GEFS:** el modelo desarrollado de forma conjunta por *National Centres for Environmental Prediction* y *The National Center for Atmospheric Research* en EEUU tiene 21 miembros.

En adelante, sólo se hará referencia al ECMWF EPS, ya que será el que suministrará los datos con los que se trabajará en este proyecto. Concretamente, el ECMWF EPS emplea 50 escenarios perturbados más 1 escenario de control, que tienen una resolución razonable, y que pueden llegar a dar predicciones con 32 días de antelación. También utiliza un escenario de alta resolución, permitiendo sólo 10 días de horizonte de predicción. Algunos científicos como Rodwell (en [2]) encontraron que la mejor predicción se establecía con una combinación entre el de alta resolución y los 50 de resolución razonable. Surge así la integración de ambas en la denominada *Integrated Forecasting System* (IFS), la cual no será considerada en este trabajo.

Añadir a lo que se ha dicho anteriormente, que el ECMWF EPS proporciona los valores del viento zonal y meridional (w_u y w_v) en un día, hora y horizonte temporal (antelación de la predicción) concreto, para cada punto de una malla de longitud y latitud determinada. La malla se considera contenida en el plano horizontal en el que vuela la aeronave.

3.3 Evaluación de modelos de predicción

En las subsecciones sucesivas se desarrollan herramientas que permiten evaluar si un modelo de predicción es bueno o malo para representar la realidad. También se establecen métricas con el objetivo de comparar entre distintos modelos de predicción. Durante todo el desarrollo, se ilustran ejemplos sencillos que ayudan a entender la aplicación de los conceptos estudiados.

3.3.1 PIT histograms

Es muy utilizada la representación de histogramas PIT (*Probability Integral Transform*). Dado un conjunto de observaciones $y \in \{y_1, y_2 \dots y_N\}$ y un modelo de predicción que viene dado por una función de probabilidad acumulada F , el cálculo del PIT consiste en transformar dichas observaciones mediante la función F . Llamaremos z , al conjunto $z \in \{z_1, z_2 \dots z_N\}$ resultante de la transformación

$$z_1 = F(y_1); z_2 = F(y_2); \dots z_N = F(y_N); \quad (3.1)$$

Evaluando la forma del histograma se puede tener una idea del ajuste del modelo. Si el resultado tiene forma de \cup , le falta dispersión; si tiene forma de \cap , es que tiene más dispersión de la cuenta (sobredispersa) y si es triangular es que hay una desviación en la media o sesgo.

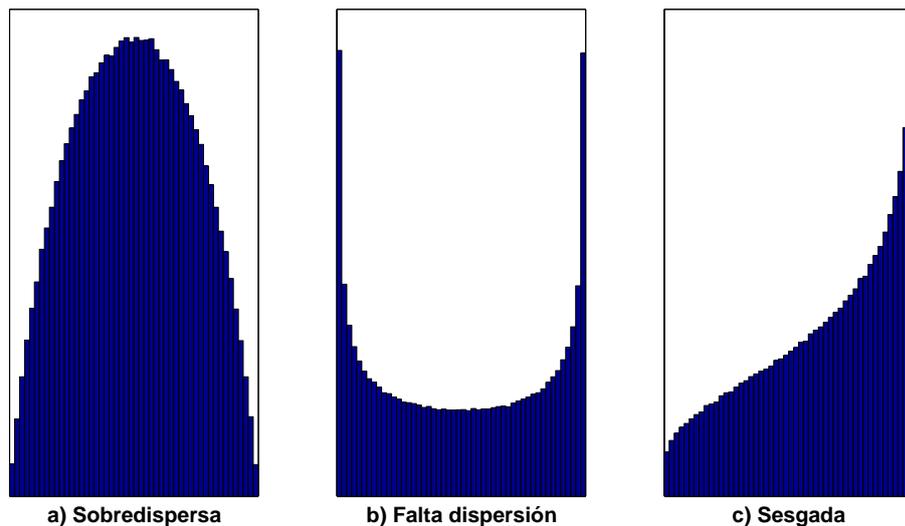


Figura 3.1 Forma de los PIT según el modelo. Los modelos se han generado siguiendo una normal de forma $\mathcal{N}(\mu, \sigma^2)$. En este caso, las observaciones se distribuyen como $\mathcal{N}(0, 1)$. Los modelos en bruto tomados para cada caso son: a) $\mathcal{N}(0, 0.5)$, b) $\mathcal{N}(0, 2)$ y c) $\mathcal{N}(0.5, 1)$.

A continuación se ilustrarán varios ejemplos para poder mostrar resultados que se pueden encontrar a la hora de realizar dichos histogramas. Los resultados descritos se han obtenido en base a códigos numéricos implementados en Matlab, los cuales se recogen en el apartado de Anexos.

Ejemplo 1. El primer ejemplo será muy sencillo y es para ver cómo tiene que ser la forma de un histograma PIT, y comprobar si el modelo tiene un buen ajuste con las observaciones. Se puede resumir en los siguientes pasos (basados en [3]):

1. Se generan las observaciones, y , de la siguiente forma:
 - Se saca un valor aleatorio de la media de la normal, μ , que se comporta a su vez como una

normal $\mathcal{N}(\mu, \sigma^2)$, cuyos parámetros son $\mathcal{N}(25, 1)$.
 -Se saca un valor observado y , mediante una normal $\mathcal{N}(\mu, 9)$.
 Este proceso se repite 1000 veces y 1e6 veces.

2. Se define un modelo de predicción X , como una normal $\mathcal{N}(25, 10)$. Este modelo se caracteriza por una distribución de probabilidad acumulada F .
3. Se transforman las observaciones mediante la función F , obteniéndose $z = F(y)$.
4. Se lleva a cabo la representación del histograma de las z considerando barras equiespaciadas (en nuestro caso 50).

Se obtienen gráficas como las siguientes:

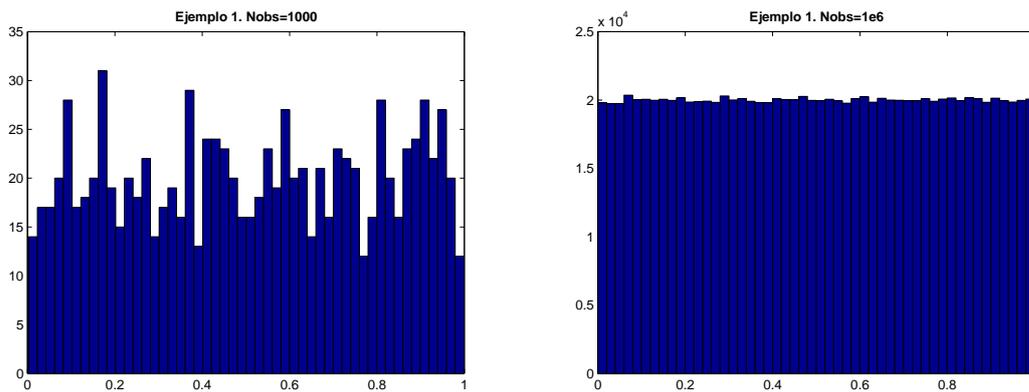


Figura 3.2 Gráficas Ejemplo 1.

Se pueden comentar varias cosas. En primer lugar, se ve que con sólo 1000 observaciones el comportamiento es mucho más caótico y no permite sacar ninguna conclusión ya que cada probabilidad tiene una repetición distinta. En cambio, cuando el número de observaciones va aumentando y se va acercando a infinito el comportamiento del modelo se asemeja al de las observaciones. Con un millón, todas las probabilidades se repiten con la misma frecuencia, lo que nos indica que el modelo coincide con el que se han generado las observaciones.

Ejemplo 2. En este ejemplo se usará otra función de distribución aparte de la normal, la denominada Gumbel. La función Gumbel es muy utilizada para modelar valores máximos y valores mínimos, por lo que su empleo para predecir terremotos, inundaciones, sequías, etc es habitual. Su función de densidad de probabilidad para modelar el máximo viene dada por

$$g(x; \mu, \sigma) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{3.2}$$

y su función de probabilidad acumulada es igual a

$$G(x; \mu, \sigma) = \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{3.3}$$

así como para modelar el mínimo simplemente hay que cambiar de signo el argumento de las exponenciales.

Antes de proceder con el ejemplo en sí, es necesario definir un método que nos permita, en base a

las observaciones, estimar los parámetros de los que dependen sus distribuciones. Existen varios métodos para realizar estimaciones. Los más conocidos son el Método de los Momentos y el Método de Máxima Verosimilitud . En nuestro caso, se utiliza el primero de ellos.

Método de los Momentos. Sea (X_1, X_2, \dots, X_N) una muestra aleatoria de una población X con función de probabilidad f_θ , con parámetros desconocidos $(\theta_1, \theta_2, \dots, \theta_k)$. Los estimadores de θ se obtienen de resolver el sistema de ecuaciones:

$$\begin{cases} E_\theta[X] &= \frac{1}{N} \sum_{i=1}^N X_i \\ \vdots &= \vdots \\ E_\theta[X^k] &= \frac{1}{N} \sum_{i=1}^N X_i^k \end{cases} \quad (3.4)$$

donde $E_\theta[X^j]$ se denomina momento de orden j .

Simplemente resolviendo el sistema para los dos parámetros de la normal se obtiene,

$$\mu_n = \frac{1}{N} \sum_{i=1}^N X_i \quad (3.5)$$

$$\sigma_n^2 = s^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \tilde{X})^2 \quad (3.6)$$

igualmente para la gumbel los parámetros son los siguientes (según [12])

$$\sigma_g = \frac{\sqrt{6}}{\pi} s \quad (3.7)$$

$$\mu_g = \tilde{X} - \gamma \sigma_g \quad (3.8)$$

siendo γ la constante de Euler-Mascheroni $\gamma = 0.5772156649015328606$.

Una vez que se sabe como estimar los parámetros de las distribuciones en cuestión, se está en condiciones suficientes para desarrollar los pasos a seguir en el segundo ejemplo. Esos pasos están detallados a continuación (basados en [3]).

1. La realidad se va a comportar como una normal (Normal Truth), $\mathcal{N}(\mu, \sigma^2)$, con $\mu \sim \mathcal{N}(25, 1)$ y $\sigma^2 = 9$.
2. Se generan 300 observaciones aleatorias de acuerdo a esa realidad.
3. Se realiza un ajuste de los parámetros de la Normal y de la Gumbel, mediante la estimación basada en el Método de los Momentos, usando las observaciones anteriores.
4. Se simulan otras 1000 o 1e6 observaciones de la realidad.
5. Se transforman las observaciones anteriores mediante la función de probabilidad acumulada de los modelos ajustados anteriormente. Luego se representan los histogramas considerando 50 barras equiespaciadas.
6. Se repiten los pasos del 2 al 5, pero considerando ahora la realidad como una función Gumbel (Gumbel Truth), $G(\mu, \sigma)$, de media $\mu + \gamma\sigma \sim \mathcal{N}(25, 1)$, y de varianza $\frac{\pi^2\sigma^2}{6} = \frac{3\pi^2}{2}$.

Algunos de los resultados obtenidos se muestran a continuación, tanto para la distribución Normal como para la Gumbel.

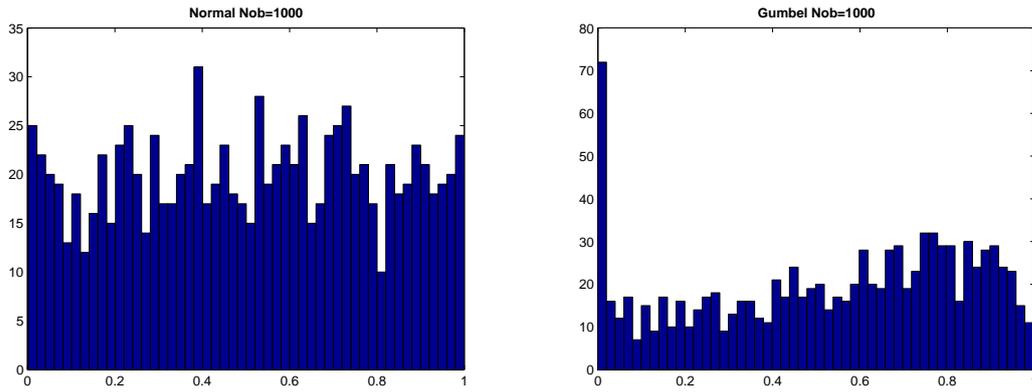


Figura 3.3 Gráficas de Normal y Gumbel con 1000 observaciones para Normal Truth.

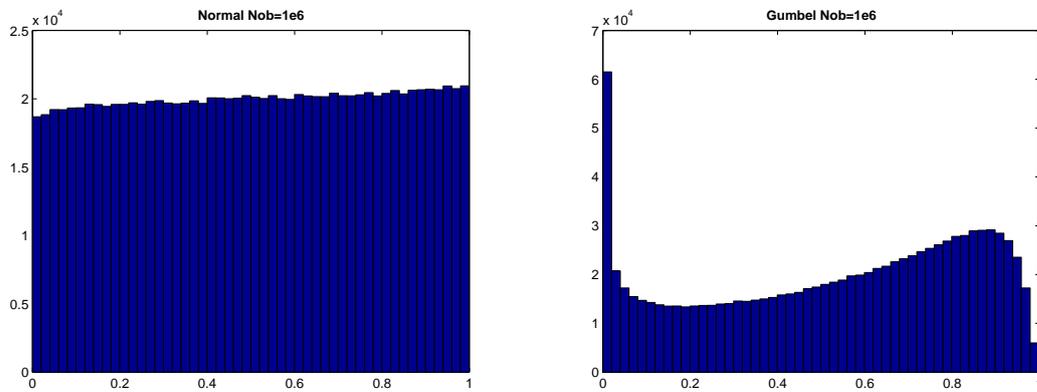


Figura 3.4 Gráficas de Normal y Gumbel con 1e6 observaciones para Normal Truth.

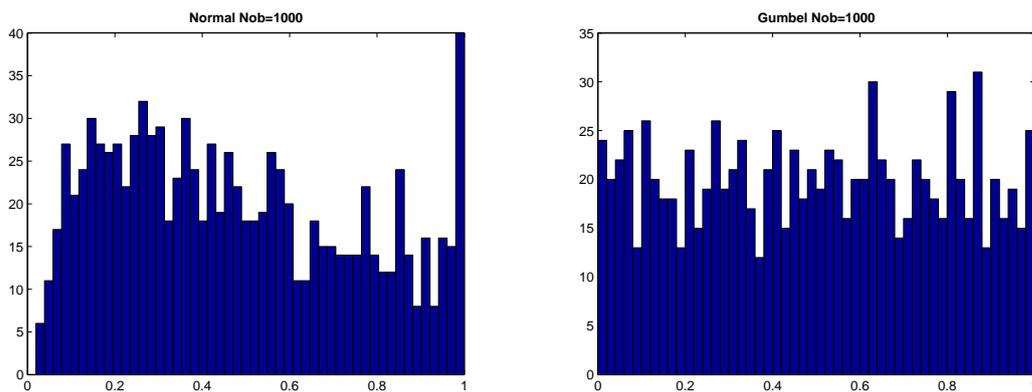


Figura 3.5 Gráficas de Normal y Gumbel con 1000 observaciones para Gumbel Truth.

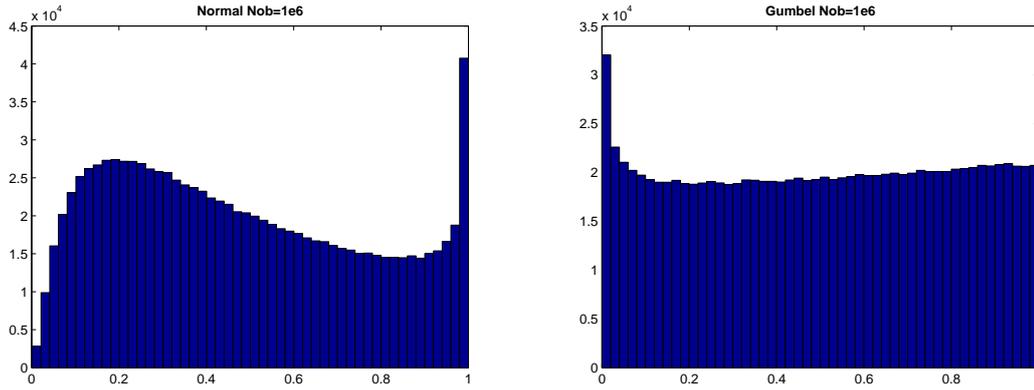


Figura 3.6 Gráficas de Normal y Gumbel con 1e6 observaciones para Gumbel Truth.

Cabe comentar varias conclusiones en base a los resultados obtenidos. Por una parte, se observa que cuando la realidad se comporta como una Normal o como una Gumbel, sus respectivas gráficas tienden a un valor uniforme (salvo errores en la estimación del Método de los Momentos). Sin embargo, esto sólo se puede ver cuando el número de observaciones es muy alto, por ejemplo un millón. Las gráficas obtenidas transformando las observaciones por modelos que no coinciden con la realidad, presentan un comportamiento extraño, que nos induce a pensar que no son buenas estimaciones. Es muy importante tener presente, que cuando el número de observaciones es bajo, tanto la Normal como la Gumbel, presentan un comportamiento del que no se puede extraer ningún patrón. Esto nos supone un problema, ya que en la realidad no se dispone de un número alto de observaciones para ajustar los modelos, y a la vista de lo resultados sería bastante difícil diferenciar entre un modelo Gumbel y un modelo Normal.

Por tanto, aunque los histogramas PIT permitan ver de forma rápida si un modelo está bien o mal encaminado para representar a una realidad concreta, no es lo más adecuado para realizar buenos ajustes, si no sólo para una primera aproximación.

Dicho esto, interesa ahora establecer métricas a la hora de comparar diferentes modelos, así como para comparar los modelos con las observaciones. Surge la necesidad de usar los *scores*. Los *scores* no son más que una serie de indicadores definidos con cierto razonamiento lógico que comparan las características de los modelos de predicción con las características de las observaciones, teniendo en cuenta diversos criterios, como pueden ser sus valores, la forma de sus modelos de predicción o cualquier otro, pero siempre traduciendo dicho criterio a un valor cuantitativo de una magnitud definida mediante una fórmula matemática. Aunque existen muchos *scores*, sólo se mostrarán tres: el *Continuous Ranked Probability Score* (CRPS), el *Absolute Error* (AE) y el *Ignorance Score* (IGN), que se estudiarán en las sucesivas secciones.

3.3.2 Continuous Ranked Probability Score (CRPS)

El primero de los *scores* que definiremos es el CRPS. Su definición (en [5]) es la siguiente.

$$CRPS(F,y) = E_F |X - y| - \frac{1}{2} E_F |X - X'| \quad (3.9)$$

$$CRPS(F,y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{I}\{y \leq x\})^2 dx \quad (3.10)$$

La ecuación (3.9) se utiliza cuando la distribución de probabilidad es discreta, siendo X y X' variables independientes aleatorias que tienen la misma función de probabilidad acumulada. La ecuación (3.10) hace referencia al caso de una distribución continua y será la que se estudiará en este trabajo. El integrando de la ecuación (3.10) tiene un nombre específico denominado *Brier Probability Score*.

El CRPS trata de medir la diferencia que existe entre la función de probabilidad acumulada del modelo de predicción y una función de probabilidad empírica de salto unidad. La función $\mathbb{I}\{y \leq x\}$ es aquella que cuenta o acumula el número de observaciones que son menores o iguales a cada valor considerado de nuestro conjunto de predicción. Su implementación se lleva a cabo mediante la función de Heaviside cuyo valor es unidad si se cumple que $y \leq x$ y nulo si no se da esa condición.

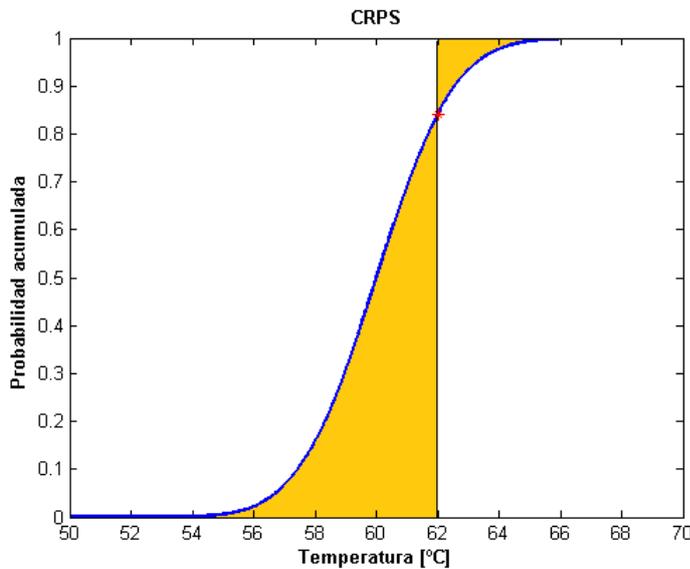


Figura 3.7 Definición gráfica del CRPS. Dada una observación, por ejemplo de temperatura, el área sombreada coincide con el valor del indicador CRPS.

Las ventajas de este indicador es que la función de probabilidad empírica cumple el Teorema de Clivenko-Cantelli. Este teorema dice que si el tamaño de la muestra, $n \rightarrow \infty$, entonces la función de probabilidad empírica converge a la función de probabilidad acumulada de nuestro modelo. Por lo tanto, sirve para cuantificar el parecido de nuestro modelo a las observaciones. Cuanto menor sea el valor del *score* más acertada será la predicción.

A pesar de ser un indicador muy utilizado presenta varios problemas a la hora del cálculo numérico. Uno de ellos, es que el intervalo de integración no es acotado $(-\infty, \infty)$ y hay que definir subintervalos para la integración numérica en los que el peso del integrando no sea significativo. Surge la idea de integrar en probabilidad, pero para ello hay que hacer un cambio de variable en el integrando, el cual tiene ahora el nombre de *Quantile Probability Score*.

$$CRPS(F,y) = 2 \int_0^1 (F^{-1}(\tau) - y)(\mathbb{I}\{y \leq F^{-1}(\tau)\} - \tau)d\tau \tag{3.11}$$

Para trabajar con vientos, presiones o temperaturas, es muy habitual que se modelen dichas variables mediante distribuciones normales. Debido a que la distribución normal es muy utilizada es lógico

considerar una expresión analítica que permita calcular el CRPS, sin necesidad de hacer la integral. Esta expresión es la que sigue (en [1]) :

$$crps[\mathcal{N}(\mu, \sigma^2), y] = \sigma \left\{ \frac{y - \mu}{\sigma} \left[2\Phi\left(\frac{y - \mu}{\sigma}\right) - 1 \right] + 2\varphi\left(\frac{y - \mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \right\} \quad (3.12)$$

siendo Φ y φ la CDF y la PDF, respectivamente, de una normal de media 0 y desviación típica 1, evaluadas en $\left(\frac{y - \mu}{\sigma}\right)$. La media del CRPS será por tanto

$$CRPS = \frac{1}{n} \sum_{i=1}^n crps[\mathcal{N}(\mu, \sigma^2), y_i] \quad (3.13)$$

con n el número de observaciones realizadas.

De esta manera, la sencillez de cálculo numérico se reduce notablemente dando más rapidez en su implementación.

3.3.3 Absolute Error e Ignorance Score (AE/IGN)

Otro *score* muy sencillo de implementar es el Absolute Error (AE). En primer lugar, se define la siguiente *función de Score* del Absolute Error (en [3]) como:

$$S(x, y) = |x - y| \quad (3.14)$$

Sin embargo, cuando tenemos una función de probabilidad acumulada basada en un modelo de predicción, hay que elegir con respecto a que x vamos a medir ese error. La teoría dice que para que el *score* sea consistente debe de calcularse en base al estimador de Bayes definido por,

$$\hat{x} = \operatorname{argmin}_x E_F[S(x, Y)] \quad (3.15)$$

que para el Absolute Error de una función de probabilidad acumulada F , ese valor coincide con su mediana. Por tanto la definición del AE inducida de su *función de Score* es la que sigue.

$$AE(y) = |\operatorname{median}(F) - y| \quad (3.16)$$

Para terminar esta sección, queda explicar el último de los *scores* que se utilizará en los ejemplos sucesivos. Este es el Ignorance Score (IGN). Su definición es la siguiente (en [3]):

$$IGN(F, y) = -\log(f(y)) \quad (3.17)$$

donde f , es la función de densidad de probabilidad correspondiente al modelo de predicción que se está evaluando.

La información contenida en el logaritmo, decrece monótonamente con el incremento de la probabilidad de las observaciones, lo que es muy intuitivo. Por tanto con este método, la calidad del modelo se basa en las observaciones. Hay que destacar también que el valor de probabilidad 0 no debe ser asignado ya que su logaritmo diverge.

Como en el CRPS, existen expresiones equivalentes para calcular el IGN score (en [1]) que permiten ahorrar tiempo de cálculo en el caso en el que la distribución sea normal.

$$\text{ign}[\mathcal{N}(\mu, \sigma^2), y] = \frac{1}{2} \ln(2\pi\sigma^2) + \frac{(y - \mu)^2}{2\sigma^2} \quad (3.18)$$

$$\text{IGN} = \frac{1}{n} \sum_{i=1}^n \text{ign}[\mathcal{N}(\mu, \sigma^2), y_i] \quad (3.19)$$

Por último, solamente comentar que el Ignorance Score es muy útil como medida de verosimilitud de las observaciones. Esta se puede ver haciendo el logaritmo de esas probabilidades y multiplicándolos entre sí. De esta manera, aunque no es un mal indicador, sería mejor si la estimación de los parámetros se realiza con el Método de la Máxima Verosimilitud.

3.3.4 Resultados de la aplicación de los Scores

Cuando se han definido todos los indicadores que se usarán a la hora de cuantificar el grado de aproximación de un modelo con las observaciones tomadas de la realidad, es la hora de realizar ejemplos que ayuden a entender su aplicabilidad.

Antes de proceder a utilizar los scores para la calibración y ajuste de los modelos, se añadirán su cálculo al Ejemplo 2. Los pasos a seguir son los siguientes (basados en [3]):

1. La realidad se va a comportar como una Normal (Normal Truth), $\mathcal{N}(\mu, \sigma^2)$, con $\mu \sim \mathcal{N}(25, 1)$ y $\sigma^2 = 9$.
2. Se generan 300 observaciones aleatorias de acuerdo a esa realidad.
3. Se realiza un ajuste de los parámetros de la Normal y de la Gumbel, mediante la estimación basada en el Método de los Momentos, usando las observaciones anteriores.
4. Se simulan otras 1000 observaciones de la realidad.
5. Mediante las observaciones simuladas por segunda vez, se calcula el CRPS, el Absolute Error y el Ignorance Score. Los cálculos se realizan según las fórmulas presentadas en la sección anterior, utilizando las funciones de probabilidad y de probabilidad acumulada de los modelos estimados y de la realidad ($F(\mu_n, \sigma_n^2)$, $F(\mu_g, \sigma_g)$, $F(\mu, \sigma^2)$, $f(\mu_n, \sigma_n^2)$, $f(\mu_g, \sigma_g)$ y $f(\mu, \sigma^2)$).
6. Se repiten los pasos del 2 al 5, pero considerando ahora la realidad como una función Gumbel (Gumbel Truth), $G(\mu, \sigma)$, de media $\mu + \gamma\sigma \sim \mathcal{N}(25, 1)$, y de varianza $\frac{\pi^2\sigma^2}{6} = \frac{3\pi^2}{2}$.

En primer lugar se muestran las curvas de cada una de las observaciones frente al valor de los indicadores.

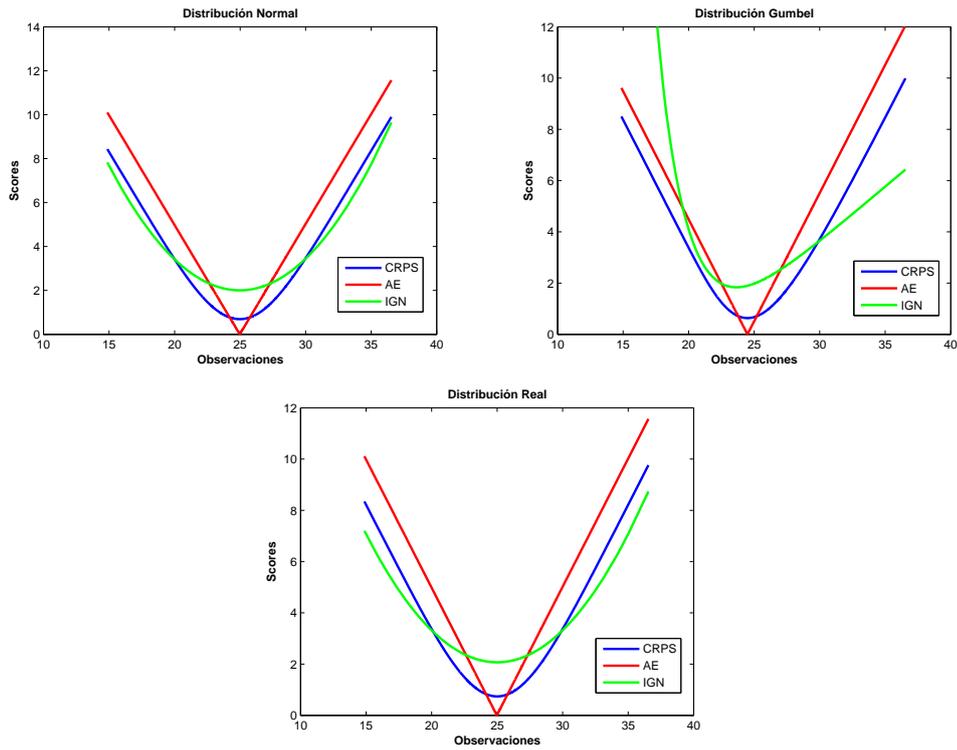


Figura 3.8 Cálculo de Scores para la Normal Truth siguiendo distintas distribuciones.

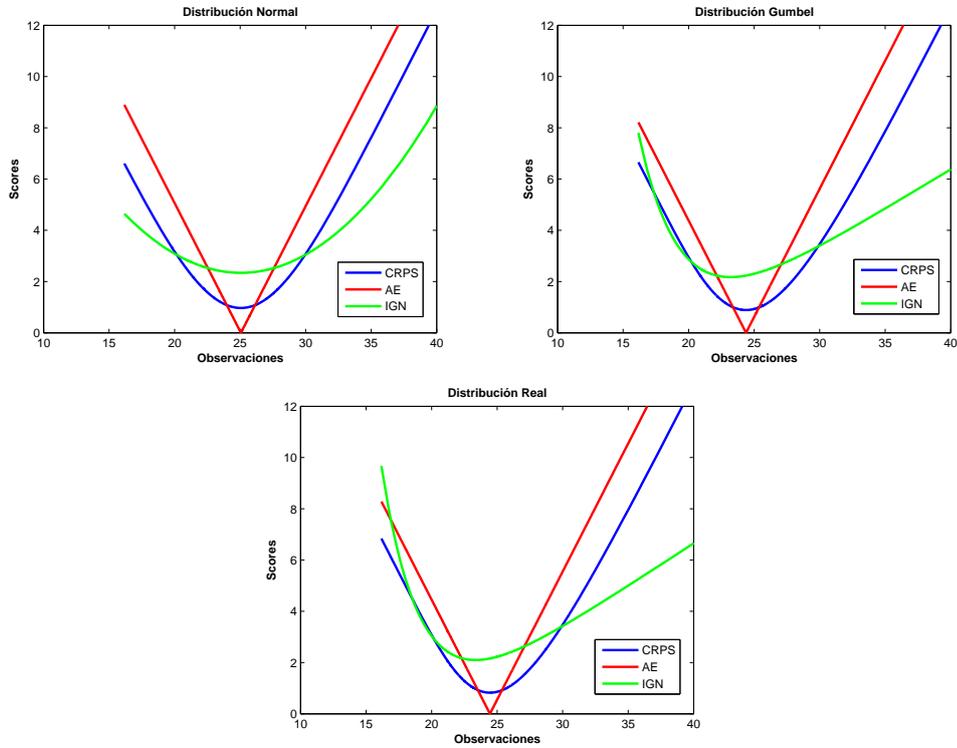


Figura 3.9 Cálculo de Scores para la Gumbel Truth siguiendo distintas distribuciones.

Se puede ver la diferente forma que presentan las curvas según el tipo de indicador. Sin embargo, no difieren mucho en cuanto a su comportamiento (tendencia, crecimiento, etc). Si que es verdad que se puede notar que en las gráficas cuya distribución del modelo empleado es similar al real (el que se ha tomado para generar las observaciones), tienen un parecido muy similar por no decir casi idéntico a la realidad (Normal-True en la figura 8 y Gumbel-True, en la figura 9). Posteriormente, se verá que eso implica que la media de los indicadores extendida a las observaciones realizadas no estará muy lejos de la distribución real o truth.

Una vez vista la representación de las observaciones con respecto al valor que se obtiene de calcular los indicadores, es necesario comparar ahora el valor medio de los Scores según la distribución empleada. Verdaderamente, esta representación será la que nos aporte una medida más útil para poder comparar por ejemplo, entre una distribución Normal y una Gumbel, tendiendo detrás herramientas que justifiquen la elección de una u otra según un mejor ajuste a la realidad. La comparación se realizará siguiendo el siguiente procedimiento:

-Primero se calcula la media de todas las observaciones utilizadas según lo dicho anteriormente.

-Esa media se obtiene para los tres Scores: CRPS, Absolute Error e Ignorance Score. El cálculo se repite para cada una de las distribuciones y a su vez para cada una de las realidades: Normal y Gumbel, (en total 18 medias).

-Para dar cierta seguridad a la media calculamos un intervalo de confianza del 95 % siguiendo la siguiente fórmula

$$\left[\bar{x} - \frac{s}{\sqrt{n}} t_{\alpha/2}, \bar{x} + \frac{s}{\sqrt{n}} t_{\alpha/2} \right] \quad (3.20)$$

siendo $t_{\alpha/2}$ el valor de la variable aleatoria de distribución t-student con probabilidad $1 - \alpha$, igual a la probabilidad del intervalo que se quiere, en este caso en cuestión 95 %. Obviamente para usar dicho intervalo se supone la hipótesis de normalidad al ser muy elevado el número de observaciones consideradas.

-Representamos tanto la media como el intervalo de confianza para cada uno de los indicadores, considerando las distintas distribuciones. Se realiza este paso dos veces: una con la Normal como realidad y otra con la Gumbel.

Los resultados de las mismas se muestran a continuación:

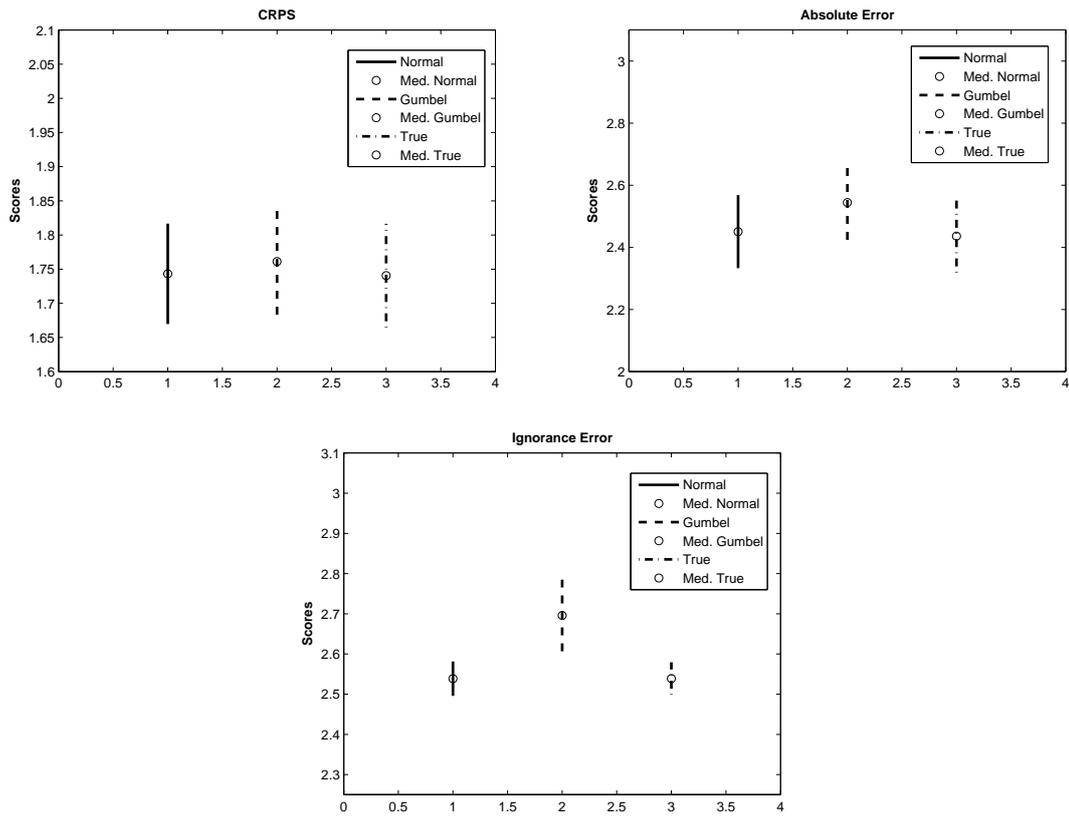


Figura 3.10 Media de los Scores con un intervalo de confianza del 95 %. Caso Normal Truth.

Tabla 3.1 Resultados de las medias de los diferentes Scores según las tres distribuciones consideradas. Caso Normal Truth.

NORMAL TRUTH	CRPS	AE	IGN
Normal	1.7432	2.4508	2.5386
Gumbel	1.7612	2.5446	2.6958
True	1.7406	2.4363	2.5390

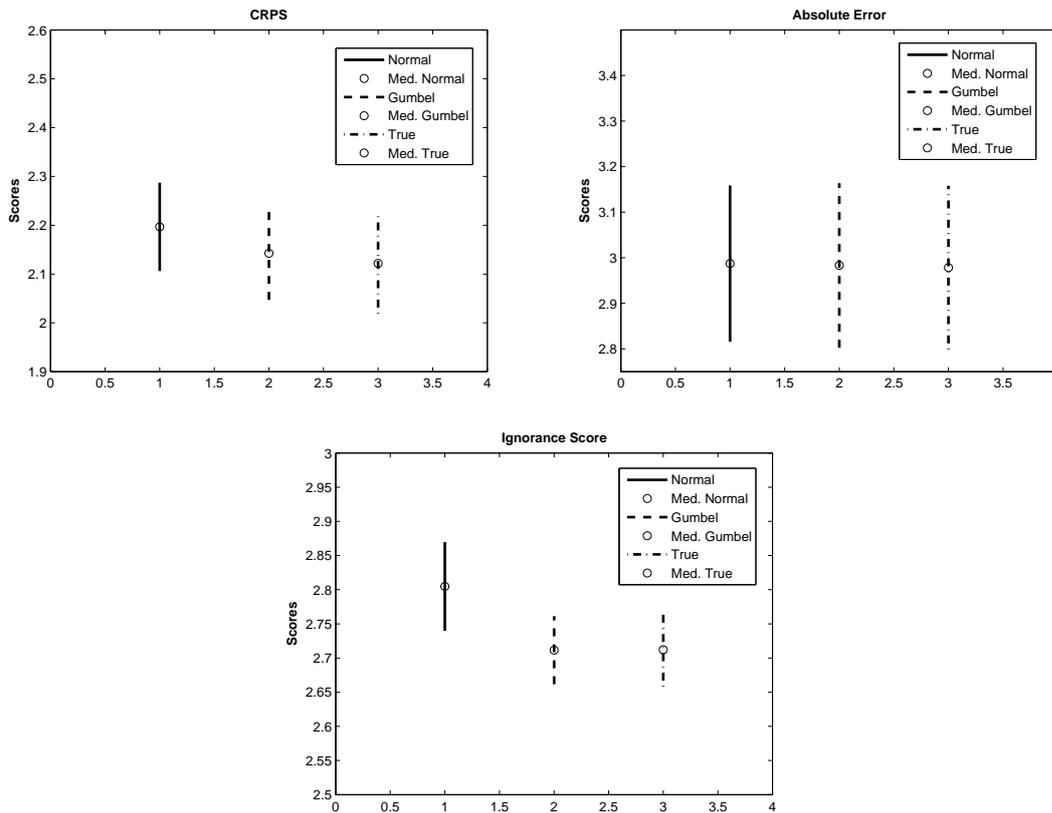


Figura 3.11 Media de los Scores con un intervalo de confianza del 95 %. Caso Gumbel Truth.

Tabla 3.2 Resultados de las medias de los diferentes Scores según las tres distribuciones consideradas. Caso Gumbel Truth.

GUMBEL TRUTH	CRPS	AE	IGN
Normal	2.1966	2.9874	2.7919
Gumbel	2.1424	2.9831	2.7154
True	2.1217	2.9784	2.7095

De estos resultados llaman la atención varias cosas. Por un lado, si la distribución considerada es la que se corresponde con la distribución real, el valor de la media de los scores será teóricamente más pequeño. Sin embargo, se puede ver que la diferencia entre la media de cada una de las distribuciones consideradas para cada indicador en la mayoría de los casos es pequeña. Esto puede inducir a pensar que usando un modelo u otro su desviación va a ser muy pequeña con respecto a la real. Además, el intervalo de confianza podría hacer que se obtengan valores menores de los Scores para una distribución que en realidad se sabe que no es la que mejor ajusta la realidad. También hay que pensar que estas gráficas han sido obtenidas considerando que las observaciones de entrada para cada distribución son las mismas. Como consecuencia, es importante tener en cuenta que puede haber casos realizados con distintas muestras que den resultados mejores para un modelo que se sabe que en realidad se ajusta peor. La única forma de acabar con los inconvenientes es que el número de observaciones sea muy elevado, por ejemplo 1 millón, cosa que en la realidad muchas veces no es posible.

3.4 Post-procesado estadístico de modelos de predicción

El contenido de esta sección se basa en desarrollar técnicas que permitan realizar la calibración o post-procesado de los modelos en bruto proporcionados por el ECMWF. En primer lugar, se explica el EMOS y su aplicación a los modelos de una variable. A continuación, se extiende el desarrollo a múltiples variables teniendo en cuenta los problemas que surgen al aplicar el EMOS por sí sólo. Finalmente, se dan pautas de cómo se pueden paliar de una forma lógica dichos inconvenientes.

3.4.1 Non-homogeneous regression o ensemble model output statistics (NR/EMOS)

Una vez vistas todas las herramientas necesarias para comprender el EMOS, se está en condiciones de proceder a su explicación para una posterior implementación numérica. Este método parte de un modelo de predicción que considera 50 escenarios distintos (en [2]). Estos escenarios se han obtenido con distintas condiciones iniciales. Aunque parezca que la idea de considerar varios casos distintos es muy acertada, resulta sorprendente que muchas veces las observaciones están incluso fuera de nuestro rango de predicción. Este problema necesita una resolución, ya que no se puede establecer modelos de predicción para ajustar observaciones dentro de un intervalo, y luego a la hora de realizar la medición de las mismas, éstas estén contenidas fuera del mismo.



Figura 3.12 Ejemplo gráfico de lo descrito en el párrafo anterior. En azul, se muestran las diferentes predicciones; en rojo su valor medio; y en verde un valor posible de una observación del día posterior. Se puede ver claramente la necesidad de calibrar el modelo para que la observación no caiga fuera del rango de las predicciones.

Una técnica que se puede usar para corregir nuestro modelo en bruto es realizar transformaciones del mismo, de manera que el modelo resultante se ajuste a las observaciones medidas en la realidad. Por tanto, se considera que la variable de predicción es una función paramétrica de nuestro ensemble forecast, es decir del conjunto de pronósticos que se han tomado para la variable

$$y|X_1, \dots, X_m \sim f(y|X_1, \dots, X_m) \quad (3.21)$$

donde f es una función de densidad de probabilidad de la variable y , condicionada por nuestro conjunto de pronósticos.

Suponemos ahora que las variables meteorológicas son esencialmente normales, una hipótesis que es bastante acertada. Por tanto, su función de densidad vendrá dada por su media μ y por su varianza σ^2 de la forma $\mathcal{N}(\mu, \sigma^2)$. En caso de que la distribución no fuera normal podría hacerse un desarrollo similar al descrito, cambiando los parámetros considerados por los de la distribución de densidad de probabilidad en cuestión.

Los parámetros de la normal pueden escribirse en función de una serie de coeficientes y del

modelo en bruto de la siguiente forma (en [1]):

$$\mu = a + b_1X_1 + \dots + b_mX_m \quad (3.22)$$

$$\sigma^2 = c + dS^2 \quad (3.23)$$

donde los coeficientes $a, b_1 \dots b_m, c$ y d son una serie de pesos que habrá que calcular siguiendo un cierto razonamiento lógico, y de cuyo carácter se hablará a continuación. La varianza muestral se define como

$$S^2 = \frac{1}{m} \sum_{i=1}^m (X_i - \tilde{X})^2 \quad (3.24)$$

con \tilde{X} la media muestral, $\tilde{X} = \frac{1}{m} \sum_{i=1}^m X_i$.

Para estimar la varianza se ha usado un modelo lineal $c + dS^2$, en el cual el coeficiente d , nos da una idea del error en la dispersión de nuestro modelo corregido con respecto a nuestro modelo de predicción original. A mayor coeficiente, obviamente mayor desviación. Los coeficiente a y c se utilizan para tener cierta capacidad de realizar modificaciones para cuadrar la media y la varianza en los límites deseados. El resto de coeficientes: $b_1 \dots b_m$, sirven para proporcionarnos una relación de cuál de los modelos "pondera más" o tiene una mayor contribución en la media de nuestra distribución corregida.

Como se ha indicado anteriormente, la dificultad consiste en elegir cuáles tienen que ser los coeficientes que mejor corrigen nuestro modelo de predicción, de manera que al corregirlo se ajuste de la mejor manera, a las observaciones de la realidad. Para realizar ese cometido, se necesita llevar a cabo un conjunto de entrenamiento, el cual va a proporcionar los valores adecuados de los coeficientes en cuestión. En secciones posteriores se estudiará cómo utilizando un conjunto de entrenamiento, se puede calibrar la predicción de partida. La respuesta se basará en la minimización del CRPS, opción que se desarrollará posteriormente.

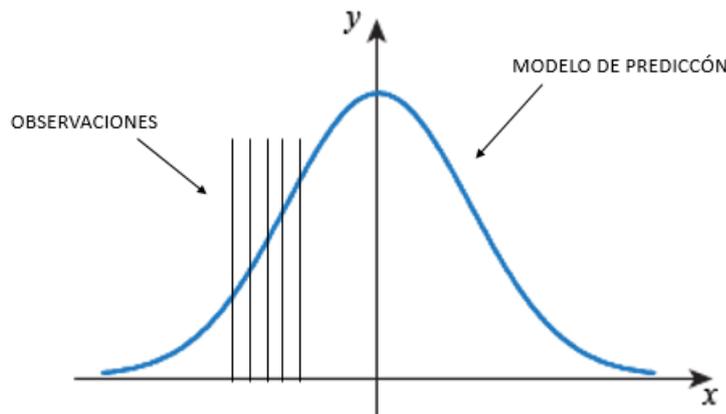


Figura 3.13 La figura muestra un modelo de predicción, en este caso continuo, que hay que ajustar para que se corresponda con las observaciones realizadas.

3.4.2 Calibración de modelos de una variable

El objetivo de esta sección es dar respuesta a los problemas planteados hasta ahora. Para ello, se aplicará el EMOS junto con los conceptos descritos en apartados anteriores.

Antes de empezar se definirá algunas de las variables a utilizar durante los desarrollos que se detallan a continuación. En primer lugar, se llamará i a cada uno de los casos o instancias consideradas, $i \in \{1 \dots N\}$, y j , al número de escenarios de los que consta la predicción $j \in \{1 \dots M\}$. Para entender que significa cada variable, considérese el siguiente ejemplo. Se quiere predecir cuál es la temperatura mañana en una cierta zona. Los datos históricos de los que se disponen son 4 valores de temperatura que marcó el termómetro justo en los días anteriores del que se quiere predecir la temperatura. Por tanto, el número de casos de los que se disponen son esos 4 valores, $N = 4$. A su vez, también se sabe que cada uno de los días de los que nos proporcionan datos, se contemplaron 8 posibilidades de valores de temperatura para cada día. Cada una de esas posibilidades es un escenario distinto de temperatura, $M = 8$.

Aclarado el significado de cada uno de los índices, se procede a desarrollar el método a emplear para la calibración de una variable. El modelo en bruto viene dado por $X_{i,j}$ donde j se corresponde con cada escenario e i con el caso equivalente a dicha predicción. Para calibrar, se necesitan una serie de observaciones, Y_i , tantas como datos históricos se quiera contemplar en la calibración. A partir de dichos datos, los pasos a realizar se resumen en los siguientes:

1. Se calcula para cada caso o instancia la media, \tilde{X}_i , y la varianza muestral, S_i^2 , para todos los escenarios. Se definen de la siguiente forma:

$$\tilde{X}_i = \frac{1}{M} \sum_{j=1}^M X_{i,j} \quad (3.25)$$

$$S_i^2 = \frac{1}{M} \sum_{j=1}^M (X_{i,j} - \tilde{X}_i)^2 \quad (3.26)$$

2. Se supone que cada caso se distribuye como una normal, $Z_i \sim \mathcal{N}(a + b\tilde{X}_i, c + dS_i^2)$. Aplicando esto a cada uno de los casos queda lo siguiente:

$$Z_1 \sim \mathcal{N}(a + b\tilde{X}_1, c + dS_1^2) \quad (3.27)$$

$$\vdots \quad (3.28)$$

$$Z_i \sim \mathcal{N}(a + b\tilde{X}_i, c + dS_i^2) \quad (3.29)$$

$$\vdots \quad (3.30)$$

$$Z_N \sim \mathcal{N}(a + b\tilde{X}_N, c + dS_N^2) \quad (3.31)$$

Cabe destacar que a diferencia del EMOS que considera los coeficientes $a, b_1 \dots b_m, c, d$ distintos para cada escenario, nuestro estudio asumirá que todos los coeficientes $b_1 \dots b_m$ son iguales. Esto puede argumentarse porque dichos escenarios han sido generados de la misma forma mediante una perturbación de las condiciones iniciales. Es decir si los escenarios son intercambiables se necesita asumir que $b_1 = \dots = b_m$ (Roman Schefzik en [8]). Se denotarán por b de aquí en adelante.

3. Siguiendo la expresión (3.12), se puede calcular el valor del CRPS para cada uno de los casos, suponiendo que se distribuyen a través de las funciones Z_i anteriores. Los resultados son cada uno de los valores del $CRPS(Z_i, Y_i)$. Una vez obtenido cada uno de los valores del CRPS para cada caso, se realiza la media de todos ellos $\xi = \frac{1}{N} \sum_{i=1}^N CRPS(Z_i, Y_i)$.

Es hora de hacer una recapitulación hasta el paso 3. Se ha partido teniendo como variables de entrada a $X_{i,j}, Y_i$ y a los coeficientes a, b, c, d ; y se ha obtenido el valor de ξ , que representa el valor medio del CRPS para todos los casos.

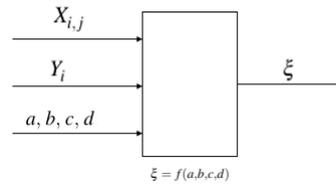


Figura 3.14 Variables de entrada y salida en la calibración de modelos de una variable.

Como $X_{i,j}$ e Y_i son datos de los que tenemos constancia, ξ será una función de cada uno de los coeficientes, $\xi = f(a,b,c,d)$.

4. Por tanto, el siguiente paso será buscar cuáles son los coeficientes a^* , b^* , c^* , d^* que minimizan ξ . El objetivo de la minimización de este score es buscar los parámetros que hacen que cada distribución en promedio se parezca lo máximo posible a las observaciones que se tienen en ese caso concreto. De esta manera, se ajustará mucho mejor el modelo en bruto a las observaciones de la realidad. Para resolver el problema de optimización de varias variables es necesario imponer unas condiciones iniciales y unas restricciones a los coeficientes. Como condición inicial, se tomará el modelo sin corregir, es decir $a = 0$, $b = 1$, $c = 0$ y $d = 1$. En cuanto a las restricciones, hay que imponer que $b \geq 0$ y $d \geq 0$, ya que un coeficiente negativo supondría un cambio en la tendencia del escenario al calibrar. Puede entenderse si se vuelve al ejemplo anterior de las temperaturas en una zona. Si mediante la predicción, la media de temperaturas para el caso 1 es la mayor (por ejemplo es la que predice más calor de todas), tras calibrar ponderando con un coeficiente b , el término $b\tilde{X}_i$; jamás podrá ser menor para el caso 1 que para el resto de casos. Es decir, no puede ser que la calibración invierta la tendencia de la predicción (sea el que predice más frío de todos). Lo mismo ocurre con la varianza muestral, y por tanto con el coeficiente d . El coeficiente a en principio servirá para corregir la media en sesgo y podrá tener cualquier valor. Además, por definición la varianza muestral tiene que ser positiva, por lo que según lo dicho, el coeficiente c podría ser negativo, siempre y cuando se respete que $c + dS_i^2 \geq 0$. Esto implica, introducir en la optimización a su vez la varianza muestral cosa que puede complicar el problema. Con el objeto de simplificar se asumirá que el coeficiente $c \geq 0$, ya que según la literatura no suele dar muchos problemas ([1]).
5. Una vez hecho el entrenamiento hay que calcular lo más importante del método. No cabe olvidar que el objetivo de este método era el siguiente: dada una predicción de cualquier variable en una zona para un día posterior, tratar de encontrar cuál debe ser la nueva predicción de manera que sea más realista. Llamaremos \hat{X}_j a la predicción del día o del caso siguiente. Ahora ya no disponemos de ninguna observación, porque es eso lo que se quiere predecir. Hay que realizar la hipótesis de que el conjunto de coeficientes de entrenamiento calculados anteriormente a^* , b^* , c^* , d^* , se mantiene para el día siguiente o para el caso siguiente. Siendo así, la probabilidad de que la temperatura sea una determinada viene dada por la siguiente distribución $\hat{Z} \sim \mathcal{N}(a^* + b^*\hat{X}, c^* + d^*\hat{S}^2)$.

Finalmente hay que comentar algunas de las ventajas e inconvenientes de este método tomando como base las suposiciones que se han hecho.

-Basar el método en el EMOS supone considerar modelos normales cuyos parámetros varían en función de coeficientes lineales. Esto es bastante simple, pero nos sirve para una primera aproximación al problema. En la siguiente tabla se recogen las distribuciones que en la práctica se deben escoger según el tipo de variable que se esté estudiando. En este proyecto, con objeto de simplificar, sólo se considerarán funciones normales.

Tabla 3.3 Distribuciones según variables. Tabla basada en [8].

VARIABLE	RANGO	DISTRIBUCIÓN
Temperatura	$y \in \mathbb{R}$	Normal
Presión	$y \in \mathbb{R}$	Normal
Cantidad de precipitación	$y \in \mathbb{R}^+$	Valor extremo
Componentes del viento	$y \in \mathbb{R}$	Normal
Velocidad del viento	$y \in \mathbb{R}^+$	Normal Truncada

-La simplicidad de este método implica realizar muchas optimizaciones que para grandes conjuntos de variables puede ser problemático.

-La suposición de que al transcurrir un período de tiempo, los coeficientes óptimos se mantienen constantes es sólo válido siempre que se desprece la evolución de la atmósfera con el tiempo. Como se sabe que la atmósfera siempre está en continuo cambio, y aunque éste sea lento, se tienen que realizar nuevos procesos de entrenamiento a lo largo del tiempo.

-No se tiene en cuenta en este trabajo la fiabilidad del método según el número de casos, N , que se considere para el conjunto de entrenamiento (véase [1]). Lo único que se realizará será una comparación entre dos números de casos diferentes, de la que se sacarán conclusiones, pero nunca con carácter general.

-De todos los scores utilizados para minimizar, se ha utilizado el CRPS por ser el más completo de los tres. Además las estimaciones de la media muestral y de la varianza muestral son las utilizadas para el Método de los Momentos, mientras que el IGN da mejores resultados si se utiliza el Método de Máxima Verosimilitud (véase subsección 3.3.3). También, se puede apreciar en las expresiones (3.12 y 3.18) que el CRPS es proporcional al error $\frac{y-\mu}{\sigma}$, mientras que el Ignorance Score es proporcional a su cuadrado, $(\frac{y-\mu}{\sigma})^2$. Por otra parte, el AE es bastante simple y no se ajusta a la forma de las observaciones.

Llegados a este punto nos parecería que todo está resuelto de cierta forma lógica y con cierto sentido físico. Sin embargo, muchas preguntas e inquietudes pueden salir a la luz a la hora de calibrar los modelos. En la formulación del método anterior las observaciones se han medido en un instante determinado y en una localización concreta. Es necesario para ciertas ramas de la ciencia como la meteorología, la comunicación o la ingeniería establecer mapas meteorológicos espaciales de las diferentes variables, como la presión, velocidad, temperatura etc. Para llevar a cabo los mapas, hay que establecer una serie de estaciones de medición, que dependerán de cada país o región, y que estarán localizadas en un entorno determinado. La pregunta que a uno se le viene a la cabeza es ¿cómo se puede calcular o predecir el valor de una determinada variable en una región, conociendo sólo el valor en ciertos puntos de la misma?.

Visto lo anterior, se podría realizar una minimización del CRPS para cada punto del espacio considerado y a partir de ahí, sacar una función de probabilidad para la variable en cuestión. Si se hace esto para cada punto, luego se sacará un valor correspondiente a una probabilidad (por ejemplo: el viento en este punto vale 7 m/s con una probabilidad del 99%). El problema surge si los valores que se elijan entre dos puntos relativamente cercanos difieran mucho, ya que supondría no sólo un valor surrealista de la variable si no que la interpolación para las regiones comprendidas entre los dos puntos arrastraría un mayor error y se perdería toda la potencia de cálculo del método. Lo razonable y por otra parte lo buscado es un modelo en que no haya muchas variaciones bruscas

entre puntos cercanos del espacio de manera que se pueda aceptar físicamente.

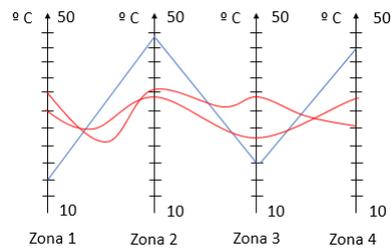


Figura 3.15 La figura muestra los diferentes modelos de temperaturas de 4 zonas cercanas, así como la temperatura para zonas intermedias. En azul vemos un modelo obtenido suponiendo que la interpolación se ha realizado a través de una minimización del CRPS en cada punto de manera independiente. En rojo, se aprecian modelos más suaves de temperatura que tienen más sentido físico. Los modelos se han exagerado mucho para poder entender el error que existe al aplicar el método descrito en esta sección.

A raíz de lo comentado, se introduce el concepto de la calibración de modelos de varias variables, que nos servirá para dar sentido al problema propuesto anteriormente. Este concepto se tratará con detenimiento en la sección posterior.

3.4.3 Calibración de modelos de varias variables

Introducción

Como se comentó en la sección anterior, se pueden cometer ciertos errores importantes a la hora de calibrar los modelos simplemente minimizando el CRPS en cada estación de medición de forma independiente, y luego interpolando para zonas intermedias. Para solucionar este problema rápidamente, uno puede pensar en colocar muchas estaciones de medición de variables por todas las zonas posibles y así reducir ese error de interpolación. Sin embargo, eso no es posible ya que un aumento de estaciones de medición supondría un aumento considerable de variables del problema, que repercutiría en un incremento del tiempo de cálculo que haría inabordable el problema. No sólo eso sería el único inconveniente, si no que el hecho de incrementar el número de estaciones en una región supone mucha más inversión, y por tanto, un mayor coste que depende en cierta medida de decisiones políticas, sociales, técnicas o legales en algunos casos. Lo segundo que uno puede proponerse es intentar hacer que los valores en las regiones intermedias cumplan las ecuaciones de Navier-Stokes para los fluidos o la de los campos de temperatura etc. La complejidad que aportaría meter de manera explícita toda la física del problema para reducir la incertidumbre no sólo dificultaría encontrar una solución aceptable, si no que posiblemente, esa solución llegaría más tarde del día en el que nosotros necesitamos conocer dicha predicción. Traduciendo el problema a lenguaje matemático, la covarianza entre las diferentes variables se pierde a la hora de calibrar. Es decir, si la relación entre la predicción de dos puntos es lineal, tras calibrar, deja de serlo, pudiendo ser otra cualquiera. Eso es algo que la calibración de modelos de una variable por si sola no puede resolver.

Dicho esto, se necesita buscar otro método basado en los anteriores que nos permita reducir esos errores, teniendo en cuenta cómo se ha obtenido el modelo en bruto original. Para ello, la calibración de varias variables puede ser una solución más que fructífera. En el subapartado siguiente se estudiará un método muy utilizado llamado *Ensemble Copula Coupling (ECC)*. La base del planteamiento consiste en integrar todas las muestras de las diferentes estaciones en un optimizador

cuyo resultado recoja de cierta manera por decirlo así, la covarianza de todas las zonas de medición entre sí, permitiendo obtener un modelo con cierta claridad física. Su desarrollo en profundidad se analiza a continuación.

Ensemble Copula Coupling (ECC)

Hablar de Ensemble Copula Coupling, es lo mismo que hablar de postprocesado de múltiples variables. Para entender la forma en la que el ECC integra todas las variables independientes en un mismo paquete, se estructurará su descripción en una serie de pasos que servirán de aclaración de cara a una posible programación futura.

Se denotará ahora por el subíndice $k \in \{1 \dots K\}$ a las distintas variables del problema. Por variable no sólo se entiende a valores de magnitudes en distintos puntos del espacio, si no que también puede referirse a magnitudes diferentes en un mismo punto, como pueden ser el viento zonal, el viento meridional, la temperatura, etc.

En primer lugar, el proceso consiste en repetir los pasos del 1 al 4 del método anterior para todas las variables de forma independiente. Se consideran como variables de entrada, la predicción del modelo en bruto, $X_{i,j,k}$, las observaciones, $Y_{i,k}$ y los coeficientes a_k, b_k, c_k, d_k . Como salida se obtiene el CRPS medio para cada variable, ξ_k . Se buscan cuáles son los coeficientes $a_k^*, b_k^*, c_k^*, d_k^*$ óptimos que minimizan cada uno de los ξ_k .

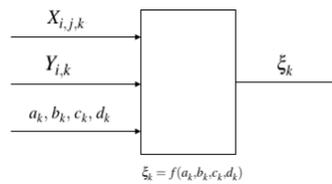


Figura 3.16 Variables de entrada y salida en el ECC.

Los siguientes pasos a dar se explican de manera detallada a continuación:

- 5. Se procede a calcular el *Multivariate Rank* para cada uno de los casos, i y para cada una de las variables, k . Lo primero que hay que hacer es ordenar de menor a mayor los valores de cada uno de los escenarios para los diferentes casos o instancias

$$X_{1,(1)} \leq X_{1,(2)} \leq \dots \leq X_{1,(M)} \tag{3.32}$$

$$\vdots \tag{3.33}$$

$$X_{i,(1)} \leq X_{i,(2)} \leq \dots \leq X_{i,(M)} \tag{3.34}$$

$$\vdots \tag{3.35}$$

$$X_{N,(1)} \leq X_{N,(2)} \leq \dots \leq X_{N,(M)} \tag{3.36}$$

donde $()$ denota los valores ordenados. Todo ello, se repite para las K variables de estudio.

- 6. Entre los numerosos Multivariate Rank que existen se elegirá el *Multivariate Rank Average* por simplicidad. Este dice que la solución es construir una tabla de ordenación, en la cual se establezca para cada uno de los casos, el nuevo orden resultante de cada escenario tras la reordenación. En una de las columnas se calcula además la media de todos los órdenes para cada uno de los casos considerados.

Tabla 3.4 Tabla de reordenación de las muestras.

ESCENARIO/ORDEN	$i = 1$	$i = 2$...	$i = N$	MEDIA
$j = 1$	2	7	4.33
$j = 2$	1	5	5.2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$j = M$

7. Se ordenan los escenarios de menor a mayor según el valor obtenido en la columna MEDIA. En caso de que dos escenarios incurran en la misma media, se supondrá que es mayor aquel con mayor índice j . El resultado se puede escribir en otra nueva tabla de la siguiente forma:

Tabla 3.5 Orden global de todos los escenarios.

ESCENARIO	NUEVO ORDEN
1	5
2	9
\vdots	?
M	?

8. En concordancia con el punto 5 de la calibración de una variable, la predicción para cada una de las variables para el día o para el caso siguiente será: $\hat{X}_{j,k}$. Por tanto, suponiendo que los coeficientes a_k^* , b_k^* , c_k^* , d_k^* se mantienen, se pueden construir las siguientes funciones de probabilidad: $\hat{Z}_k \sim \mathcal{N}(a_k^* + b_k^* \hat{X}_k, c_k^* + d_k^* \hat{S}_k^2)$, para cada una de las K variables en cuestión.
9. De cada uno de los modelos calibrados \hat{Z}_k , se extraen muestras que sean equiprobables. Por ejemplo, siguiendo la siguiente fórmula.

$$\hat{Z}_{(j),k} = \hat{Z}_k^{-1} \left(\frac{m}{M+1} \right) \quad (3.37)$$

con $m \in \{1 \dots M\}$.

Además como las muestras se han extraído de forma equiprobable, los valores obtenidos ya están ordenados de menor a mayor:

$$\hat{Z}_{(1),k} \leq \hat{Z}_{(2),k} \leq \dots \hat{Z}_{(M),k} \quad (3.38)$$

para todo $k \in \{1 \dots K\}$.

10. Finalmente, sólo hay que entrar para cada muestra, $\hat{Z}_{(j),k}$ en la columna NUEVO ORDEN de la Tabla 3.5, e identificarla con el escenario de partida, restableciéndose el orden original.

Comentar que la potencia de este método es que nos permite combinar, por una parte el hecho de que se minimiza el CRPS, ajustándose el modelo más a la realidad, y por otra parte que el reordenamiento de la muestra obtenida nos mantiene las relaciones del modelo original entre las distintas variables de cada región. Así, si por ejemplo, la temperatura o el valor del viento en Madrid siempre es superior al de Barcelona para un escenario concreto, después de calibrar, debe de continuar siéndolo para el mismo escenario, ya que si no se rompería con los principios físicos

del problema. Con sólo la minimización, esa relación se pierde cambiando totalmente los valores de dispersión entre variables. Con este planteamiento, los mapas de vientos, temperaturas o de cualquier otra variable serán más que aceptables para poder usarlos en la práctica.

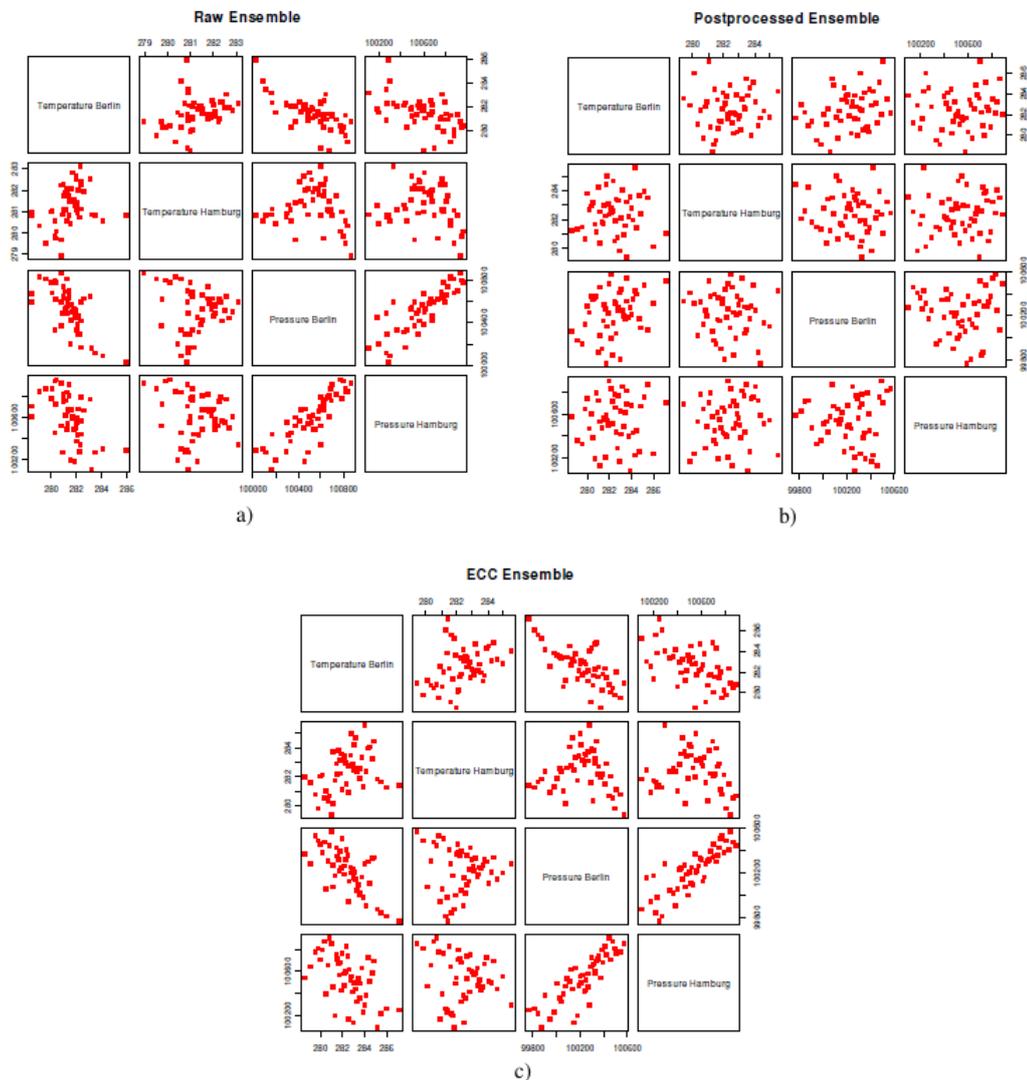


Figura 3.17 Relaciones de presión y temperatura entre Berlín y Hamburgo. Ilustraciones tomadas de [7].

En la figura anterior, se puede ver un claro ejemplo de la pérdida de la relación entre variables cuando se optimiza el CRPS. El caso a) muestra las relaciones originales entre las presiones y temperatura de dos puntos pertenecientes a Berlín y Hamburgo. En el b) se observa como la minimización del CRPS provoca una dispersión de los valores de dichas variables. La solución de esto es emplear el ECC como en el caso c), lo que nos permite mantener las relaciones originales a la misma vez que se consigue un modelo de predicción más realista.

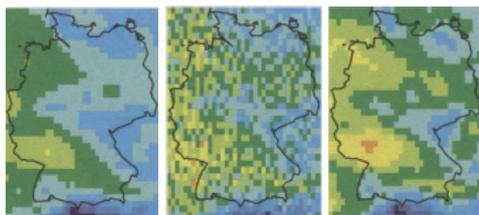


Figura 3.18 Mapas de temperaturas sobre Alemania. Ilustración tomada de [8].

Otro ejemplo quizás más significativo, es el de los mapas de temperatura de una determinada región. En la Figura 3.18, se observan tres mapas de temperaturas de Alemania. Cada mapa se corresponde con el modelo en bruto, el modelo tras calibrar y el resultado tras aplicar el ECC, respectivamente. Llama claramente la atención el segundo mapa, en el cual, la calibración hace que se pierda la relación entre las temperaturas de cada zona resultando un mapa mucho menos uniforme.

4 Predicción de trayectorias

El objetivo del presente capítulo es presentar al lector de una forma sencilla cómo se puede abordar la predicción de una trayectoria sujeta a determinadas restricciones.

4.1 Introducción

La predicción de una trayectoria es una de las aplicaciones más inmediata, pero a la vez más importante, para la utilización de los modelos de vientos, los cuales serán usados a lo largo de este capítulo. El estudio realizado de ahora en adelante, es muy útil para dar una primera aproximación de las variaciones que introducen en las actuaciones los vientos, que se sabe que están presentes en la realidad. Mediante consideraciones como la anterior se pueden determinar cuáles son los incrementos o reducciones en el tiempo que se tarda en recorrer la ruta determinada. Este es el objetivo primordial del capítulo ya que obteniendo esos incrementos con respecto al valor nominal se tiene una idea del combustible de más o de menos para volar que se necesita, repercutiendo a su vez en un coste mayor o menor. También desde el punto de vista operativo, conocer con más precisión el tiempo que tarda una aeronave en ir de un sitio a otro permite elaborar una planificación del tráfico mucho más acotada, por lo que puede servir de vía para acelerar la gestión del tráfico aéreo. No sólo se queda la cosa aquí en lo que respecta a aplicaciones. Uno puede pensar en resolver el problema inverso y preguntarse cuál debe ser la mejor ruta para ir de una región a otra con el menor consumo posible o tardando lo mínimo posible. A su vez, los modelos de viento utilizados para la trayectoria han sido predichos minimizando indicadores como el CRPS, descrito en el capítulo anterior.

La estructura de este capítulo será la siguiente: por una parte, se describirá cuáles son las hipótesis más importantes que simplifican la trayectoria, y por otra parte, se ilustrará, en base a dichas hipótesis, cuáles son las ecuaciones necesarias para su cálculo y cómo proceder si se quiere implementar el mismo. A su vez se hará hincapié durante todo el capítulo en las ventajas e inconvenientes que tienen las simplificaciones que se hacen, y cuáles son sus limitaciones al respecto.

4.2 Caracterización de la trayectoria

Para facilitar el estudio de la trayectoria completa la dividiremos en un número determinado de segmentos. Cada uno de los segmentos tendrá un punto inicial y un punto final, que vendrán dados por su longitud y por su latitud (λ, ϕ). Estos puntos se denominan *Waypoints* y servirán de apoyo para discretizar la trayectoria. Un waypoint es el resultado de la intersección de un paralelo con un meridiano terrestre. En nuestro estudio se supondrá la superficie terrestre como una cuadrícula delimitada por ϕ_{min} ϕ_{max} λ_{min} λ_{max} , con un paso de 1° en latitud y en longitud (según [9]).

A la hora de establecer las trayectorias que unen un punto de origen con un punto de destino, se trabajará con loxodrómicas. Esto es una ecuación de una curva terrestre del tipo $\phi = \phi(\lambda)$. Será muy útil poder conocer la distancia entre cualquiera dos puntos de la loxodrómica, así como el curso que hay que imponer para alcanzar el punto final. Esas dos magnitudes se pueden calcular a partir de la loxodrómica mediante desarrollos matemáticos (los cuales se omiten en este trabajo), obteniéndose:

$$\chi_{g_j} = \arctan \left(\frac{\lambda_j - \lambda_{j-1}}{\ln \left[\frac{\tan(\pi/4 - \phi_{j-1}/2)}{\tan(\pi/4 - \phi_j/2)} \right]} \right) \quad (4.1)$$

$$r_j \begin{cases} \frac{(R_E + h)(\phi_j - \phi_{j-1})}{\cos \chi_{g_j}} & \text{si } \phi_j \neq \phi_{j-1} \\ (R_E + h) \cos \phi_{j-1} |\lambda_j - \lambda_{j-1}| & \text{si } \phi_j = \phi_{j-1} \end{cases} \quad (4.2)$$

Estas expresiones no pueden usarse tal cual, ya que hay que distinguir varios casos problemáticos (en [10]). En primer lugar, si $|\lambda_j - \lambda_{j-1}| < \pi$ tenemos lo siguiente:

-Para $\lambda_j \geq \lambda_{j-1}$, si $\phi_j < \phi_{j-1}$ hay que corregir el curso con $\chi_g = \chi_g + \pi$.

-Para $\lambda_j < \lambda_{j-1}$, si $\phi_j < \phi_{j-1}$ hay que corregir con $\chi_g = \chi_g + \pi$; en caso contrario corregimos con $\chi_g = \chi_g + 2\pi$, para tener un curso entre 0 y 360 grados.

En cambio, si se da el caso de que $|\lambda_j - \lambda_{j-1}| > \pi$ podemos recorrer la tierra cruzando el meridiano 180° (por el camino corto). Las expresiones modificadas son:

-Para $\lambda_j < 0$, la expresión del curso es

$$\chi_{g_j} = \arctan \left(\frac{\lambda_j - \lambda_{j-1} + 2\pi}{\ln \left[\frac{\tan(\pi/4 - \phi_{j-1}/2)}{\tan(\pi/4 - \phi_j/2)} \right]} \right) \quad (4.3)$$

donde si $\phi_j < \phi_{j-1}$, se corrige el curso con $\chi_g = \chi_g + \pi$. Cuando el curso obtenido sale negativo se le vuelve a sumar 2π , para obtener uno entre 0 y 360 grados.

-Para $\lambda_j \geq 0$, la expresión del curso es

$$\chi_{g_j} = \arctan \left(\frac{\lambda_j - \lambda_{j-1} - 2\pi}{\ln \left[\frac{\tan(\pi/4 - \phi_{j-1}/2)}{\tan(\pi/4 - \phi_j/2)} \right]} \right) \quad (4.4)$$

donde si $\phi_j < \phi_{j-1}$, se corrige con $\chi_g = \chi_g + \pi$. Cuando el curso obtenido sale negativo, se le vuelve a sumar 2π , para obtener un curso entre 0 y 360 grados.

En vista a las ecuaciones presentadas, se inducen varias hipótesis que se presentan a continuación:

- Se asume que la Tierra es esférica de radio constante $R_E = 6371 \text{ km}$.
- Los segmentos que forman parte de la trayectoria serán segmentos rectos. Para simular segmentos curvos se tomarán muchos puntos en las zonas donde se quiera realizar una curva más pronunciada. Como consecuencia de lo anterior sólo se considerarán segmentos de crucero, pudiéndose hacer estos a diferentes altitudes h . Ningún segmento de subida, descenso, aterrizaje, despegue o viraje, se tendrá en cuenta, así como la transición entre distintos segmentos de crucero. Se supone que el tiempo y el consumo de combustible en dichos segmentos es mucho menor que en los segmentos de crucero, siendo esto más acertado cuanto mayor alcance tenga la misión.
- Se supondrá que no se puede volar justo en los polos donde el curso a tomar siempre tiende a cero, según las expresiones anteriores.

4.3 Ecuaciones de la trayectoria

En la figura se puede ver un ejemplo de trayectoria definida por un punto inicial y un punto final. Si se quiere ir de un waypoint a otro de la trayectoria, se tendrá que recorrer una cierta distancia r en línea recta, con un curso χ_g . La velocidad aerodinámica de la aeronave vendrá dada por $V = M a$, siendo M , el número de Mach al que se quiere volar y a , la velocidad del sonido a dicha altitud, $a = \sqrt{\gamma_a R_a T}$.

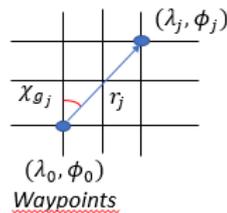


Figura 4.1 Ejemplo de segmento de trayectoria.

En cada punto de la cuadrícula definida en el apartado anterior se consideran tres variables: un valor de temperatura, T , y dos valores de vientos, un viento zonal que viene del este y otro meridional en dirección norte, w_u y w_v , respectivamente.

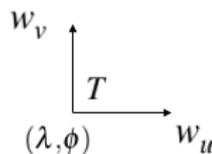


Figura 4.2 Vientos y temperaturas para cada punto de la malla.

Quizás uno de los asuntos más importantes del tema de cálculo de trayectorias es integrar los vientos en nuestro modelo de cálculo. La idea es descomponer el vector global de vientos, w , en dos componentes: una en la dirección de la trayectoria, w_l , y otra de viento cruzado, w_x (positivas

en la dirección de la figura). Esa descomposición es bastante simple, ya que es solamente un giro de ángulo el correspondiente al curso de la trayectoria de la aeronave. La transformación queda como

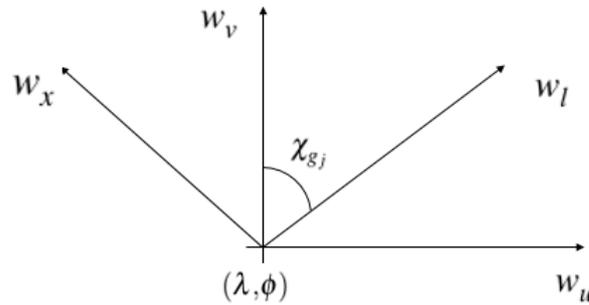


Figura 4.3 Cálculo del viento longitudinal y cruzado.

$$w_l = w_v \cos \chi_{gj} + w_u \sin \chi_{gj} \tag{4.5}$$

$$w_x = w_v \sin \chi_{gj} - w_u \cos \chi_{gj} \tag{4.6}$$

Por tanto, para poder volar con el curso de la trayectoria requerida, orientamos el rumbo de nuestra aeronave, χ , de manera que la composición del vector velocidad aerodinámica, V , y del viento w , coincida en módulo, dirección y sentido con la velocidad con respecto de tierra, V_g , la cual lleva la dirección del curso a seguir. Dicho esto, la velocidad con respecto de tierra puede calcularse mediante el siguiente triángulo, dando lugar al siguiente resultado.

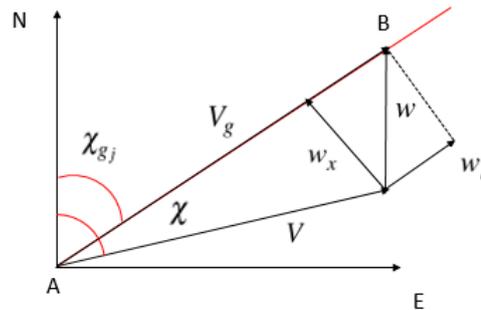


Figura 4.4 Triángulo de vientos.

$$V_g = \sqrt{V^2 - w_x^2} + w_l = \sqrt{(\gamma R_a T) M^2 - w_x^2} + w_l \tag{4.7}$$

El número de Mach se supondrá constante en cada segmento, y el resto de variables será función de cada waypoint y del tiempo: $T(\lambda, \phi, t)$ $w_l(\lambda, \phi, t)$ $w_x(\lambda, \phi, t)$. En este proyecto, no se considerará la dependencia de estas variables con el tiempo, por lo que se supondrá que el valor de las variables estará tomado en un instante de tiempo fijo. De aquí en adelante, se omitirá la dependencia del mismo en las ecuaciones. Como en las ecuaciones de la Mecánica del Vuelo, las variables que

aparecen son el tiempo y la distancia, r y t , hay que escribir la longitud y la latitud en función de r : $\lambda(r)$, $\phi(r)$. Para obtener estas expresiones, hay que invertir la loxodrómica de la siguiente forma

$$\phi = \phi_{j-1} + \frac{r}{R_E + h} \cos \chi_{g_j} \quad (4.8)$$

$$\lambda = \lambda_{j-1} + \tan \chi_{g_j} \ln \left[\frac{\tan(\pi/4 - \phi_{j-1}/2)}{\tan(\pi/4 - \phi/2)} \right] \quad (4.9)$$

teniendo en cuenta que si el curso es $\chi_g = 90^\circ$ o $\chi_g = 270^\circ$, estamos recorriendo un paralelo $\phi_j = \phi_{j-1}$, por lo que la ecuación de la longitud hay que sustituirla por esta otra:

$$\lambda = \lambda_{j-1} + \frac{r \sin \chi_{g_j}}{(R_E + h) \cos \phi_j} \quad (4.10)$$

Según las simplificaciones anteriores, la ecuación de la trayectoria viene dada según la Mecánica del Vuelo como:

$$\frac{dr}{dt} = V_g \quad (4.11)$$

pero como lo que nos interesa es el tiempo de vuelo, no tenemos más que invertir la ecuación anterior para obtenerlo.

$$\frac{dt}{dr} = \frac{1}{V_g} \quad (4.12)$$

El tiempo total de la trayectoria será la suma de los tiempos que se tarda en recorrer cada uno de los segmentos.

$$t_{vuelo} = \sum_{j=1}^N t_j \quad (4.13)$$

4.4 Resolución Numérica de la trayectoria

El objetivo de este apartado es dar al lector unas nociones y/o instrucciones necesarias a tener en cuenta a la hora de implementar numéricamente la trayectoria. Los siguientes pasos resumen lo más destacado para realizar su cálculo numérico, una vez vistas las ecuaciones del apartado anterior. En primer lugar se considera que sólo se disponen de datos de un escenario de viento concreto. A continuación, lo que hay que realizar es:

1. Se crea una malla en ϕ y en λ , sabiendo que el ϕ_{min} λ_{min} ϕ_{max} λ_{max} son las coordenadas de los puntos que tengan menor latitud y longitud, y mayor latitud y longitud de la trayectoria respectivamente. El paso de esta malla será 1° . Si por simplicidad, se quiere trabajar con números enteros, se puede tomar 1° por debajo o por encima de las coordenadas anteriores. El número de puntos será el suficiente para que la malla no sea demasiado gruesa, pudiéndose modificar los límites de la misma si fuese necesario.
2. Se calcula el χ_{g_j} y el r_j para cada uno de los segmentos según la loxodrómica.
3. En cada punto de la malla se conocen las coordenadas del viento zonal y del meridional, w_u w_v . Por tanto, como en la ecuación (4.12) la variable independiente es la distancia recorrida, se necesita saber cual es el valor del viento en puntos intermedios de la malla. Ese valor puede obtenerse mediante una interpolación lineal en 2D. Su cálculo puede realizarse si se conocen los valores de las coordenadas del punto a interpolar (expresiones 4.8 4.9 4.10).

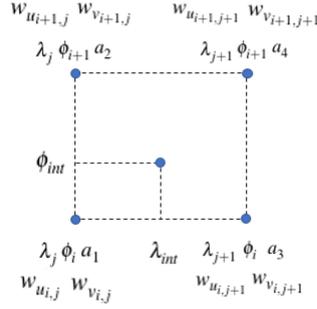


Figura 4.5 Interpolación de los valores del viento en la malla.

Para la interpolación, se necesitan saber los valores de las siguientes constantes:

$$a_1 = \frac{(\lambda_{j+1} - \lambda_{int})(\phi_{int} - \phi_{i+1})}{(\lambda_{j+1} - \lambda_j)(\phi_i - \phi_{i+1})} \quad (4.14)$$

$$a_2 = \frac{(\lambda_{j+1} - \lambda_{int})(\phi_i - \phi_{int})}{(\lambda_{j+1} - \lambda_j)(\phi_i - \phi_{i+1})} \quad (4.15)$$

$$a_3 = \frac{(\lambda_{int} - \lambda_i)(\phi_{int} - \phi_{i+1})}{(\lambda_{j+1} - \lambda_j)(\phi_i - \phi_{i+1})} \quad (4.16)$$

$$a_4 = \frac{(\lambda_{int} - \lambda_i)(\phi_i - \phi_{int})}{(\lambda_{j+1} - \lambda_j)(\phi_i - \phi_{i+1})} \quad (4.17)$$

En el caso de que el paso de la malla sea unidad, por ejemplo, 1° , las expresiones anteriores se reducen a:

$$a_1 = (1 - \rho_\lambda)(1 - \rho_\phi) \quad (4.18)$$

$$a_2 = \rho_\lambda(1 - \rho_\phi) \quad (4.19)$$

$$a_3 = (1 - \rho_\lambda)\rho_\phi \quad (4.20)$$

$$a_4 = \rho_\lambda\rho_\phi \quad (4.21)$$

con

$$\rho_\lambda = \lambda_{int} - \lambda_i \quad (4.22)$$

$$\rho_\phi = \phi_{int} - \phi_i \quad (4.23)$$

A partir de las constantes, los valores del viento interpolados en la malla se obtienen mediante una combinación de las mismas.

$$w_{u,int} = a_1 w_{u_{i,j}} + a_2 w_{u_{i+1,j}} + a_3 w_{u_{i,j+1}} + a_4 w_{u_{i+1,j+1}} \quad (4.24)$$

$$w_{v,int} = a_1 w_{v_{i,j}} + a_2 w_{v_{i+1,j}} + a_3 w_{v_{i,j+1}} + a_4 w_{v_{i+1,j+1}} \quad (4.25)$$

4. Se realiza un giro de ángulo χ_{g_j} para obtener el viento cruzado y el viento longitudinal, w_x w_l .
5. Se resuelve la ecuación diferencial (4.12), empleando el comando *ode45* en Matlab con la condición inicial $t_j(0) = 0$.

6. Por último, para hallar el tiempo de vuelo total, simplemente se suman los tiempos de los diferentes segmentos de los que consta la trayectoria.

$$t_{vuelo} = \sum_{j=1}^N t_j \quad (4.26)$$

En segundo lugar, hay que contemplar el caso en el que se tienen datos de diferentes escenarios de viento. Cada uno de los escenarios se denotará con el subíndice $i = \{1, 2 \dots N_{esc}\}$. Con ello, el tiempo de vuelo cambiará en función de los vientos que se hayan predicho que ocurrirán en cada punto, según el escenario en cuestión.

La solución más sencilla para resolver el problema de muchos escenarios es realizar una iteración del problema anterior cambiando el valor de los vientos zonales y meridionales, w_{u_i} w_{v_i} que ahora, son funciones del escenario del que se quiere hallar el tiempo de vuelo. El número de iteraciones será igual al número de escenarios total, N_{esc} . El resultado de esto es un vector de tiempos de vuelo correspondientes a cada uno de los escenarios, $t_{vuelo} = [t_{v_1} \ t_{v_2} \ \dots \ t_{v_i}]$.

Sin embargo, implementar numéricamente lo descrito en el párrafo anterior, puede suponer un problema desde el punto de vista computacional. Algunos comandos como el *ode45*, son computacionalmente muy costosos, por lo que no es bueno que el número de iteraciones a resolver sea muy elevado. Hacerlo de esta manera, implica que el número de ecuaciones diferenciales que tiene que resolver el programa son $N_{esc} \times N$, con N el número de segmentos de la trayectoria.

Para reducir el tiempo computacional, se puede escribir el problema como un sistema de ecuaciones diferenciales, en lugar de una ecuación diferencial

$$\left\{ \begin{array}{l} \frac{dt_1}{dr} = \frac{1}{V_{g1}} \\ \frac{dt_2}{dr} = \frac{1}{V_{g2}} \\ \vdots = \vdots \\ \frac{dt_{N_{esc}}}{dr} = \frac{1}{V_{g_{N_{esc}}}} \end{array} \right. \quad (4.27)$$

donde cada una de las variables se corresponde con el tiempo determinado para cada escenario. La condición inicial de este sistema será

$$\vec{t}(0) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4.28)$$

De esta forma, sólo habrá que resolver este sistema tantas veces como segmentos conste la trayectoria. Es decir, el número de *ode45* a resolver es ahora mucho más pequeño, igual a N .

5 Aplicación. Trayectoria sujeta a incertidumbre

Como su propio nombre indica, el presente capítulo ilustrará cómo se puede aplicar la calibración de los modelos y la predicción de trayectorias a una ruta determinada, con unos puntos de origen y destino específicos. Se desarrollará centrándose en los aspectos más importantes a la hora de su implementación numérica, así como en los resultados que deriven de ésta. El mayor interés del capítulo será estudiar las posibles diferencias entre los tiempos de vuelo de una aeronave con viento, sin viento, y con unos modelos de viento calibrados siguiendo las directrices anteriores. A lo largo del capítulo, se recurrirá a mencionar apartados de capítulos anteriores pero siempre dando una visión más específica y no tanto teórica, con el objetivo de entender las diferencias entre modelos. Sin más que hablar, se está en condiciones de empezar a desarrollar el capítulo.

5.1 Vuelo de estudio

La trayectoria que se va a considerar en el proyecto es una ruta que recorra el Océano Atlántico. Concretamente, su origen es Berlín (Alemania) y su destino Washington D. C. (EE.UU.). El vuelo no es directo, si no que se utilizará como waypoints ciertas capitales europeas y/o ciudades importantes como son París (Francia), Madrid (España), Lisboa (Portugal), Ponta Delgada (Azores, Portugal), Bermudas (Reino Unido) y Miami (Florida, EE.UU.). Las coordenadas tanto de latitud como de longitud de las diferentes ciudades, se recogen en la siguiente tabla.

Tabla 5.1 Coordenadas geográficas de los diferentes waypoints.

CIUDADES/COORDENADAS	λ	ϕ
1.BERLÍN	13° 24' 29" E	52° 31' 7" N
2.PARÍS	2° 19' 60" E	48° 52' 0" N
3.MADRID	3° 42' 12" O	40° 25' 0" N
4.LISBOA	9° 8' 84" O	38° 42' 26" N
5.PONTA DELGADA	25° 40' 0.8" O	37° 44' 46" N
6.BERMUDAS	64° 45' 0" O	32° 19' 60" N
7.MIAMI	80° 12' 38" O	25° 46' 30" N
8.WASHINGTON DC	77° 1' 2" O	38° 54' 15" N

Como datos adicionales se supondrá que la altitud y el número de Mach son constantes (la velocidad también lo es), a la que se realizan los cruceros es de $h = 12000\text{ m}$ y de $M = 0.8$. Se

desprecian los tramos de despegue, aterrizaje, subidas descensos etc, como ya se dijo en la sección 4.2. El recorrido de vuelta será el mismo que el de ida pero al contrario, siendo la velocidad y el número de Mach idénticos, con el objetivo de comparar los diferentes tiempos de vuelo. Las distancias y los cursos necesarios para llegar de un punto a otro de la trayectoria, pueden verse en la siguiente tabla.

Tabla 5.2 Distancias y cursos entre los diferentes waypoints tanto en ida como en vuelta.

	1-2/2-1	2-3/3-2	3-4/4-3	4-5/5-4	5-6/6-5	6-7/7-6	7-8/8-7
r (km)	880.8	1055.2	504.5	1450.5	3612.8	1671.9	1491.9
$\chi_{g,ida}$	242.50°	206.86°	247.81°	265.77°	260.38°	244.12°	11.56°
$\chi_{g,vuelta}$	62.51°	26.86°	67.81°	85.77°	80.38°	64.12°	191.56°

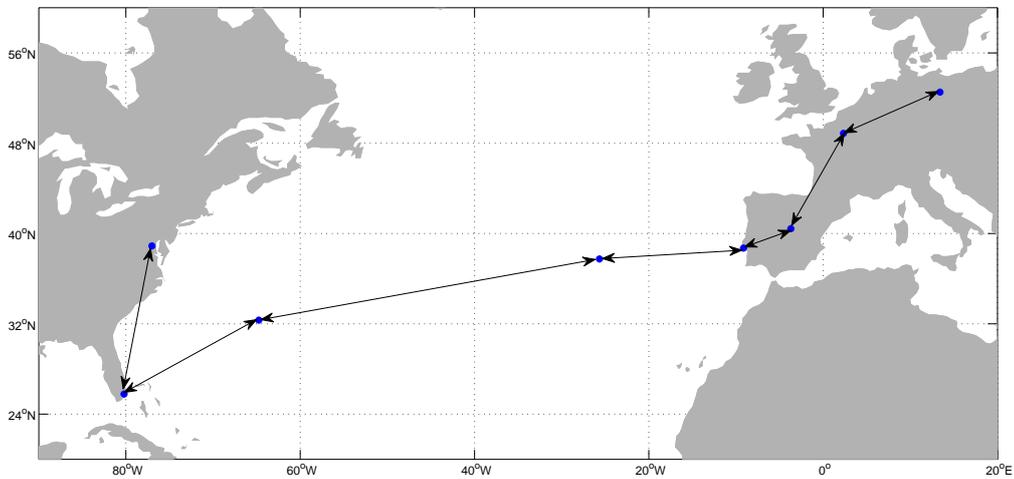


Figura 5.1 Representación de la trayectoria en proyección cilíndrica equidistante.

Será importante para la calibración definir el día y la hora en la que se desea recorrer la trayectoria. En nuestro caso, la trayectoria será recorrida el día 15 de Diciembre del año 2017 a las 00:00 UT. Por tanto, el ECMWF proporciona un conjunto de predicciones para ese día, que contienen 50 valores posibles para el viento zonal, w_u y otros 50 valores para el viento meridional, w_v , más los escenarios de control y de alta resolución, los cuales no serán tenidos en cuenta para la trayectoria. Dichos valores están definidos para cada punto (λ, ϕ) de la superficie terrestre, con un paso de 1° . Como la ruta de estudio abarca gran parte del Océano Atlántico, sólo serán necesarios los vientos comprendidos entre los siguientes valores de longitud y latitud máximos y mínimos, respectivamente $[\phi_{min}, \lambda_{min}, \phi_{max}, \lambda_{max}] = [25^\circ, -81^\circ, 53^\circ, 14^\circ]$.

Cabe destacar que los vientos que serán usados para una posterior calibración están definidos a las 00:00 UT, por lo que se hará la hipótesis de que mientras se recorre la trayectoria los vientos no han sufrido ningún cambio con respecto al valor de los mismos a las 00:00 UT. Por decirlo de otro modo, no se considerará la variación temporal de los vientos a lo largo de la trayectoria, como ya se dijo en el apartado 4.3. Además, los vientos con los que se trabajará son los previstos a una presión

de 200 hPa, lo que es equivalente a una altitud de 11784 m. Como la ruta se recorre a una altitud de 12000 m se supondrá que las variaciones tanto de los tiempos de vuelo como de los vientos usando los datos de entrada mencionados anteriormente será muy pequeña con respecto al valor real en las condiciones adecuadas.

Una vez dejado claro el problema que se quiere resolver se procederá a seguir los siguientes pasos. En primer lugar, se explicarán y se darán resultados para los conjuntos de entrenamientos que se utilizarán para calibrar el tiempo de vuelo durante el 15 de Diciembre de 2017. En segundo lugar, se mostrará qué cambios aparecen al realizar la calibración usando cada uno de los conjuntos de entrenamiento, comparando sus resultados con criterio. Para terminar, se reflexionará sobre el método aplicado al problema concreto.

5.2 Conjuntos de calibración

En nuestro problema se va a trabajar con dos conjuntos de entrenamiento diferentes. Un conjunto de entrenamiento está formado por los valores del viento zonal y meridional en unos puntos concretos, un día y en una hora específica. El objetivo de un conjunto de entrenamiento es usar un histórico de predicciones junto con valores observados de una variable para realizar el postprocesado de los modelos originales. El primer conjunto abarca 50 valores para el viento zonal y para el viento meridional los días 1 y 15 de cada mes hasta el 1 de Diciembre del año 2017, todos a las 00:00 UT, definidos en los puntos de la malla descrita en la sección anterior. En cambio, el segundo conjunto será igual que el primero, salvo los días considerados, que serán ahora del 1 al 14 del mes de Diciembre de 2017. Comentar que, los modelos de predicción que proporciona el ECMWF se han realizado el día anterior del que se quiere predecir qué ocurrirá, es decir, con 24 h de antelación. Por otra parte, el ECMWF también da información de los valores reales del viento zonal y del viento meridional medidos en cada zona para cada uno de los días utilizados para la calibración.

Por tanto, los modelos de predicción lo forman vectores de vientos w_u y w_v , los cuales se pueden definir mediante 4 índices, $w_u(i, j, k, l)$ $w_v(i, j, k, l)$. Los dos primeros índices hacen referencia al punto en cuestión donde se quiere conocer el viento (latitud y longitud), y los dos últimos, al escenario y al día que se esté analizando. Los valores observados son únicos y sólo constan de 3 índices: $w_{u_{obs}}(i, j, l)$ $w_{v_{obs}}(i, j, l)$.

Para la temperatura se tomará la obtenida por la atmósfera estándar internacional, por lo que será independiente de los puntos (λ, ϕ) de la malla de estudio.

En este capítulo se trabajará con el tiempo de vuelo total de la trayectoria, y no con el tiempo que se tarda en recorrer cada segmento de la misma. En primer lugar, si se supone que el viento es nulo, el tiempo que se tarda en llegar al destino es:

$$t_v|_{w=0} = 45191 \text{ s} = 12 \text{ h } 33 \text{ min} \quad (5.1)$$

Según lo dicho anteriormente, para cada día el vector de tiempos tiene 50 componentes, cada una correspondiente a cada escenario: $t_v = [t_{v_1} \dots t_{v_{50}}]$. Como paso previo, se recogerá en una serie de tablas los valores del tiempo de vuelo previstos y observados utilizando cada uno de los conjuntos de entrenamiento. Para simplificar, sólo se muestra la media del tiempo de vuelo de todos los escenarios, el valor máximo y el valor mínimo para cada uno de los días. El objetivo será doble, por un lado comparar los valores del tiempo en la ida y en la vuelta, y por otro lado, ver si los tiempos reales (usando los vientos medidos) se parecen mucho a los tiempos medios previstos.

Tabla 5.3 Tiempos de vuelos previstos para la ida de los 6 primeros meses. Primer conjunto de entrenamiento.

DÍAS: ENERO-JUNIO/TIEMPOS(s)	$t_{v_{medio}}$	$t_{v_{maximo}}$	$t_{v_{minimo}}$
1 ENERO	48219	48747	47775
15 ENERO	46476	46851	46198
1 FEBRERO	49653	50131	49367
15 FEBRERO	49278	49597	49002
1 MARZO	48378	48671	48038
15 MARZO	45857	46331	45301
1 ABRIL	50064	50802	49598
15 ABRIL	49999	50512	49369
1 MAYO	48067	48419	47683
15 MAYO	50413	51764	49208
1 JUNIO	48558	49004	48052
15 JUNIO	45179	45543	44790

Tabla 5.4 Tiempos de vuelos previstos para la ida de los 6 últimos meses. Primer conjunto de entrenamiento.

DÍAS: JULIO-DICIEMBRE/TIEMPOS(s)	$t_{v_{medio}}$	$t_{v_{maximo}}$	$t_{v_{minimo}}$
1 JULIO	47036	47283	46708
15 JULIO	45673	46284	45222
1 AGOSTO	47639	48045	47150
15 AGOSTO	45612	46573	44925
1 SEPTIEMBRE	46894	47267	46498
15 SEPTIEMBRE	47185	48060	46559
1 OCTUBRE	43307	43652	42958
15 OCTUBRE	45302	46673	43692
1 NOVIEMBRE	49970	51536	48424
15 NOVIEMBRE	47478	48754	46799
1 DICIEMBRE	47483	48259	46638

Tabla 5.5 Tiempos de vuelos correspondientes a cada uno de los días 1 y 15 del 2017 según los vientos observados en cada punto de la trayectoria en el recorrido de ida. Primer conjunto de entrenamiento.

DÍAS: ENERO-JUNIO/TIEMPOS(s)	$t_{v_{wobs}}$	DÍAS: JULIO-DICIEMBRE/ TIEMPOS(s)	$t_{v_{wobs}}$
1 ENERO	48014	1 JULIO	47063
15 ENERO	46318	15 JULIO	45629
1 FEBRERO	49494	1 AGOSTO	47954
15 FEBRERO	49202	15 AGOSTO	45448
1 MARZO	48391	1 SEPTIEMBRE	46962
15 MARZO	45921	15 SEPTIEMBRE	46731
1 ABRIL	50186	1 OCTUBRE	43178
15 ABRIL	50171	15 OCTUBRE	45312
1 MAYO	47838	1 NOVIEMBRE	49473
15 MAYO	49916	15 NOVIEMBRE	47495
1 JUNIO	48657	1 DICIEMBRE	47529
15 JUNIO	45229		

Tabla 5.6 Tiempos de vuelos previstos y observados para los primeros 14 días del mes de Diciembre. Recorrido de ida. Segundo conjunto de entrenamiento.

DÍAS: DICIEMBRE/TIEMPOS(s)	$t_{v_{medio}}$	$t_{v_{maximo}}$	$t_{v_{minimo}}$	$t_{v_{wobs}}$
1 DICIEMBRE	47483	48259	46638	47529
2 DICIEMBRE	48333	49138	47692	48311
3 DICIEMBRE	47373	47744	46897	47387
4 DICIEMBRE	48521	49499	47824	48060
5 DICIEMBRE	47174	47653	46669	46742
6 DICIEMBRE	47652	48481	47062	46884
7 DICIEMBRE	46413	46960	45743	46348
8 DICIEMBRE	46401	46902	46059	46252
9 DICIEMBRE	46384	47072	45705	46339
10 DICIEMBRE	45430	45823	44841	45401
11 DICIEMBRE	48593	49665	47754	47954
12 DICIEMBRE	47480	48013	47055	47478
13 DICIEMBRE	47431	47826	46955	47485
14 DICIEMBRE	49733	50345	49397	49660

En vista a los resultados se puede comprobar que el máximo de todos los tiempos se obtiene para el 15 de Mayo, siendo el tiempo igual a $51764 s = 14 h 23 min$. El tiempo más pequeño se da el día 1 de Octubre, marcando un tiempo de $42958 s = 11 h 56 min$. Es importante remarcar la diferencia en el tiempo de vuelo cuando se tiene un cierto viento en cada punto de la trayectoria. El viento provoca incrementos y reducciones más que notables del tiempo de vuelo que afectan directamente en el consumo de combustible de la aeronave, y por tanto, en el coste del vuelo. En nuestro caso, se obtienen unos incrementos y reducciones relativas con respecto al tiempo de vuelo sin viento que pueden calcularse como:

$$\Delta t = \frac{|t_{v_{max}} - t_{v,w=0}|}{t_{v,w=0}} = 0.14 \Rightarrow 14\% \quad (5.2)$$

$$\Delta t = \frac{|t_{v_{min}} - t_{v,w=0}|}{t_{v,w=0}} = 0.049 \Rightarrow 4.9\% \quad (5.3)$$

Tabla 5.7 Tiempos de vuelos previstos para la vuelta de los 6 primeros meses. Primer conjunto de entrenamiento.

DÍAS: ENERO-JUNIO/TIEMPOS(s)	$t_{v_{medio}}$	$t_{v_{maximo}}$	$t_{v_{minimo}}$
1 ENERO	42658	43081	42298
15 ENERO	44140	44394	43892
1 FEBRERO	40499	40674	40189
15 FEBRERO	41505	41941	41249
1 MARZO	41757	42051	41483
15 MARZO	42101	42622	41687
1 ABRIL	41706	42158	41416
15 ABRIL	42148	42622	41783
1 MAYO	43637	44201	42933
15 MAYO	40789	41582	40051
1 JUNIO	44546	44993	44145
15 JUNIO	45242	46176	44457

Tabla 5.8 Tiempos de vuelos previstos para la vuelta de los 6 últimos meses. Primer conjunto de entrenamiento.

DÍAS: JULIO-DICIEMBRE/TIEMPOS(s)	$t_{v_{medio}}$	$t_{v_{maximo}}$	$t_{v_{minimo}}$
1 JULIO	44893	45437	44506
15 JULIO	46747	47778	45186
1 AGOSTO	43967	44403	43572
15 AGOSTO	46278	47031	45727
1 SEPTIEMBRE	45201	45757	44692
15 SEPTIEMBRE	43924	44232	43599
1 OCTUBRE	46667	47204	46188
15 OCTUBRE	45416	46318	44879
1 NOVIEMBRE	42553	42985	42209
15 NOVIEMBRE	41706	42333	40985
1 DICIEMBRE	43268	43729	42537

Tabla 5.9 Tiempos de vuelos correspondientes a cada uno de los días 1 y 15 del 2017 según los vientos observados en cada punto de la trayectoria en el recorrido de vuelta. Primer conjunto de entrenamiento.

DÍAS: ENERO-JUNIO/TIEMPOS(s)	$t_{v_{wobs}}$	DÍAS: JULIO-DICIEMBRE/ TIEMPOS(s)	$t_{v_{wobs}}$
1 ENERO	42698	1 JULIO	45038
15 ENERO	44369	15 JULIO	46957
1 FEBRERO	40448	1 AGOSTO	43620
15 FEBRERO	41515	15 AGOSTO	45960
1 MARZO	41853	1 SEPTIEMBRE	45252
15 MARZO	42084	15 SEPTIEMBRE	43922
1 ABRIL	41637	1 OCTUBRE	46708
15 ABRIL	42189	15 OCTUBRE	45603
1 MAYO	43718	1 NOVIEMBRE	42491
15 MAYO	41049	15 NOVIEMBRE	41771
1 JUNIO	44580	1 DICIEMBRE	43422
15 JUNIO	45326		

Tabla 5.10 Tiempos de vuelos previstos y observados para los primeros 14 días del mes de Diciembre. Recorrido de vuelta. Segundo conjunto de entrenamiento.

DÍAS: DICIEMBRE/TIEMPOS(s)	$t_{v_{medio}}$	$t_{v_{maximo}}$	$t_{v_{minimo}}$	$t_{v_{wobs}}$
1 DICIEMBRE	43268	43729	42537	43423
2 DICIEMBRE	42681	43054	42244	42732
3 DICIEMBRE	43812	44002	43615	43842
4 DICIEMBRE	42163	42415	41900	42167
5 DICIEMBRE	41997	42260	41745	42119
6 DICIEMBRE	44690	45115	44258	45066
7 DICIEMBRE	44628	45282	44096	44608
8 DICIEMBRE	44951	45662	44528	45199
9 DICIEMBRE	45163	45774	44700	44813
10 DICIEMBRE	46483	47010	45908	46513
11 DICIEMBRE	41617	42694	40882	42148
12 DICIEMBRE	42134	42597	41697	42079
13 DICIEMBRE	42831	43129	42374	42860
14 DICIEMBRE	40585	40831	40314	40561

La importancia de implementar el recorrido de vuelta y compararlo con el ida radica en ilustrar con números como afecta la dirección del viento, y no su módulo, en el tiempo total de vuelo.

El día 15 de Mayo se predice que el tiempo de vuelo para la vuelta será mínimo e igual a $40051 s = 11 h 8 min$. Si nos fijamos en el máximo de los tiempos máximos de vuelo para el recorrido de ida, que es igual a $51764 s = 14 h 23 min$, se puede comprobar que la variación del tiempo según el sentido de la trayectoria puede ser bastante amplia por el hecho del cambio de dirección de los vientos. En términos de porcentajes relativos el tiempo de vuelo puede llegar a ser un 30% mayor en la ida que en la vuelta.

Para ver más claramente cómo funcionan las predicciones para cada uno de los días, se puede representar en una serie de figuras el valor del tiempo medio, máximo y mínimo de vuelo, así como el tiempo de vuelo correspondiente a las observaciones reales, para cada uno de los conjuntos de entrenamiento. De esta manera, se puede ver gráficamente, donde cae el tiempo real, es decir, si está dentro del rango previsto, muy alejado o no de la media, o incluso fuera de rango (en consonancia con la figura 3.12).

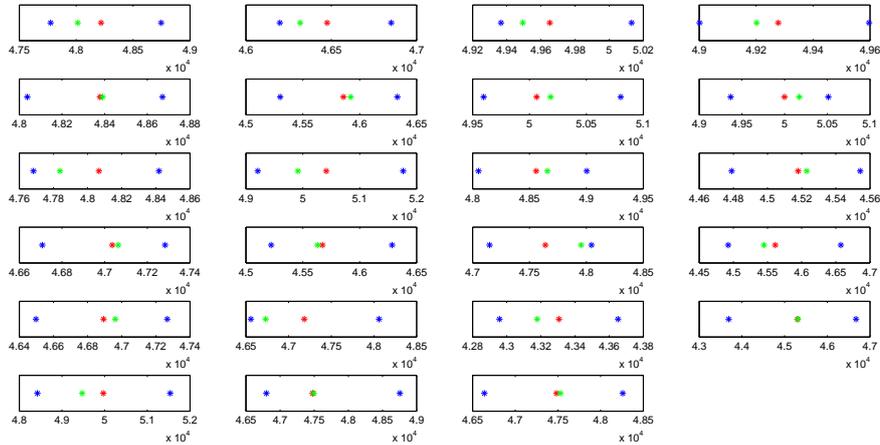


Figura 5.2 Representación de los tiempos de vuelo durante la ida para el primer conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales.

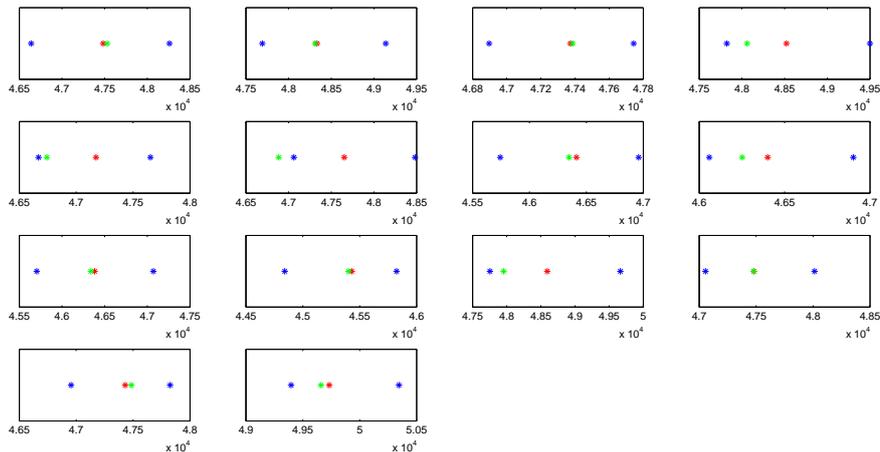


Figura 5.3 Representación de los tiempos de vuelo durante la ida para el segundo conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales.

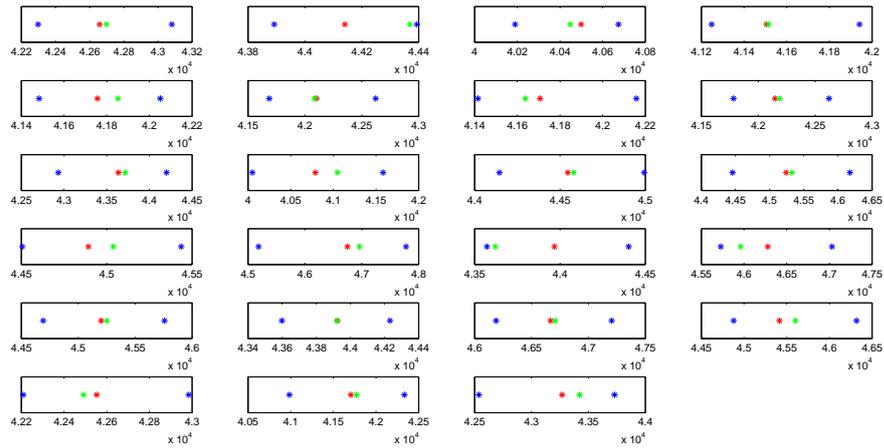


Figura 5.4 Representación de los tiempos de vuelo durante la vuelta para el primer conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales.

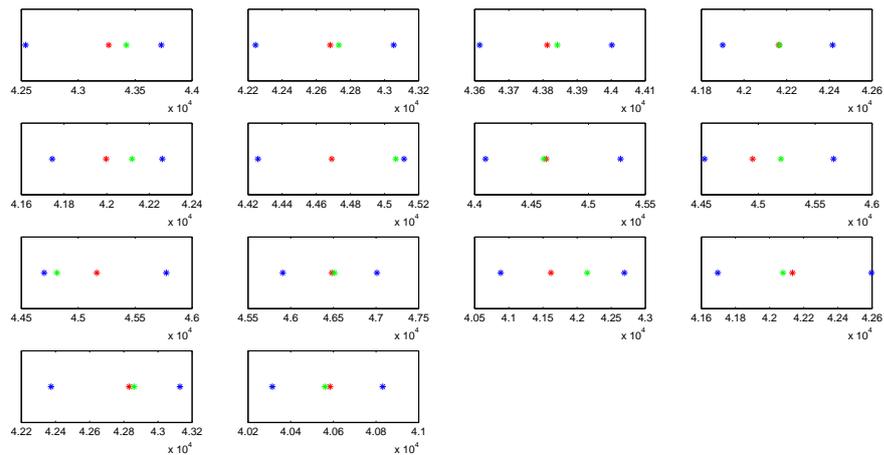


Figura 5.5 Representación de los tiempos de vuelo durante la vuelta para el segundo conjunto de entrenamiento. En azul, los tiempos máximos y mínimos previstos para cada día; en rojo, la media de los tiempos de cada escenario; y en verde, el tiempo real basado en observaciones reales.

Con la representación de las figuras anteriores se puede ver claramente, las diferentes posiciones de las observaciones en función del día correspondiente. En general, se observa que el valor medio del tiempo de vuelo de las predicciones es bastante cercano al valor observado. En la gran mayoría de los casos la observación cae dentro del intervalo formado por los tiempos máximos y mínimos de los 50 escenarios, lo que induce a pensar que con bastante probabilidad ocurrirá lo mismo para el día 15 de Diciembre, del cual se quiere conocer el tiempo de vuelo. Aunque el histórico de predicciones sea muy acertado, no quiere decir que el día 15 de Diciembre la observación tenga que caer dentro del intervalo. Es más, si se mira con detenimiento, uno puede darse cuenta que para el recorrido de ida del día 6 de Diciembre (2ª fila y 2ª columna de la figura 5.3), el tiempo correspondiente a la observación es menor que el mínimo tiempo previsto por los 50 escenarios ese día. Si ocurriera esto para nuestro día de estudio, la calibración no sólo sería útil para ganar precisión, si no que sería imprescindible.

5.3 Resultados de los vientos tras la calibración

Antes de mostrar los resultados obtenidos es necesario decir que la optimización del CRPS se ha realizado utilizando el comando *fmincon* de Matlab. Las condiciones iniciales y las restricciones ya fueron aclaradas anteriormente (véase sección 3.4). Falta por mencionar cuál ha sido la tolerancia elegida para la optimización de las variables. Se ha escogido un valor de tolerancia de 10^{-5} , ya que se ha comprobado que eligiendo un valor mayor no se consigue mucha más precisión, y sí un aumento en el tiempo de cálculo.

Dicho esto, se procede a estudiar los resultados obtenidos cuando se utiliza el primer conjunto de entrenamiento. Como se tienen muchos puntos en la trayectoria, se analizan sólo dos puntos cercanos correspondientes a las ciudades de Berlín y Hamburgo. En cada uno de los puntos se tienen dos variables: viento zonal y viento meridional. Sólo se mostrarán los resultados para el viento zonal y no para el viento meridional al ser análogos. Con ello, el cálculo del CRPS para el modelo en bruto y para el modelo calibrado puede verse a continuación.

Tabla 5.11 Resultados de la optimización para el primer conjunto de entrenamiento.

w_u	ξ	ξ_{optimo}	a	b	c	d
Berlín	0.6309	0.6116	-0.116	1.0123	0.3416	0.3280
Hamburgo	0.7481	0.7193	-0.1428	1.0288	0	0.9362

De los resultados pueden extraerse varias conclusiones. Por un lado, un valor tan bajo de CRPS induce a pensar que los modelos proporcionados de predicción ya son bastante buenos de por sí, por lo que una optimización sólo ayuda a mejorarlos más bajando su valor, pero no en gran medida. Llama la atención también, que en los dos casos el valor de a es negativo, siendo necesario un sesgo negativo para corregir el modelo. En el caso de Hamburgo la optimización del CRPS se consigue haciendo cero uno de los parámetros, como es, el coeficiente c .

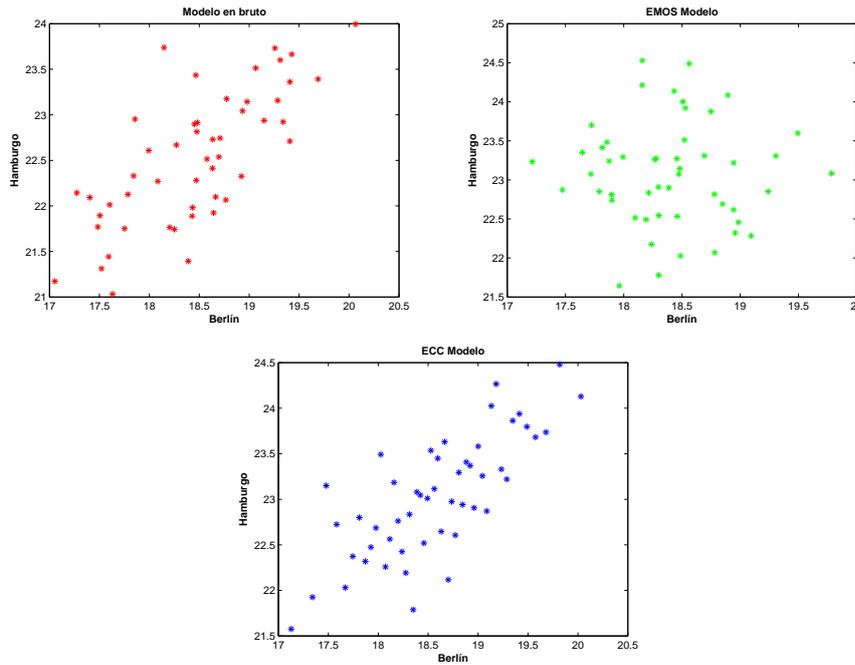


Figura 5.6 Representación del ensemble forecast antes de la calibración, tras realizar la calibración individual, EMOS y una vez aplicado el ECC .

Se puede ver que optimizando a través del EMOS se pierde la tendencia aproximadamente lineal que tenía la nube de puntos del modelo en bruto. Al aplicar el ECC a este modelo se recupera gran parte de esa tendencia motivada por el uso de la reordenación del modelo en bruto usando el Multivariate Rank Average, cómo se explicó en la sección 3.4.

Tabla 5.12 Coeficientes de correlación entre ambas ciudades para el primer conjunto de entrenamiento.

	BRUTO	EMOS	ECC
COEFICIENTES DE CORRELACIÓN	0.7341	-0.0254	0.7619

El coeficiente de correlación del EMOS es prácticamente nulo mientras que al aplicar el ECC la relación entre las variables se recupera, obteniéndose un valor de correlación similar al original.

Si se usa ahora el segundo conjunto de entrenamiento (primeros 14 días del mes de Diciembre) los valores del CRPS y de los coeficientes que se obtienen son:

Tabla 5.13 Resultados de la optimización para el segundo conjunto de entrenamiento.

w_u	ξ	ξ_{optimo}	a	b	c	d
Berlín	0.4814	0.4220	0.1235	1.005	0.5071	0.0957
Hamburgo	0.9807	0.8850	-0.7558	1.0486	0.3759	0.8108

Lo que más sorprende si se comparan estos valores con el primer conjunto de entrenamiento es que el valor del CRPS óptimo para Berlín es menor en el segundo caso que en el primero (0.6116

vs 0.4220). Sin embargo, para Hamburgo ocurre justamente lo contrario. El CRPS es menor en el primer caso y no en el segundo (0.7193 vs 0.8850). Aun así, no se pueden sacar conclusiones generales en base a estos resultados ya que sólo se muestran dos variables. En la sección siguiente, se calcularán los tiempos de vuelos, que es el objetivo de nuestra aplicación. A partir de ahí, se podrá decidir y valorar sobre cuál modelo merece más la pena usar para ajustar, o si por lo contrario, ninguno de los dos aporta demasiado, visto los valores tan pequeños del CRPS. Comentar también, que al igual que en el primer caso, se obtiene un valor de a negativo para Hamburgo, lo que implica un sesgo negativo al ajustar. Los valores de c y d en general son menores de 1 mientras que los valores de b suelen alcanzar valores mayores de 1.

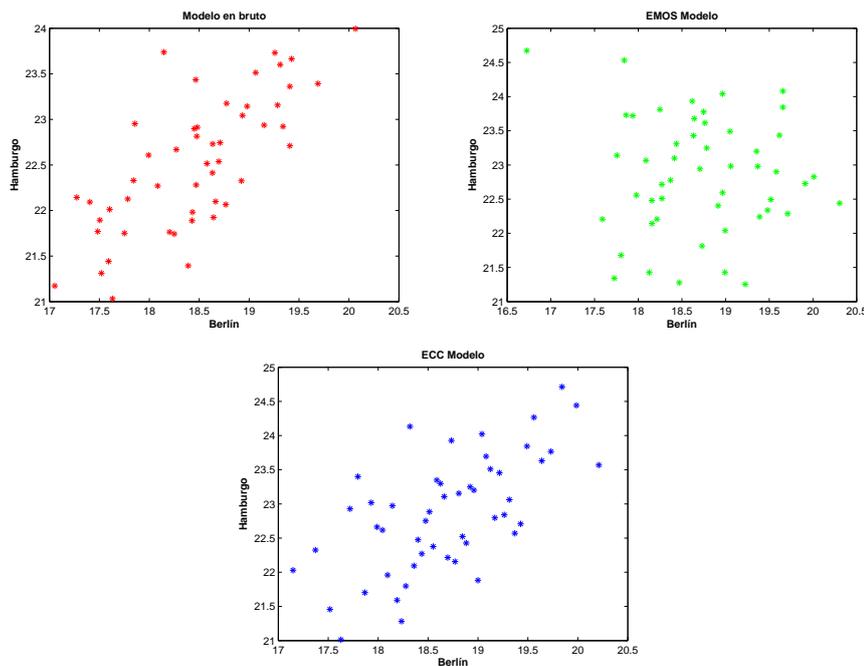


Figura 5.7 Representación del ensemble forecast antes de la calibración, tras realizar la calibración individual, EMOS y una vez aplicado el ECC .

Ocurre lo mismo que para el primer conjunto de entrenamiento. El ECC permite recuperar algo mejor la relación primitiva existente en el modelo en bruto entre las ciudades de Berlín y Hamburgo, la cual se pierde al calibrar con el EMOS de forma independiente.

Tabla 5.14 Coeficientes de correlación entre ambas ciudades para el segundo conjunto de entrenamiento.

	BRUTO	EMOS	ECC
COEFICIENTES DE CORRELACIÓN	0.7341	-0.0779	0.6309

Al igual que para el primer conjunto de entrenamiento, la tendencia se recupera bastante. Además, se puede destacar que, mientras en el primer conjunto la tendencia resultante era algo más lineal que en el modelo en bruto (0.7341 vs 0.7619), en este caso ocurre lo contrario, siendo ésta algo menos lineal que la original (0.7341 vs 0.6309). En cualquier caso, es evidente la ganancia obtenida mediante el ECC.

Para ilustrar lo que ocurre al calibrar no sólo en un punto si no en toda la malla de estudio, se presentan en una serie de figuras los valores del viento zonal y meridional para un escenario cualquiera del modelo en bruto, del modelo calibrado mediante el EMOS de forma individual y tras el ECC.

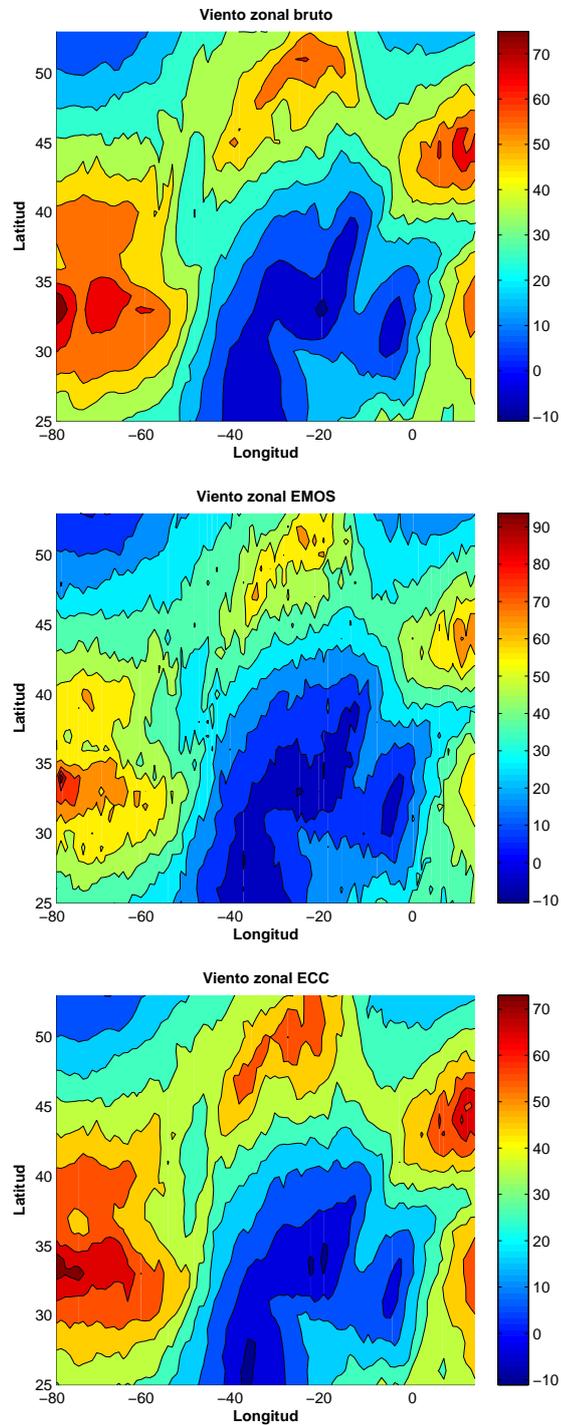


Figura 5.8 Representación de las curvas de nivel para el viento zonal. Primer conjunto de entrenamiento.

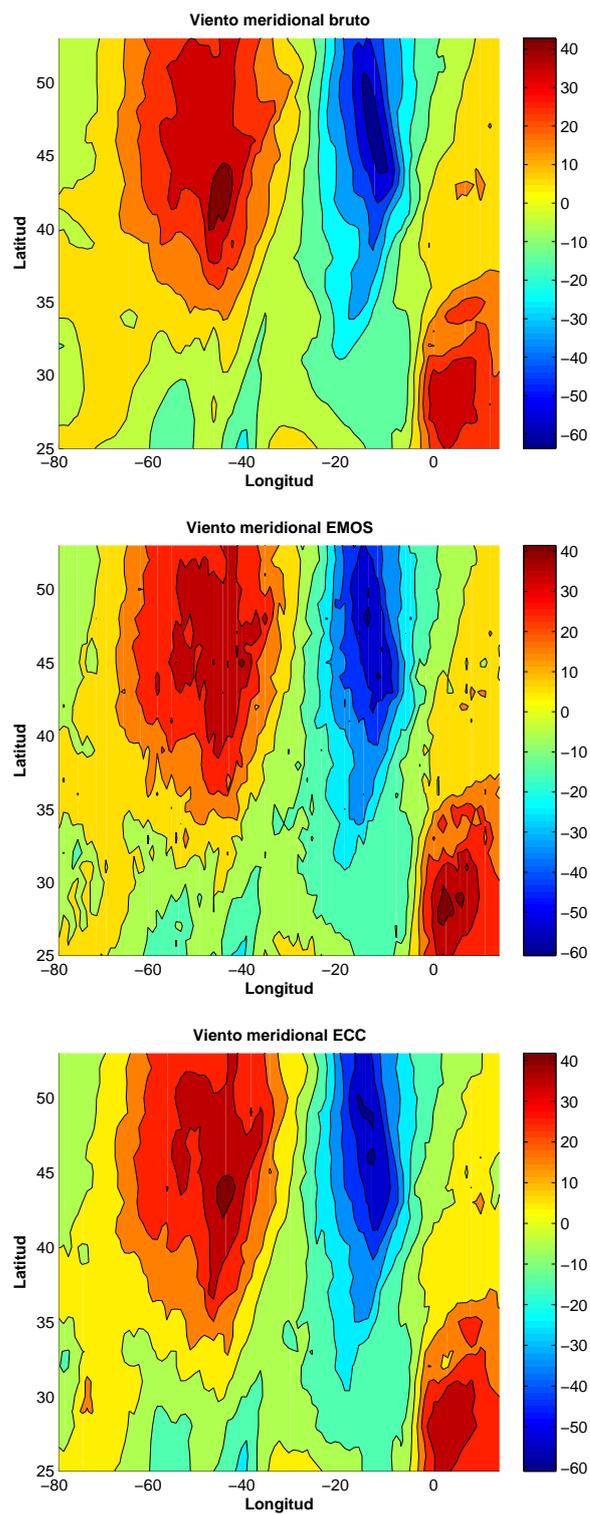


Figura 5.9 Representación de las curvas de nivel para el viento meridional. Primer conjunto de entrenamiento.

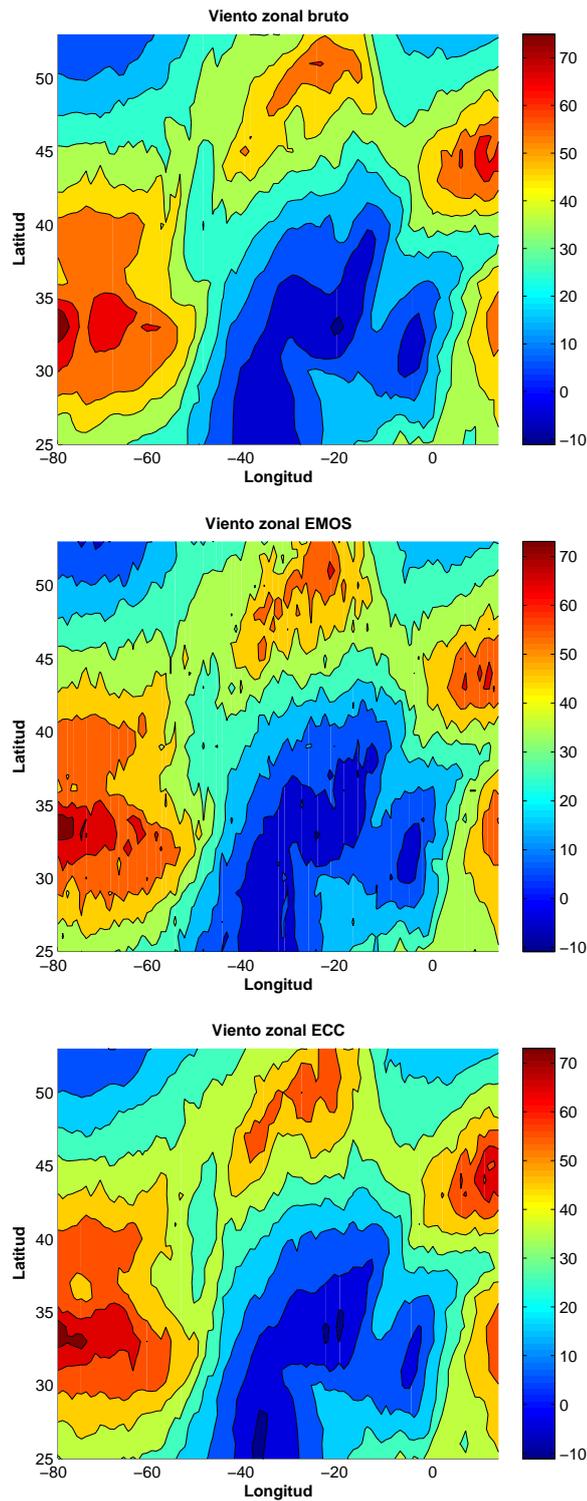


Figura 5.10 Representación de las curvas de nivel para el viento zonal. Segundo conjunto de entrenamiento.

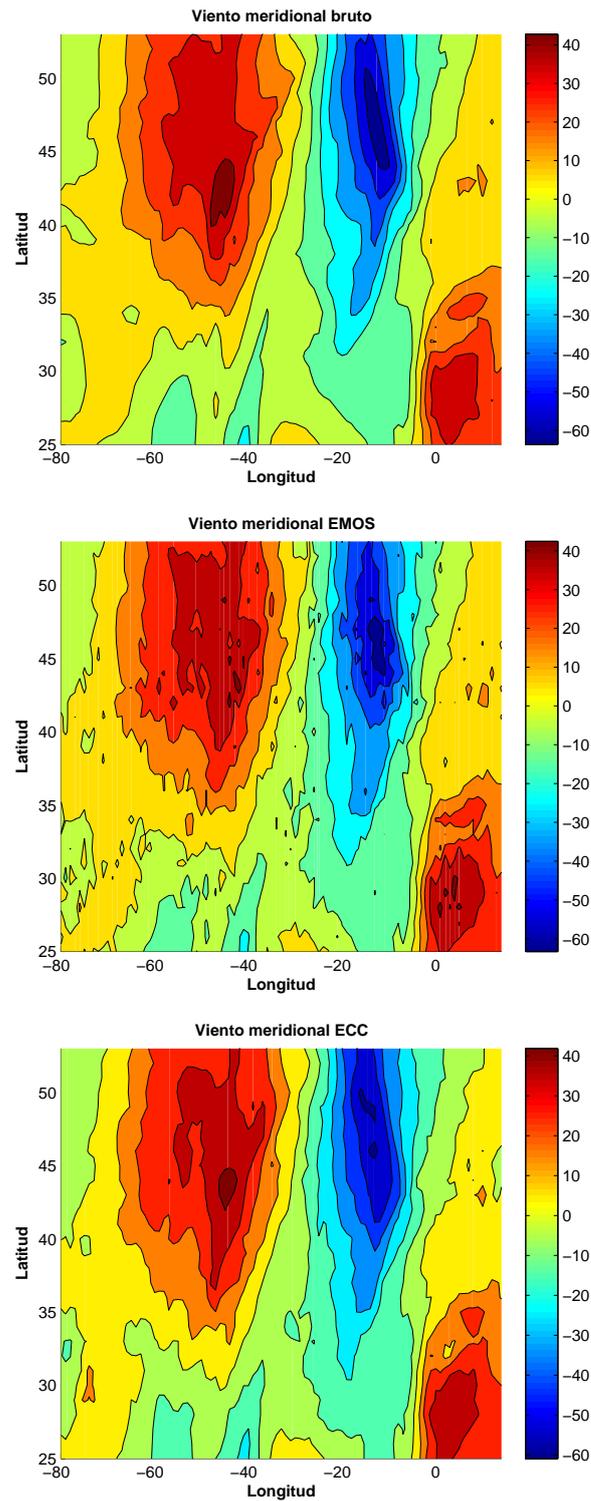


Figura 5.11 Representación de las curvas de nivel para el viento meridional. Segundo conjunto de entrenamiento.

De estas figuras, se observa la pérdida de uniformidad del modelo cuando se realiza el EMOS de forma individual. El mapa obtenido en ese caso se aleja más de ser suave, presentando números picos entre regiones cercanas.

En el segundo conjunto de entrenamiento la no uniformidad se hace más clara entre regiones cercanas, las cuales se visualizan mejor al cambiar el color de sus curvas de nivel. A simple vista, parece que el primer conjunto es algo más suave que el segundo conjunto. Aun así, es importante destacar que aunque los modelos obtenidos mediante la optimización individual son menos uniformes que los del ECC, la pérdida de su uniformidad no es del todo severa, por lo que pueden servir para tener una idea genérica de los vientos en una determinada zona (por ejemplo, el *jet stream* en el hemisferio norte, componiendo el viento zonal y el meridional). La respuesta a esto, se debe a que a pesar de haber calibrado individualmente, las predicciones de partida ya eran bastantes buenas en cuanto a su parecido a la realidad, lo que implica que el valor del CRPS en general es pequeño y con él, la desviación con respecto al modelo en bruto original.

5.4 Tiempos de vuelo tras el postprocesado

Quizás esta sección sea la más importante del proyecto, ya que se verá de una forma práctica cómo funciona el Ensemble Copula Coupling para una aplicación concreta relacionada con la Mecánica del Vuelo.

Se hablará indistintamente de modelos en bruto, calibrados y observados de los tiempos de vuelo y de los vientos, siempre y cuando se tenga presente que la optimización se ha realizado sobre los vientos y que por tanto, los tiempos de vuelo son un resultado indirecto de la ejecución de la trayectoria tomando los vientos como variables de entrada.

Lo primero que hay que mostrar es una gráfica similar a la de los apartados anteriores, para ver dónde se encuentra el tiempo observado el día 15 de Diciembre con respecto al modelo en bruto para ese día proporcionado por el ECMWF .

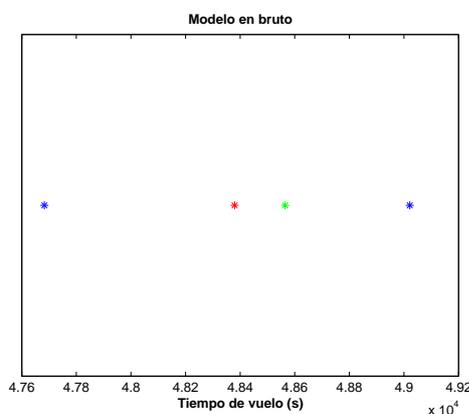


Figura 5.12 Modelo en bruto de los tiempos de vuelo para el recorrido de ida.

Las características más importantes de este modelo para el recorrido de ida se muestran a continuación. También es importante conocer la diferencia existente entre la observación y la media del modelo en bruto, para poder ver cómo queda dicho valor tras el postprocesado. Otros parámetros que pueden ser interesantes son el rango entre los valores de tiempo máximo y mínimo, y la diferencia entre esos valores y el tiempo observado.

$$t_{v_{medio}} = 48380 \text{ s} \quad t_{v_{maximo}} = 49021 \text{ s} \quad t_{v_{minimo}} = 47683 \text{ s} \quad (5.4)$$

$$Desv. \text{ Tipica } S = 284.36 \text{ s} \quad (5.5)$$

$$t_{v_{wobs}} = 48565 \text{ s} \quad (5.6)$$

$$|t_{v_{medio}} - t_{v_{wobs}}| = 185 \text{ s} \quad (5.7)$$

$$|t_{v_{maximo}} - t_{v_{minimo}}| = 1338.4 \text{ s} \quad (5.8)$$

$$|t_{v_{maximo}} - t_{v_{wobs}}| = 456 \text{ s} \quad (5.9)$$

$$|t_{v_{wobs}} - t_{v_{minimo}}| = 882 \text{ s} \quad (5.10)$$

Tras aplicar el Ensemble Copula Coupling al primer conjunto de entrenamiento formado por 23 días (días 1 y 15 de cada mes hasta el 1 de Diciembre), se obtiene un modelo calibrado que para el recorrido de ida tiene las siguientes características:

$$t_{v_{medio}} = 48428 \text{ s} \quad t_{v_{maximo}} = 48877 \text{ s} \quad t_{v_{minimo}} = 47974 \text{ s} \quad (5.11)$$

$$Desv. \text{ Tipica } S = 224.84 \text{ s} \quad (5.12)$$

$$t_{v_{wobs}} = 48565 \text{ s} \quad (5.13)$$

$$|t_{v_{medio}} - t_{v_{wobs}}| = 137 \text{ s} \quad (5.14)$$

$$|t_{v_{maximo}} - t_{v_{minimo}}| = 903 \text{ s} \quad (5.15)$$

$$|t_{v_{maximo}} - t_{v_{wobs}}| = 312 \text{ s} \quad (5.16)$$

$$|t_{v_{wobs}} - t_{v_{minimo}}| = 591 \text{ s} \quad (5.17)$$

De estos resultados se pueden sacar varias conclusiones. La más importante es el acercamiento relativo del tiempo de vuelo medio al tiempo de vuelo observado pasando de una diferencia de 185 s a 137s. En términos relativos la mejora es de

$$\% \text{ Mejora} = \frac{185 - 137}{185} \cdot 100 = 26\% \quad (5.18)$$

que puede llegar a ser importante teniendo en cuenta lo cerca que está el tiempo medio del modelo en bruto con respecto al tiempo observado.

Por otra parte, el nuevo modelo obtenido del postprocesado de los vientos, se contrae, es decir, el tiempo máximo y el tiempo mínimo se acercan más al tiempo observado (al disminuir todas las diferencias). Probablemente, si la observación se encontrara fuera del intervalo formado por el tiempo máximo y el tiempo mínimo (del modelo en bruto), el modelo calibrado habría aumentado su dispersión, así como sus valores extremos, con el objetivo de dar cabida a la observación. La disminución de la dispersión es muy pequeña debido a que el valor medio está muy cerca del valor observado. Lo descrito en este párrafo, puede entenderse gráficamente, si se superponen los dos modelos junto con la observación en la siguiente figura.

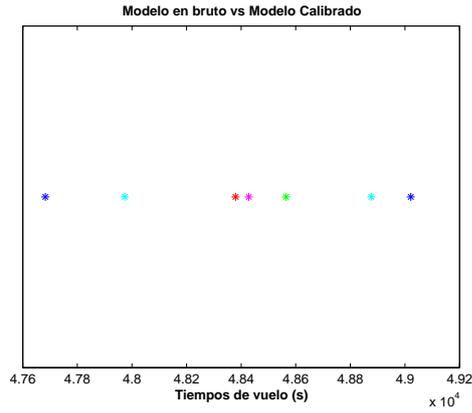


Figura 5.13 Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de ida. Primer conjunto de entrenamiento.

Para el recorrido de vuelta el modelo en bruto tiene las siguientes características:

$$t_{v_{medio}} = 41322 \text{ s} \quad t_{v_{maximo}} = 41656 \text{ s} \quad t_{v_{minimo}} = 41016 \text{ s} \quad (5.19)$$

$$Desv. \text{ Tipica } S = 151.85 \text{ s} \quad (5.20)$$

$$t_{v_{wobs}} = 41288 \text{ s} \quad (5.21)$$

$$|t_{v_{medio}} - t_{v_{wobs}}| = 34 \text{ s} \quad (5.22)$$

$$|t_{v_{maximo}} - t_{v_{minimo}}| = 640 \text{ s} \quad (5.23)$$

$$|t_{v_{maximo}} - t_{v_{wobs}}| = 368 \text{ s} \quad (5.24)$$

$$|t_{v_{wobs}} - t_{v_{minimo}}| = 272 \text{ s} \quad (5.25)$$

que dan lugar gráficamente a un modelo cuya forma se puede ver a continuación en la figura.

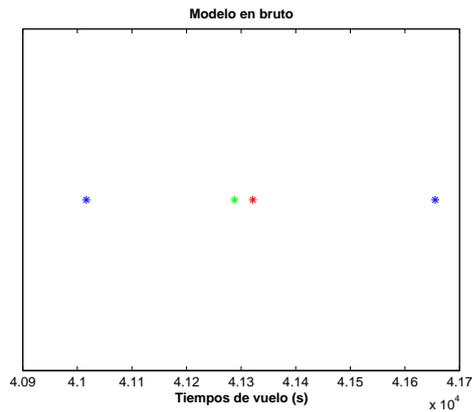


Figura 5.14 Modelo en bruto de los tiempos de vuelo para el recorrido de vuelta.

Aplicando los resultados de la optimización se obtienen unos modelos de vientos calibrados que al implementarlos en la trayectoria originan unos tiempos de vuelo cuyas características son las

siguientes:

$$t_{v_{medio}} = 41293 \text{ s} \quad t_{v_{maximo}} = 41588 \text{ s} \quad t_{v_{minimo}} = 40928 \text{ s} \quad (5.26)$$

$$Desv. \text{ Tipica } S = 152.71 \text{ s} \quad (5.27)$$

$$t_{v_{wobs}} = 41288 \text{ s} \quad (5.28)$$

$$|t_{v_{medio}} - t_{v_{wobs}}| = 5 \text{ s} \quad (5.29)$$

$$|t_{v_{maximo}} - t_{v_{minimo}}| = 660 \text{ s} \quad (5.30)$$

$$|t_{v_{maximo}} - t_{v_{wobs}}| = 300 \text{ s} \quad (5.31)$$

$$|t_{v_{wobs}} - t_{v_{minimo}}| = 360 \text{ s} \quad (5.32)$$

La mejora relativa que se tiene en el recorrido de vuelta puede calcularse como

$$\% \text{ Mejora} = \frac{34 - 5}{34} \cdot 100 = 83 \% \quad (5.33)$$

la cual es mucho más elevada que para el recorrido de ida. Es curioso observar que a pesar de que el tiempo medio se acerca muchísimo al real, la dispersión de los 50 escenarios aumenta levemente (151.85 s a 152.71 s), haciendo posible que ese acercamiento sea mucho más llamativo que en la ida. Con ello, los tiempos máximos y mínimos se expanden, por así decirlo, aumentando y disminuyendo respectivamente. Gráficamente, se puede ver en la figura siguiente.

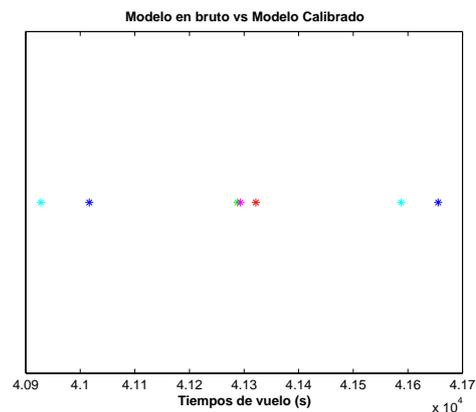


Figura 5.15 Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de vuelta. Primer conjunto de entrenamiento.

Una vez terminado el desarrollo de los tiempos de vuelos provenientes del primer conjunto de entrenamiento se prosigue con los resultados obtenidos utilizando el Ensemble Copula Coupling con el segundo conjunto de entrenamiento. A modo de recordatorio, este segundo conjunto, está formado por las previsiones de vientos para los primeros 14 días del mes de Diciembre junto con los valores reales para esos días. El modelo en bruto a calibrar será el mismo: el correspondiente al día 15 de Diciembre, cuyas características ya se detallaron anteriormente.

Para el recorrido de ida el modelo calibrado tiene la siguientes forma y características:

$$t_{v_{medio}} = 48398 \text{ s} \quad t_{v_{maximo}} = 48856 \text{ s} \quad t_{v_{minimo}} = 47977 \text{ s} \quad (5.34)$$

$$Desv. \text{ Tipica } S = 251.24 \text{ s} \quad (5.35)$$

$$t_{v_{wobs}} = 48565 \text{ s} \quad (5.36)$$

$$|t_{v_{medio}} - t_{v_{wobs}}| = 167 \text{ s} \quad (5.37)$$

$$|t_{v_{maximo}} - t_{v_{minimo}}| = 879 \text{ s} \quad (5.38)$$

$$|t_{v_{maximo}} - t_{v_{wobs}}| = 291 \text{ s} \quad (5.39)$$

$$|t_{v_{wobs}} - t_{v_{minimo}}| = 588 \text{ s} \quad (5.40)$$

El margen de mejora de este modelo con respecto al modelo en bruto puede calcularse como

$$\% \text{ Mejora} = \frac{185 - 167}{185} \cdot 100 = 9.7\%. \quad (5.41)$$

Puede apreciarse que es bastante menor que el obtenido con respecto al primer conjunto de entrenamiento 26%. También lo es la dispersión, S , siendo ésta algo menor para el primer conjunto: 224.84 s en lugar de 251.24 s. Sin embargo, el rango es mayor en el primer caso que en el segundo (903 s vs 879 s).

Comparando el modelo calibrado con el modelo en bruto el resultado es lo siguiente.

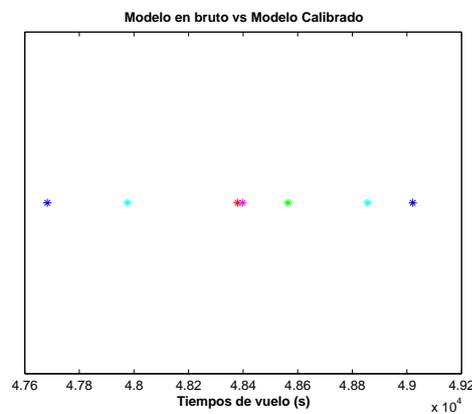


Figura 5.16 Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de ida. Segundo conjunto de entrenamiento.

En cuanto a la vuelta, el modelo calibrado con sus características se muestra a continuación:

$$t_{v_{medio}} = 41265 \text{ s} \quad t_{v_{maximo}} = 41570 \text{ s} \quad t_{v_{minimo}} = 40905 \text{ s} \quad (5.42)$$

$$\text{Desv. Tipica } S = 159.86 \text{ s} \quad (5.43)$$

$$t_{v_{wobs}} = 41288 \text{ s} \quad (5.44)$$

$$|t_{v_{medio}} - t_{v_{wobs}}| = 23 \text{ s} \quad (5.45)$$

$$|t_{v_{maximo}} - t_{v_{minimo}}| = 665 \text{ s} \quad (5.46)$$

$$|t_{v_{maximo}} - t_{v_{wobs}}| = 282 \text{ s} \quad (5.47)$$

$$|t_{v_{wobs}} - t_{v_{minimo}}| = 383 \text{ s} \quad (5.48)$$

Ahora el nuevo margen de mejora del tiempo medio con respecto al valor observado es de

$$\% \text{ Mejora} = \frac{34 - 23}{34} \cdot 100 = 32.4\%. \quad (5.49)$$

Comparado con el obtenido en la vuelta para el primer conjunto de entrenamiento, 83 % este nuevo margen es muy inferior, lo que nos hace notar que quizás el primer conjunto sea más adecuado para realizar la calibración. Es más, también la dispersión de los 50 escenarios para el primer conjunto es menor que para el segundo (152.71 s vs 159.86). La diferencia de tiempos es ahora algo mayor pero en menor medida, pasando de 660 s a 665 s.

Por último, superponiendo el modelo en bruto junto con el modelo calibrado se llega a

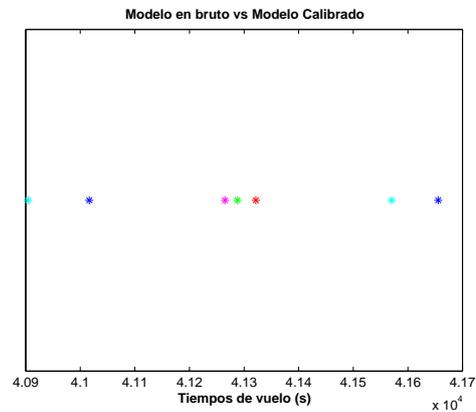


Figura 5.17 Modelo en bruto y calibrado de los tiempos de vuelo para el recorrido de vuelta. Segundo conjunto de entrenamiento.

5.5 Conclusiones generales

Con vista a los resultados anteriores se incluye una sección en la que se habla de las conclusiones generales que pueden sacarse tras aplicarse el postprocesado a los modelos en bruto de partida. Se analizan las mismas en base a lo obtenido con su aplicación a la predicción de trayectorias teniendo en cuenta diversos puntos que puedan tener cierta influencia en un futuro. A continuación, se resumen algunas de ellas:

- En primer lugar, una vez terminado el desarrollo del Ensemble Copula Coupling es importante discutir de la rentabilidad de su aplicación a la predicción de trayectorias a la hora de ganar precisión en el ajuste del tiempo de vuelo. Si que es verdad que aunque el ECC proporciona bastante precisión, también supone un gasto desde el punto de vista numérico. Lo que hay que mirar en este caso es si la predicción de trayectorias requiere de gran precisión a la hora de su operatividad en la práctica. No cabe olvidar que son muchas las simplificaciones que se asumen en su cálculo. Algunas de ellas son la eliminación de subidas, descensos, aterrizajes y la velocidad constante suponiendo tramos rectos. Por tanto, los valores obtenidos pueden variar en el torno de minutos o incluso de horas si se meten las esperas para aterrizar, o las emergencias. La precisión que nos da el ECC es del orden de varios minutos, por lo que en teoría no sería muy necesario hacerla porque los modelos proporcionados por el ECMWF se ajustan de por sí bastante bien a la realidad.
- Con respecto a la necesidad de aplicar el ECC comentada en el párrafo anterior hay que añadir ciertas reseñas. Con los datos de partida proporcionados en este proyecto no se puede establecer una valoración más profunda sobre la rentabilidad de la aplicación del método. La justificación es que la media de los 50 escenarios para el día 15 de Diciembre cae muy cerca del valor real, y en ningún caso, fuera del intervalo de los valores máximos y mínimos, que es donde se puede

notar más la utilidad del método. También, el ECC se ha aplicado al viento zonal y al viento meridional y no a los tiempos de vuelo. Los tiempos de vuelo se obtienen de una forma indirecta donde entran en juego otras variables como la velocidad aerodinámica que en muchos casos es de un orden superior a los vientos, por lo que un cambio de precisión en los mismos, no tiene una gran repercusión en el tiempo de vuelo. Quizás en otras aplicaciones de la estadística, meteorología o incluso de la aeronáutica donde los vientos repercutan de manera más directa o donde su precisión afecte de manera determinante para el cálculo de una magnitud, sea mucho más útil este método.

-Por último es muy importante destacar las diferencias existentes al aplicar el método con conjuntos de entrenamiento diferentes. En el primer caso el número de casos, instancias o días N es mayor $N = 23$, pero están más separados en el tiempo al ser previsiones sobre los días 1 y 15 de cada mes. En el segundo caso, $N = 14$, pero las previsiones están más juntas en tiempo por ser del mes de Diciembre. Es curioso que visto lo visto con los resultados de la sección anterior en el que el primer conjunto de entrenamiento daba valores mucho más precisos que el segundo, sea en el cual las previsiones están realizadas durante días más espaciados en el tiempo. Nuestro sentido común diría lo contrario. Aunque el número de casos sea más pequeño, al estar éstos más cercanos en tiempo sería más fácil captar la tendencia de las observaciones y por tanto realizar un mejor ajuste del modelo. El caso es que en nuestro ejemplo, un número mayor de N , nos aporta más precisión captando mejor el valor de las observaciones aunque las predicciones estén más equiespaciadas en el tiempo.

6 Futuras Mejoras. Recomendaciones

En este capítulo final se abordan líneas que quedan abiertas tras este proyecto, y en las cuales se puede influir en un futuro. Cabe destacar algunas muy interesantes como pueden ser las siguientes:

- Probar a realizar la calibración optimizando otros tipos de Scores, por ejemplo, el Ignorance Score. Para ello, se pueden usar otros ajustes como el método de Máxima Verosimilitud, y ver si existen diferencias significativas. Por lo que se vio en el capítulo 3 no es esperable encontrar grandes diferencias siempre y cuando la hipótesis de normalidad de los vientos se siga cumpliendo, cosa que se comprueba que es así en la realidad.
- Realizar la optimización de varias variables de manera más realista teniendo en cuenta la restricción real de $c + dS^2 \geq 0$, para comprobar que no hay ningún problema al igual que señala la literatura.
- Cambiar las variables de optimización o añadir otras variables nuevas como pueden ser la presión y la temperatura. Incluso al cambiar las variables, puede probarse a modificar las distribuciones normales por otras que recojan mejor la tendencia real (véase tabla 3.3).
- Variar mucho más el número de casos, N , incluso en un horizonte temporal mucho más extenso, para ver si existen grandes diferencias con valores más pequeños del parámetro N .
- Incorporar el escenario de alta resolución a la calibración y comprobar si es verdad que éste aporta mucha más precisión que el resto de escenarios.
- Emplear otros tipos de Multivariate Rank que no sea el Multivariate Rank Average. El hecho de utilizar otro Rank cambiaría el criterio de reordenación del modelo calibrado, y por tanto, variaría el resultado del ECC.
- Utilizar otros métodos de calibración en lugar del Ensemble Copula Coupling. Uno de ellos es conocido en la literatura como el Ensemble Member a Member, contenido en la bibliografía de este proyecto (véase [4]).
- En cuanto a la trayectoria, puede probarse a variar el paso de la malla, para tener una idea de la variación que pueda suponer en los tiempos de vuelo cualquier refinamiento de la misma.
- Modificar la ruta de destino y origen para probar qué ocurre con otras combinaciones.
- Incorporar otros tipos de segmentos a la trayectoria: rectos y curvos con subidas, descensos y

aceleraciones, teniendo en cuenta sus respectivas ecuaciones y la presencia del viento en las mismas.

-Hacer que en la trayectoria se tengan en cuenta la variación de los parámetros con el tiempo, para distintos horizontes temporales, y comparar los resultados de unos con otros.

-Mezclar la optimización de variables realizadas con el ECC junto con la optimización de trayectorias, para encontrar una trayectoria que sea óptima, desde un punto de vista quizás algo más realista.

-Por último, se pueden emplear los métodos descritos en este proyecto, para numerosas aplicaciones en la que intervengan los vientos o cualquier variable atmosférica, y que además, se requiera de gran precisión.

Todas estas posibles modificaciones hacen que el proyecto sea bastante abierto y con una buena perspectiva de futuro, siempre siendo consciente de la dificultad que entraña realizar algunas de las propuestas.

7 Anexos

Se recogen aquí algunos códigos numéricos en Matlab que pueden ser útiles o interesantes.

A. Anexo PIT

```
%CÁLCULO DE HISTOGRAMAS PARA UNA NORMAL Y PARA UNA GUMBEL ASUMIENDO QUE  
LA  
%REALIDAD SE COMPORTA COMO UNA NORMAL  
clear all; close all; clc;  
%GENERAMOS 300 OBSERVACIONES  
mu_mu=25;  
var_mu=1;  
var_Y=9;  
mu_Y=mu_mu+sqrt(var_mu).*randn(300,1);  
Y=mu_Y+sqrt(var_Y).*randn(300,1);  
%AJUSTAMOS LA NORMAL Y LA GUMBEL  
mu_n=mean(Y)  
var_n=var(Y,1)  
%=====  
var_g=(6/pi^2)*var_n;  
mu_g=mu_n-double(eulergamma)*sqrt(var_g);  
%TRANSFORMAMOS MEDIANTE 1000 Y 1E6 SIMULACIONES CON LA CDF DEL MODELO  
%AJUSTADO  
N=1000000;  
mu_Y=mu_mu+sqrt(var_mu).*randn(N,1);  
Y=mu_Y+sqrt(var_Y).*randn(N,1);  
% Y=mu_n+sqrt(var_n).*randn(N,1);  
z_n=normcdf(Y,mu_n,sqrt(var_n));  
z_g=1-evcdf(-Y,-mu_g,sqrt(var_g));  
figure(1)  
hist(z_n,50)  
title('\bf{Normal Nob=1e6}')  
figure(2)  
hist(z_g,50)  
title('\bf{Gumbel Nob=1e6}')
```

```

%CÁLCULO DE HISTOGRAMAS PARA UNA NORMAL Y PARA UNA GUMBEL ASUMIENDO QUE
  LA
%REALIDAD SE COMPORTA COMO UNA GUMBEL
clear all; close all; clc;
%GENERAMOS 300 OBSERVACIONES
var_Y=9;
mu_Y=(25-double(eulergamma)*sqrt(var_Y))+sqrt(1).*randn(300,1);
Y=-evrnd(-mu_Y,sqrt(var_Y));
%AJUSTAMOS LA NORMAL Y LA GUMBEL
mu_n=mean(Y);
var_n=var(Y,1);
%=====
var_g=(6/pi^2)*var_n
mu_g=mu_n-(double(eulergamma)*sqrt(var_g))
%TRANSFORMAMOS MEDIANTE 1000 Y 1E6 SIMULACIONES CON LA CDF DEL MODELO
%AJUSTADO
N=1000000;
mu_Y=(25-double(eulergamma)*sqrt(var_Y))+sqrt(1).*randn(N,1);
Y=-evrnd(-mu_Y,sqrt(var_Y));
% Y=-evrnd(-mu_g,sqrt(var_g),N,1);
z_n=normcdf(Y,mu_n,sqrt(var_n));
z_g=1-ecdf(-Y,-mu_g,sqrt(var_g));
% z_g=1-ecdf(-Y,-(25-double(eulergamma)*3),3);
figure(1)
hist(z_n,50)
title('\bf{Normal Nob=1e6}')
figure(2)
hist(z_g,50)
title('\bf{Gumbel Nob=1e6}')

```

7.1 B. Cálculo de Scores

```

function [m,CRPS]=calcula_crps_normal(Yob,par_n)
Yob=sort(Yob);
N=length(Yob);
%CALCULO SU CRPS PARA EL CASO DE LA NORMAL
for j=1:N
CRPS(j)=quadgk(@(tau)myint_normal(tau,Yob(j),par_n),0,1,'AbsTol',1e-2,'
  RelTol',1e-2);
% CRPS1(j)=quadgk(@(tau)myint_normal(tau,Yob(j),par_n),0,1,'AbsTol',1e
  -4,'RelTol',1e-4);
end
% err=abs((norm(CRPS-CRPS1))/(norm(CRPS)))
m=mean(CRPS);
% plot(Yob,CRPS)
end
function int=myint_normal(tau,Yob,par_n)

```

```
int=2.*(norminv(tau,par_n(1),sqrt(par_n(2)))-Yob).*((1+sign(norminv(tau,
    par_n(1),sqrt(par_n(2)))-Yob)).*0.5-tau);
end
```

```
function [m,IGN]=calcula_ignorance_error_normal(Yob,par_n)
Yob=sort(Yob);
N=length(Yob);
%CALCULAMOS EL IGNORANCE SCORE PARA LA NORMAL
for i=1:N
    IGN(i)=-log(normpdf(Yob(i),par_n(1),sqrt(par_n(2))));
end
m=mean(IGN);
% plot(Yob,IGN)
end
```

```
function [m,AE]=calcula_absolute_error_normal(Yob,mediana_n)
Yob=sort(Yob);
N=length(Yob);
%CALCULAR EL ABSOLUTE ERROR ASOCIADO A UNA NORMAL
for i=1:N
    AE(i)=abs(mediana_n-Yob(i));
end
m=mean(AE);
% plot(Yob,AE)
end
```

7.2 C. ECC

```
% CALIBRACIÓN DE LOS MODELOS
clear all; clc; close all;
% ESTABLECEMOS EL NÚMERO DE VARIABLES, NUMERO DE CASOS Y NUMERO DE
    ESCENARIOS
K=5568; %número de variables
N=23;%número de casos
M=50; %número de escenarios

%1º Cargo las observaciones y 2º Me genero el conjunto de training y la
    muestra a calibrar
load genera_entrada.mat
% Xc(1,j)=(20+j*0.2)+3.*randn;
% Xc(2,j)=(17+j*0.2)+1.*randn;

%3º Me cálculo la media de cada una de la variables y la varianza
    muestral
Xb=mean(X,2);
Sc=var(X,1,2);
```

```

% 4º Entro en la caja negra que me calcula el crps y lo optimizo una vez
% para cada variable obteniendo distintos coeficientes óptimos
tic
for k=1:K
    problem = createOptimProblem('fmincon',...
    'objective',@(cf)calculo_crps(Xb,Y,Sc,k,cf),...
    'x0',[0 1 0 1],'lb',[-Inf 0 0 0],'ub',[],'options',...
    optimoptions(@fmincon,'Algorithm','sqp','Display','off','TolX',1e
    -5,'TolFun',1e-5));
    [cfopt(:,k),mopt(k)] = fmincon(problem);
%    gs = GlobalSearch('Display','iter','NumTrialPoints',5000)
%    [cfopt1(:,k),mopt1(k)] = run(gs,problem)
%    problem1 = createOptimProblem('fmincon',...
%    'objective',@(cf)calculo_crps(Xb,Y,Sc,k,cf),...
%    'x0',[0 1 0 1],'lb',[-Inf 0 0 0],'ub',[],'options',...
%    optimoptions(@fmincon,'Algorithm','interior-point','Display','off
%    ','TolX',1e-16,'TolFun',1e-9));
%    [cfopt1(:,k),mopt1(k)] = fmincon(problem1);
end

%4º Construyo una tabla de ordenación.
[X1,I]=sort(X,2);
[X2,Ord]=sort(I,2);

%5º Hago la media de la tabla de ordenación
Med=mean(Ord);

%6º Hago una reaordenacion de la media
[Med1,I1]=sort(Med,2);
[Med2,Ord1]=sort(I1,2);
t=toc
%7º Calculamos media y varianza de la muestra a calibrar y generamos la
%muestra equiprobable
Xbc=mean(Xc,2);
Scc=var(Xc,1,2);
for k=1:K
    mucal(k)=cfopt(1,k)+cfopt(2,k)*Xbc(k,1);
    varcal(k)=cfopt(3,k)+cfopt(4,k)*Scc(k,1);
    z(k,1:M)=norminv([1:M]/(M+1),mucal(k),sqrt(varcal(k)));
end
%
% 8º Reordenamos la muestra de manera que coincida con la muestra
original
for k=1:K
    zc(k,1:M)=z(k,Ord1(1,1:M,k));
end
save('resultados_optimizacion.mat','Xb','Sc','cfopt','mopt','Xbc','Scc
','mucal','varcal','z','zc');

```

```

%Función que me realiza el cálculo del CRPS para cada variable.
function chi=calculo_crps(Xb,Y,Sc,k,cf)
N=length(Y(:,k));
mu=cf(1)+cf(2).*Xb(:, :, k);
sigmac=cf(3)+cf(4).*Sc(:, :, k);
err=(Y(:,k)-mu)./sqrt(sigmac);
CRPS=sqrt(sigmac).*(err.*(2.*normcdf(err,0,1)-1)+2.*normpdf(err,0,1)-1/
sqrt(pi));
chi=mean(CRPS);
end

```

7.3 D. Trayectoria

```

%%TRAYECTORIA
function trayectoria_ode_mejorada
clear all; clc; close all;
tic
%COORDENADAS DE LOS PUNTOS EN LONGITUD Y LATITUD
long=[13.41 2.33 -3.70 -9.14 -25.67 -64.75 -80.21 -77.02]*pi/180;
lat=[52.52 48.87 40.42 38.71 37.75 32.33 25.78 38.90]*pi/180;
%PARA VER LA VUELTA
% long=long(8:-1:1);
% lat=lat(8:-1:1);
N=length(long);

%DATOS DE ALTURA DEL RADIO DE LA TIERRA Y DE LA ALTURA DE LOS CRUCEROS
R_E=6371*1000;
h=[12000 12000 12000 12000 12000 12000 12000];

%CÁLCULO DE LA LOXODRÓMICA
[chi,r]=loxodromica(long,lat,h,R_E);

%CREAMOS UNA MALLA EN PHI Y EN LAMBDA DE MANERA QUE ABARQUE TODAS LAS LOCALIZACIONES
Nl=96;
Np=29;
pa=1;
[phi,lambda]=meshgrid([25:pa:53]*pi/180, [-81:pa:14]*pi/180);
%=====
Nesc=50;
%=====

%CREAMOS LOS VALORES DE LA VELOCIDAD AERODINAMICA
M=[0.8 0.8 0.8 0.8 0.8 0.8 0.8];
[T,p,rho]=atmosfera_internacional(h);
gamma=1.4;
Ra=287.05;
a=sqrt(gamma*Ra*T);

```

```

V=M.*a;
%=====
%DEFINO LOS VECTORES DE VIENTO
load predicciones.mat
wu=wu(21:4:133,201:4:581,,:);
wv=wv(21:4:133,201:4:581,,:);
%ME PLANTEO RESOLVER PARA CADA UNO DE LOS SEGMENTOS EL ODE O LA INTEGRAL
for c=1:24
s=1;
for j=1:N-1
t0=zeros(1,Nesc);
r_x=[0 r(j)];
misopciones=odeset('AbsTol',1e-4,'RelTol',1e-4);
[R,T]=ode45(@velocidad_tierra,r_x,t0,misopciones,V,s,phi,lambda,R_E,h,
chi,lat,long,pa,wu,wv,c);
tam=size(T);
tseg(:,j)=T(tam(1),:);
s=s+1;
end
tseg;
tvuelo(:,c)=(sum(tseg'))';
end
tvuelo
mean(tvuelo)
sqrt(var(tvuelo,1))
save tvuelo.mat tvuelo
t1=toc
end
function f=velocidad_tierra(t,d,V,s,phi,lambda,R_E,h,chi,lat,long,pa,wu,
wv,c)
phi_v=lat(s)+d(1)./(R_E+h(s))*cos(chi(s));
lambda_v=long(s)+tan(chi(s))*(log(tan(pi/4-lat(s)/2)./tan(pi/4-phi_v
/2)));
if chi(s)==90*pi/180 || chi(s)==270*pi/180
lambda_v=long(s)+d(1).*sin(chi(s))/cos(lat(s))/(R_E+h(s));
end

%CALCULAR LOS VALORES DEL VIENTO EN CADA PUNTO
%CALCULAMOS SEGÚN LAS COORDENADAS DE LAS 4 ESQUINAS
phi_1=fix(phi_v*180/pi);
% phi_2=phi_1+pa;
lambda_1=fix(lambda_v*180/pi);
% lambda_2=lambda_1+pa;
%CALCULAMOS LOS VALORES DE I Y J QUE OCUPAN DICHOS PUNTOS DE LAS
%ESQUINAS
i=1+(phi_1-25)./pa;
j=1+(lambda_1+81)./pa;
%CALCULAMOS LAS 4 CONSTANTES DE INTERPOLACION
rho_x=lambda_v*180/pi-lambda_1;
rho_y=phi_v*180/pi-phi_1;

```

```

a1=(1-rho_x)*(1-rho_y);
a2=rho_x*(1-rho_y);
a3=(1-rho_x)*rho_y;
a4=rho_x*rho_y;
%=====
%CALCULAMOS EL VALOR DE LOS VIENTOS EN EL PUNTO DE INTERPOLACION
wu_int=a1*wu(i,j,:,c)+a2*wu(i+1,j,:,c)+a3*wu(i,j+1,:,c)+a4*wu(i+1,j
+1,:,c);
wv_int=a1*wv(i,j,:,c)+a2*wv(i+1,j,:,c)+a3*wv(i,j+1,:,c)+a4*wv(i+1,j
+1,:,c);
%REALIZAR EL GIRO DE AMBOS VECTORES PARA OBTENER VALORES DEL VIENTO
CRUZADO
%Y DEL VIENTO LONGITUDINAL
wl=wv_int*cos(chi(s))+wu_int*sin(chi(s));
wx=wv_int*sin(chi(s))-wu_int*cos(chi(s));
%=====
%CALCULO LA FUNCIÓN NORMAL ORIGINAL
f(:,1)=1/(sqrt(V(s)^2-wx.^2)+wl);
end

```


Referencias

- [1] TILMANN GNEITING, ADRIAN E. RAFTERY, ANTON H. WESTVELD III, AND TOM GOLDMAN. Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation. VOLUME 133 (2004). *Department of Statistics, University of Washington, Seattle, Washington*.
- [2] Tilmann Gneiting. Calibration of Medium-Range Weather Forecasts. *Heidelberg Institute for Theoretical Studies and Karlsruhe Institute of Technology*. March 18, 2014.
- [3] Thordis L. Thorarinsdottir. Forecast Evaluation I and II. *Heidelberg Institute*. October 4, 2017.
- [4] Michael Scheuerer. Statistical post-processing I and II: Univariate methods. *University of Colorado, CIRES, and NOAA/ESRL, Physical Sciences Division*. October 2017.
- [5] Tilmann Gneiting and Roopesh Ranjan. Comparing Density Forecasts Using Threshold and Quantile Weighted Scoring Rules. *Department of Statistics, University of Washington*. April 2008.
- [6] B. Efron, R. Tibshirani. O. Kirchkamp. An Introduction to the bootstrap (1993) and Resampling methods (2017).
- [7] Roman Schefzik. Ensemble Copula Coupling. *Ruprecht-Karls-Universität Heidelberg Fakultät für Mathematik und Informatik*. März 2011.
- [8] Roman Schefzik, Thordis L. Thorarinsdottir and Tilmann. Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling. *Institute of Mathematical Statistics, 2013*.
- [9] Antonio Franco, Damián Rivas, and Alfonso Valenzuela. Optimal Aircraft Path Planning Considering Wind Uncertainty. *Department of Aerospace Engineering, Universidad de Sevilla, Seville, Spain*.
- [10] Miguel Sillero Navajas. Optimización de trayectorias de avión sujetas a incertidumbre meteorológica y costes de navegación. *Dep. de Ingeniería Aeroespacial y Mecánica de Fluidos. Escuela Técnica Superior de Ingeniería. Universidad de Sevilla Sevilla, 2017*.
- [11] Sylvain Robert. Ensemble methods for data assimilation. 2014.
- [12] <http://fluidos.eia.edu.co/hidrologiai/probabilidad/probabilidad.htm>

- [13] *Guidelines on Ensemble Prediction Systems and Forecasting*. World Meteorological Organization (WMO), 2012.
- [14] *The ECMWF Ensemble Prediction System*. European Centre for Medium-Range Weather Forecasts (ECMWF), 2012.
- [15] R. Mladek and F. Vitart. (2016, Sep.) Models. [Online]. Available: <https://software.ecmwf.int/wiki/display/TIGGE/Models>.
- [16] Mercedes Cordero Romero. *Optimización de trayectorias sujetas a incertidumbre meteorológica*. Trabajo Fin de Máster. Sevilla, 2016.