



## Diseño de sistemas de predicción por conjuntos: error en las condiciones iniciales

# 16

DOI: [10.31978/014-18-009-X.16](https://doi.org/10.31978/014-18-009-X.16)

PAU ESCRIBÀ AYERBE

Delegación Territorial en Cataluña, *Agencia Estatal de Meteorología (AEMET)*

Aquel que quiera construir torres altas, deberá permanecer largo tiempo en los fundamentos.

ANTON BRUCKNER

El estado de la atmósfera en un momento dado no puede conocerse a la perfección. Por un lado, este hecho se debe a los errores inherentes a los *sistemas de observación*, a la distribución imperfecta y/o densidad insuficiente de las redes correspondientes y a los intervalos de tiempo limitados en que dichas observaciones se toman. Por otro lado, los *sistemas de asimilación de datos* para introducir estas observaciones en los modelos atmosféricos tampoco son perfectos, de modo que cada sistema de asimilación está afectado por errores característicos de las observaciones y, también, del modelo de corto plazo usado para brindar una predicción previa (*background* o *first guess*) al propio sistema de asimilación, predicción previa que se ajustará con nuevas observaciones. Todo ello conforma los así llamados *errores en las condiciones iniciales*. Estos errores o incertidumbres, aunque son pequeños, tal como hemos visto en capítulos anteriores crecen de modo no lineal según la dinámica caótica que describe la atmósfera. Presentamos en este capítulo una introducción a este apasionante tema del tratamiento del error en el estado inicial en predicción por conjuntos.

**Palabras clave:** diseño de sistemas de predicción por conjuntos, incertidumbres en las condiciones iniciales, asimilación de datos, método de vectores singulares, método de asimilación por conjuntos, método LETKF.

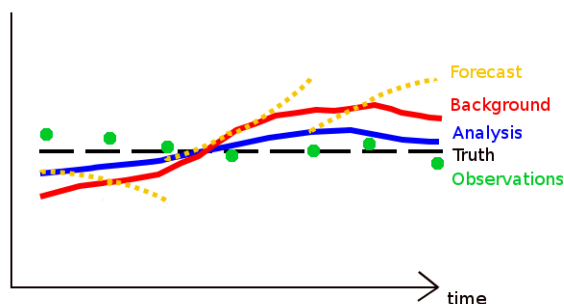


Figura 16.1: Esquema conceptual de predicción previa (background), observaciones, análisis y verdad (truth) en asimilación de datos (ver texto).

## 16.1 Análisis y su error

La resolución de la evolución de la atmósfera en el tiempo es un problema, matemáticamente hablando, de valores iniciales. Es decir, dado un *estado inicial* atmosférico, la integración numérica de las ecuaciones de evolución, nos da el estado atmosférico futuro. Por tanto, definir el estado inicial atmosférico es una cuestión fundamental. La *asimilación de datos* (sec. 10.3 en la página 133) es la parte de la predicción numérica que tiene como objetivo definir este estado inicial. Para ello, combina de una manera inteligente una predicción a corto plazo o estimación previa, con las observaciones disponibles en el instante inicial. El resultado de la asimilación es lo que llamamos *análisis*.

Existen distintos algoritmos para generar el análisis, aunque los más usados son los métodos *variacionales* (3DVAR y 4DVAR) y los métodos por *filtros de KALMAN* (ETKF, LETKF, ...). En la sección 16.2 se presenta la formulación básica de estos métodos, en este apartado únicamente se presenta una descripción conceptual. Un aspecto importante de estos algoritmos es que lo que todos persiguen es combinar *observaciones* y *predicción previa* teniendo en cuenta sus errores respectivos, de tal modo que el resultado final, el análisis, tenga un error mínimo.

La Figura 16.1 describe de una manera conceptual los distintos componentes del análisis. Cabe destacar, como se ve en la figura que, con el análisis, lo que queremos es acercarnos lo más posible a la *verdad*, que es siempre desconocida. Otro aspecto interesante de la figura es entender el concepto de *error* en los distintos componentes del análisis. Tanto para las

observaciones, como para la predicción previa o el análisis, el error respectivo es aquello que separa su valor de la verdad. Aquí debemos pararnos en el razonamiento, porque la última frase es lo que conecta el análisis con la predicción por conjuntos. Una de las cosas que queremos hacer en predicción por conjuntos es *estimar* de alguna manera el *error del análisis*. Para ello definimos distintos estados iniciales posibles dentro del margen de error del análisis. Esto constituye el problema de estimar la incertidumbre del estado inicial en *predicción por conjuntos*.

Llegado a este punto uno se debe preguntar *cómo* podemos estimar el error del análisis o cómo podemos construir los distintos estados iniciales que deberían estar en este margen de error. Resolver este problema no es sencillo, de hecho, se han propuesto multitud de técnicas con más o menos éxito tanto desde el punto de vista de mejora de la predicción como del uso de recursos computacionales. Dado que la lista es larga, en este capítulo vamos a resumir de manera breve las técnicas usadas en nuestro entorno de trabajo operativo más cercano, que básicamente es el ECMWF (cap. 19 en la página 289), los consorcios HIRLAM-ALADIN (cap. 20 en la página 301) y la Agencia Estatal de Meteorología (AEMET).

## 16.2 Técnicas de generación de estados iniciales

Tal como hemos dicho, las técnicas que describimos a continuación son las que se utilizan operativamente en nuestros centros de trabajo de referencia. Es importante destacar que siempre estamos hablando de estado inicial *atmosférico*, o sea, no estamos teniendo en cuenta, por ejemplo, el estado inicial en la superficie o en el mar (u océano). De cualquier modo estas técnicas o equivalentes se podrían usar también en estos casos.

### 16.2.1 Multimodelo

Esta técnica también conocida como *poor man's ensemble* o en español como del *conjunto de hombre pobre* [3, 16], es una tecnificación de la clásica práctica usada por los predictores para hacer la predicción (sec. 13.2.2 en la página 168), práctica que a menudo llevaba el mismo nombre. Se trata de consultar varios modelos para una misma situación y hacerse una composición de la predicción más probable del



conjunto. Básicamente, los estados iniciales se crean tomando los análisis o las predicciones previas de distintos modelos globales interpolados a nuestro dominio de interés. Un problema evidente de esta técnica es que el número de estados iniciales está limitado por el número de modelos globales disponibles. En el caso de la **AEMET**, cuyos *sistemas de predicción por conjuntos* (SPC) han usado históricamente esta técnica, se combinan los modelos globales con distintos modelos de alta resolución, multiplicando el número de estados iniciales por el número de modelos locales. En la Figura 16.2 se muestra un esquema de los distintos estados iniciales creados con esta técnica en el antiguo AEMET-SREPS [6] (cap. 21 en la página 313). Tal como se muestra en el cuadro, aquí se toman predicciones de modelos globales de referencia

a nivel mundial. Los modelos locales utilizados son el propio de **AEMET** en su momento, el HIRLAM, el HRM y el LM (Servicio Meteorológico Alemán), el UM (Servicio Meteorológico Británico) y el MM5 (de la Universidad de Pennsylvania y el Centro Nacional americano para Investigaciones Atmosféricas PSU/NCAR). Esta técnica ha sido implementada en el actual sistema de alta resolución AEMET- $\gamma$ SREPS (cap. 22 en la página 333).

Esta técnica tiene tres aspectos criticables. Primeramente, es de esperar una cierta correlación entre los miembros del conjunto formados con el mismo modelo local, por lo que a priori la condición de equiprobabilidad (sec. 13.7.2 en la página 191) entre los miembros se rompe.

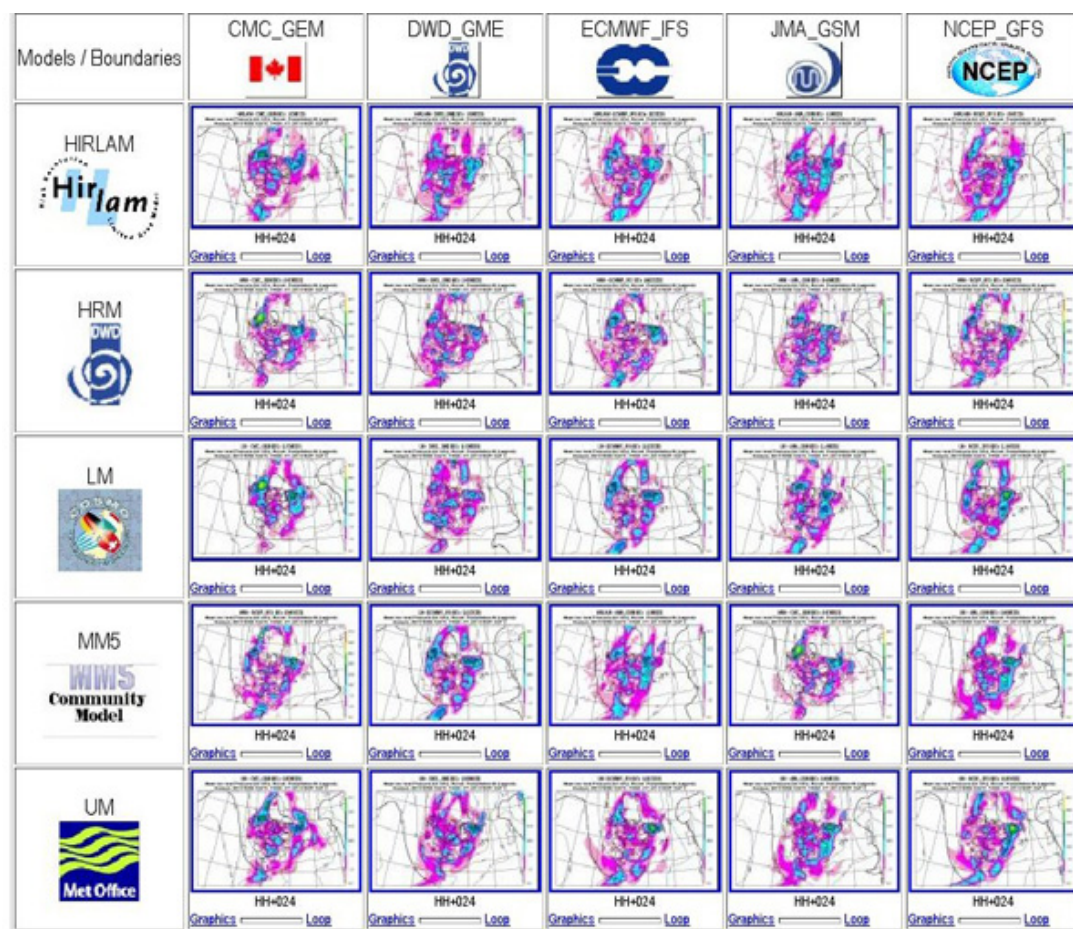


Figura 16.2: Estados iniciales resultantes de combinar 5 modelos globales operativos con 5 modelos locales. Esta es la configuración del (antiguo) AEMET-SREPS (ver texto).

Por otro lado, ciertamente el uso de distintos modelos no deja de ser en cierto sentido una *caja negra* cuando miramos el comportamiento estadístico del conjunto: no sabemos, científicamente, cómo se estima la incertidumbre de la atmósfera. Finalmente, el mantenimiento de un sistema con tantos modelos distintos es complicado desde el punto de vista técnico, así como la dependencia de la disponibilidad (y comunicación) de los modelos globales en los centros de origen.

A pesar de sus aspectos criticables esta técnica, aporta numerosas ventajas desde el punto de vista práctico, por lo que existen numerosos partidarios, entre los cuales nos contamos en AEMET. Un argumento principal es que esta técnica da excelentes resultados [6]. Un segundo argumento que apoyaría el uso de esta técnica es que, al contener cada miembro del conjunto los mismos errores sistemáticos en cada integración, la corrección de los mismos con técnicas de posproceso debería ser adecuada y, por tanto, se podrían potencialmente mejorar los resultados. Finalmente, el usar modelos globales para fines operativos significa que éstos están altamente verificados y en continua mejora en centros operativos.

### 16.2.2 Local Ensemble Transform Kalman Filter (LETKF)

El método Local Ensemble Transform Kalman Filter, LETKF [10], para asimilación de datos genera por el mismo precio un análisis para usarlo en predicción determinista y un conjunto de análisis que pueden ser usados como conjunto de estados iniciales para una predicción probabilista. El código LETKF ha sido implementado por el autor de este capítulo en el modelo HARMONIE-AROME adaptando el código desarrollado por MATS HAMRUD (ECMWF [7]) para el modelo global IFS. El LETKF pertenece a la familia de algoritmos por *filtros de KALMAN*. Para una muy buena explicación de la implantación del algoritmo en el modelo IFS del ECMWF y de sus resultados en comparación con el 4DVAR, se remite al lector a los artículos de HAMRUD Y COL. [7] y BONAVITA Y COL. [1], ambos de 2015. Como todos los algoritmos de asimilación, LETKF trata de minimizar la llamada *función de coste J* (sec. 16.1). Esta técnica consiste en encontrar aquel valor del estado atmosférico  $\mathbf{x}$  que haga el valor  $\mathbf{J}$  mínimo, dados los distintos componentes de la ecuación. Para encontrar  $\mathbf{x}$  LETK

hace unas consideraciones propias que se describen a continuación.

**Local Ensemble Transform Kalman Filter, LETKF [10].** Como todos los algoritmos de asimilación, LETKF trata de minimizar la llamada *función de coste J* (sec. 16.1). Esta técnica consiste en encontrar aquel valor del estado atmosférico  $\mathbf{x}$  que haga el valor  $\mathbf{J}$  mínimo, dados los distintos componentes de la ecuación. Dicho estado atmosférico será el análisis. Los otros componentes de la ecuación son, la estimación previa  $\mathbf{x}^b$ , el error de la estimación previa  $\mathbf{P}^b$ , las observaciones  $\mathbf{y}$ , el error de las observaciones  $\mathbf{R}$  y el operador observación  $\mathbf{H}$ .

$$\mathbf{J}(\mathbf{x}) = (\mathbf{x}^b - \mathbf{x})^T (\mathbf{P}^b)^{-1} (\mathbf{x}^b - \mathbf{x}) + (\mathbf{y} - \mathbf{H}(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}(\mathbf{x})) \quad (16.1)$$

La diferencia más importante entre los métodos de *filtros de KALMAN* y los métodos *variacionales* radica en cómo se aproximan los componentes  $\mathbf{x}^b$  y  $\mathbf{P}^b$ . En 16.2 se describen estos componentes en el caso de los Filtros de Kalman.

$$\mathbf{x}^b = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^b \quad \mathbf{P}^b \approx \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i^b - \overline{\mathbf{x}^b}) (\mathbf{x}_i^b - \overline{\mathbf{x}^b})^T \quad (16.2)$$

En este caso, la predicción previa  $\mathbf{x}^b$  es la media de un conjunto de predicciones previas. Su error  $\mathbf{P}^b$ , así mismo, se construye a partir de este conjunto de predicciones. Luego los *filtros de KALMAN* utilizan un conjunto de predicciones previas en la resolución del propio algoritmo de asimilación. En el caso de los métodos variacionales,  $\mathbf{x}^b$  es una predicción determinista a corto plazo y  $\mathbf{P}^b$  es el resultado de una estimación estadística climatológica con predicciones previas, y por lo tanto, no dependiente del flujo atmosférico.

Si nos fijamos en particular en el LETKF y sin ahondar en detalle en todas las ecuaciones del método [10], el análisis determinista, así como el conjunto de análisis resultantes para cada miembro, se describen en las ecuaciones 16.3.

$$\mathbf{x}^a = \overline{\mathbf{x}^b} + \mathbf{X}^b \overline{\mathbf{w}^a} \quad \mathbf{x}_i^a = \overline{\mathbf{x}^b} + \mathbf{X}^b (\overline{\mathbf{w}^a} + \mathbf{w}_i^a) \quad (16.3)$$

El algoritmo calcula una serie de pesos (*weights*),  $\mathbf{w}_i$ , donde está toda la información proveniente de las observaciones. Otro aspecto interesante del algoritmo es que el análisis determinista es la media de los análisis del conjunto (ecuación 16.4):

$$\mathbf{x}^a = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^a \quad (16.4)$$

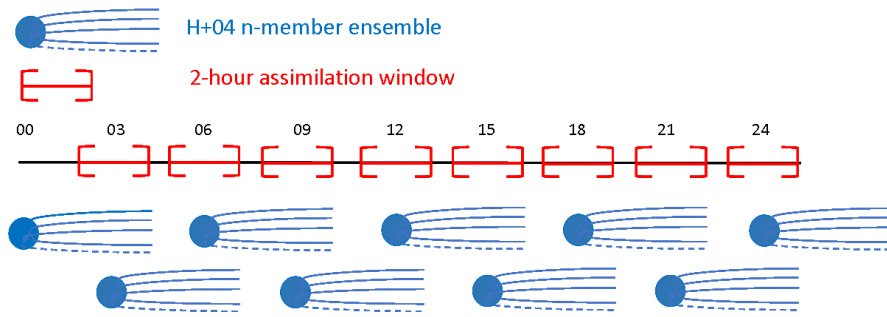


Figura 16.3: Ciclo de análisis del LETKF con el HARMONIE-AROME (ver texto).

Si nos fijamos en la implementación técnica en el modelo HARMONIE-AROME, LETKF asimila observaciones en un ciclo de análisis de 3 horas, a lo largo de una ventana de asimilación centrada de 2 horas. El parámetro *inflación* se introduce para aproximar aquellos errores no estimados en el algoritmo. Por otro lado, el parámetro *localización*, que permite seleccionar las observaciones más cercanas al punto de análisis, está en el diseño intrínseco del mismo. Así mismo, LETKF está diseñado e implementado para asegurar que su eficiencia computacional sea muy elevada en comparación con los métodos variacionales. En particular, resuelve el análisis independientemente en cada punto de la rejilla del modelo, por lo que acepta muy bien la *paralelización* (cap. 11 en la página 145). Este último aspecto es quizás el más destacable en el entorno de trabajo de alta resolución en el que nos movemos en la actualidad. En la Figura 16.3 se muestra un esquema del ciclo de análisis del LETKF. La idea es integrar una predicción por conjuntos a corto plazo (4 horas) que cubre la ventana de asimilación, donde residen las observaciones a tener en cuenta. Posteriormente se resuelve el algoritmo LETKF y se obtienen los análisis (sec. 16.4 en la página anterior) en el centro de la ventana, con los que volvemos a integrar el conjunto hasta el siguiente ciclo. Dada la flexibilidad y eficiencia del método, está planeado que LETKF sea el algoritmo que permita la ingestión de todo tipo de observaciones en el actual sistema de predicción por conjuntos AEMET-γSREPS.

### 16.2.3 Método NMC

El método NMC toma el nombre del antiguo *National Meteorological Centre* (hoy llamado *NCEP*), o Servicio Meteorológico Estadounidense. En concreto, este método consiste en estimar los errores de la predicción a partir de las *diferencias* entre dos predicciones

inicializadas en tiempos distintos pero válidas para el mismo tiempo de predicción. Podemos, entonces, aproximar el error del análisis de esta manera. Entonces, si calculamos el análisis mediante un método clásico (por ejemplo el 3DVAR), podemos añadir a este análisis este error o perturbación, con lo que obtenemos una nueva estimación del estado inicial. Este es el método de referencia con el que se construye el conjunto de predicciones en el HARMONIE-AROME-EPS.

**Método NMC.** La siguiente ecuación hace explícito el cálculo descrito previamente:

$$\mathbf{x}_i^a = \mathbf{x}_0^a + k_i (\mathbf{x}_{t_1}^b - \mathbf{x}_{t_2}^b)_i \quad (16.5)$$

donde  $\mathbf{x}_0$  es el llamado *análisis de control*, que en el caso que nos ocupa es el 3DVAR, pero podría ser otro. El factor  $k$  es un factor *ad hoc* de amplificación o reducción de la diferencia. Se pueden tomar valores de este factor positivos y negativos, construyendo así los miembros simétricos respecto al control. Finalmente  $\mathbf{x}_{t_1}^b$  y  $\mathbf{x}_{t_2}^b$  son las predicciones previas válidas ambas en el mismo tiempo (tiempo del análisis) e inicializadas en los instantes  $t_1$  y  $t_2$ .

**3DVAR.** Puesto que aquí lo usamos para generar el análisis de control, es oportuno dedicar un momento a describir brevemente el algoritmo 3DVAR. Como el LETKF, se trata de resolver la ecuación 16.1 en la página anterior. Describas las distintas componentes de la función de coste  $\mathbf{J}$ , en el caso de los métodos variacionales, después de una serie de métodos de resolución numéricos [5], el resultado del análisis 3DVAR se describe en la ecuación 16.6. En este caso la matriz  $\mathbf{K}$  se llama *matriz de ganancia* o *matriz de KALMAN* y contiene una descripción del peso que se le debe dar a la información obtenida por las observaciones, en el análisis.

$$\mathbf{x}_{3DVAR}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - \mathbf{H}(\mathbf{x})) \quad (16.6)$$

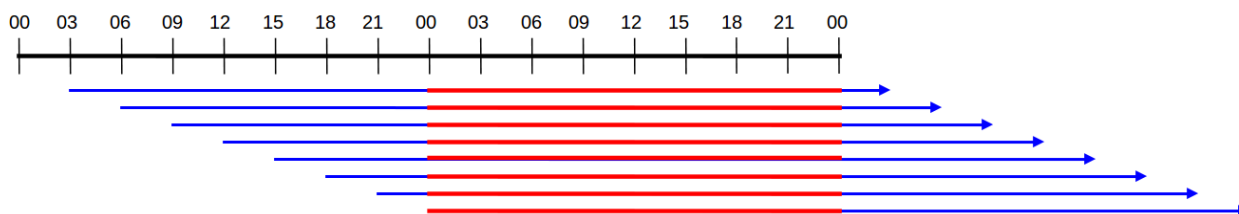


Figura 16.4: Conjunto formado con la técnica time lagged ensemble. En este caso, partimos de distintas pasadas de el mismo modelo determinista (HARMONIE-3DVAR) a 48 horas. Los segmentos rojos corresponden a las trayectorias que forma la predicción por conjuntos a 24 horas inicializada a las 00 UTC (ver texto).

Volviendo a la aplicación práctica del método NMC en el caso de HARMONIE-AROME, los valores de referencia para  $k_i$ ,  $t_1$  y  $t_2$  que estamos tomando para generar los análisis en un conjunto de 10 miembros, están resumidos en la Tabla 16.1.

#### 16.2.4 Lagged ensemble

Otro método que se usa con frecuencia para obtener distintos estados iniciales es el llamado, en inglés, *lagged ensemble* [4, 9, 13, 15]. En este método se definen directamente los estados iniciales a partir de las predicciones previas inicializadas en tiempos anteriores, validas en el mismo tiempo, el tiempo inicial del conjunto. Este método es extremadamente sencillo, y ha demostrado buenos resultados en el caso de combinar por ejemplo dos conjuntos de predicción inicializados en dos pasadas consecutivas. En este caso, el hecho de doblar el número de miembros del conjunto a coste cero, mejora la predicción probabilista.

miembro	K	$t_1$ (h)	$t_2$ (h)
1	1.75	$t_0$	$t_0 - 6$
2	-1.75	$t_0$	$t_0 - 6$
3	1.5	$t_0 - 6$	$t_0 - 12$
4	-1.5	$t_0 - 6$	$t_0 - 12$
5	1.2	$t_0 - 12$	$t_0 - 18$
6	-1.2	$t_0 - 12$	$t_0 - 18$
7	1.0	$t_0 - 18$	$t_0 - 24$
8	-1.0	$t_0 - 18$	$t_0 - 24$
9	0.9	$t_0 - 24$	$t_0 - 30$
10	-0.9	$t_0 - 24$	$t_0 - 30$

Tabla 16.1: Valores de referencia en HARMONIE-AROME-EPS para construir un conjunto de estados iniciales con el método NMC (ver texto).

En el grupo de predecibilidad de AEMET realizamos una verificación de un conjunto formado con este método a partir únicamente de la predicción determinista operativa de HARMONIE-AROME. Esta predicción está basada en ciclos de asimilación 3DVAR cada 3 horas, y las predicciones llegan hasta 48 horas. Con estos valores, podemos construir un conjunto de 8 miembros con una predicción a 24 horas, tal y como muestra la Figura 16.4.

Los resultados de dicha verificación así como de otros métodos para generar condiciones iniciales, serán presentados en la sección 16.3 en la página 250.

#### 16.2.5 Ensemble Data Assimilation

Con la Asimilación de Datos por Conjuntos (*Ensemble Data Assimilation* o EDA), se introduce un conjunto de predicciones en el proceso de asimilación, de forma parecida a como se hace con los *filtros de KALMAN*. EDA es la técnica de referencia en el ECMWF, tanto para asimilación de datos como para predicción por conjuntos [11, 12], en el segundo caso en combinación con los *vectores singulares*, que se presentan en el próximo apartado.

La idea teórica del EDA se basa en estimar las incertidumbres fundamentales en el proceso de predicción-asimilación, mediante *perturbaciones estocásticas* teniendo en cuenta una *estimación de las distintas fuentes de error*. En total existen 3 fuentes de error en la predicción: el *error de las observaciones*, el *error del modelo* y el *error en las condiciones de contorno*. Entonces, de lo que se trata es de construir varias trayectorias de predicción-asimilación independientes en las que los 3 componentes son perturbados estocásticamente.



**Perturbaciones estocásticas.** Sobre el tipo de perturbación estocástica existen distintos métodos, basándose todos ellos en la aleatoriedad. Para entender mejor este método, podemos pensar en un observación ( $o$ ) y en su error ( $\varepsilon_o$ ). Entonces, si asumimos una forma para la distribución de probabilidad de esta variable (gaussiana por ejemplo si es temperatura), podemos construir un conjunto de 10 observaciones ( $o_i$ ) igualmente válidas para ese punto del espacio-tiempo, sumándoles un parámetro estocástico proporcional a su error ( $k_i$ ), según la ecuación:

$$o_i = o + k_i \varepsilon_o \quad i = 1, \dots, n \quad (16.7)$$

Una vez tenemos una muestra de, por ejemplo,  $n = 10$  conjuntos de observaciones, 10 realizaciones del error del modelo y 10 campos superficiales todos ellos válidos para el mismo punto del espacio-tiempo, para cada uno de ellos generamos un ciclo de análisis 4DVAR independiente. El conjunto de estos análisis independientes será el conjunto de estados iniciales para nuestra predicción probabilista.

En cuanto a la estimación del error del modelo (sec. 5.3 en la página 62), presente tanto en predicción como en asimilación, el ECMWF ha propuesto una técnica también estocástica llamada Stochastic Perturbation of Parametrization Tendencies, SPPT [14], que ha dado buenos resultados [17].

### 16.2.6 vectores singulares

Los *vectores singulares* o *singular vectors* [2] conforman históricamente el primer método de generación de estados iniciales para el Sistema de Predicción por Conjuntos del ECMWF, el ECENS (sec. 19.3 en la página 293). Aunque a día de hoy este método todavía se utiliza en combinación con EDA, paulatinamente va a ser substituido por este último método en todo aquello que haga referencia a la estimación de incertidumbres del estado inicial atmosférico en el ECMWF.

**Vectores singulares.** La ecuación siguiente describe la ecuación de evolución temporal del estado atmosférico  $\mathbf{x}$ .

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \quad (16.8)$$

Si tomamos una pequeña perturbación  $\delta\mathbf{x}$  respecto al estado atmosférico  $\mathbf{x}$  y aplicamos el desarrollo en serie de TAYLOR, obtenemos:

$$\frac{d(\mathbf{x} + \delta\mathbf{x})}{dt} = \mathbf{F}(\mathbf{x}) + \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \delta\mathbf{x} + \mathbf{O}(\mathbf{x}^2) \quad (16.9)$$

Si integramos ahora entre el tiempo  $t_0$  y el tiempo  $t$ , obtenemos la siguiente ecuación:

$$\mathbf{x}(t) + \delta\mathbf{x}(t) \approx \mathbf{G}(\mathbf{x}(t_0)) + \mathbf{M}_{t_0 \rightarrow t} \delta\mathbf{x}(t_0) \quad (16.10)$$

donde  $\mathbf{M}$  es el llamado propagador lineal tangente (la linealización de las ecuaciones del modelo). Por lo tanto, la perturbación  $\delta\mathbf{x}$  evoluciona en el tiempo según:

$$\delta\mathbf{x}(t) = \mathbf{M}_{t_0 \rightarrow t} \delta\mathbf{x}(t_0) \quad (16.11)$$

Ahora podemos definir el *crecimiento de la perturbación*  $\sigma^2$  según la ecuación:

$$\begin{aligned} \sigma^2 &= \frac{\langle \delta\mathbf{x}(t), \delta\mathbf{x}(t) \rangle}{\langle \delta\mathbf{x}(t_0), \delta\mathbf{x}(t_0) \rangle} \\ &= \frac{\langle \mathbf{M}_{t_0 \rightarrow t}^T \mathbf{M}_{t_0 \rightarrow t} \delta\mathbf{x}(t_0), \delta\mathbf{x}(t_0) \rangle}{\langle \delta\mathbf{x}(t_0), \delta\mathbf{x}(t_0) \rangle} \end{aligned} \quad (16.12)$$

Donde  $\langle, \rangle$  es el producto escalar entre vectores. Resulta que el *crecimiento máximo* de las perturbaciones en 16.12 se dará en las direcciones definidas por los vectores singulares del producto de matrices  $\mathbf{M}^T \mathbf{M}$ . Estas direcciones nos dan los *vectores singulares*.

Para una explicación matemática completa de este método, donde se añade el concepto de norma en la definición de los vectores singulares, se remite a la lectura del artículo de BUIZZA y PALMER de 1995 [2].

Una explicación sencilla del significado de los *vectores singulares* es la siguiente. Si pensamos en un estado inicial atmosférico y en su evolución temporal, podemos identificar qué zonas de ese estado contienen un mayor crecimiento de energía, asociado a la inestabilidad, del sistema atmosférico. Podemos pensar por ejemplo en zonas de inestabilidad baroclina tales como sistemas de bajas presiones o estructuras frontales. Entonces, los vectores singulares son una herramienta matemática que permite identificar estas zonas de mayor crecimiento energético. Una vez calculados estos vectores, a partir de ellos construimos perturbaciones del estado inicial que sumaremos al análisis de control (4DVAR) para obtener los distintos estados iniciales de nuestro conjunto.

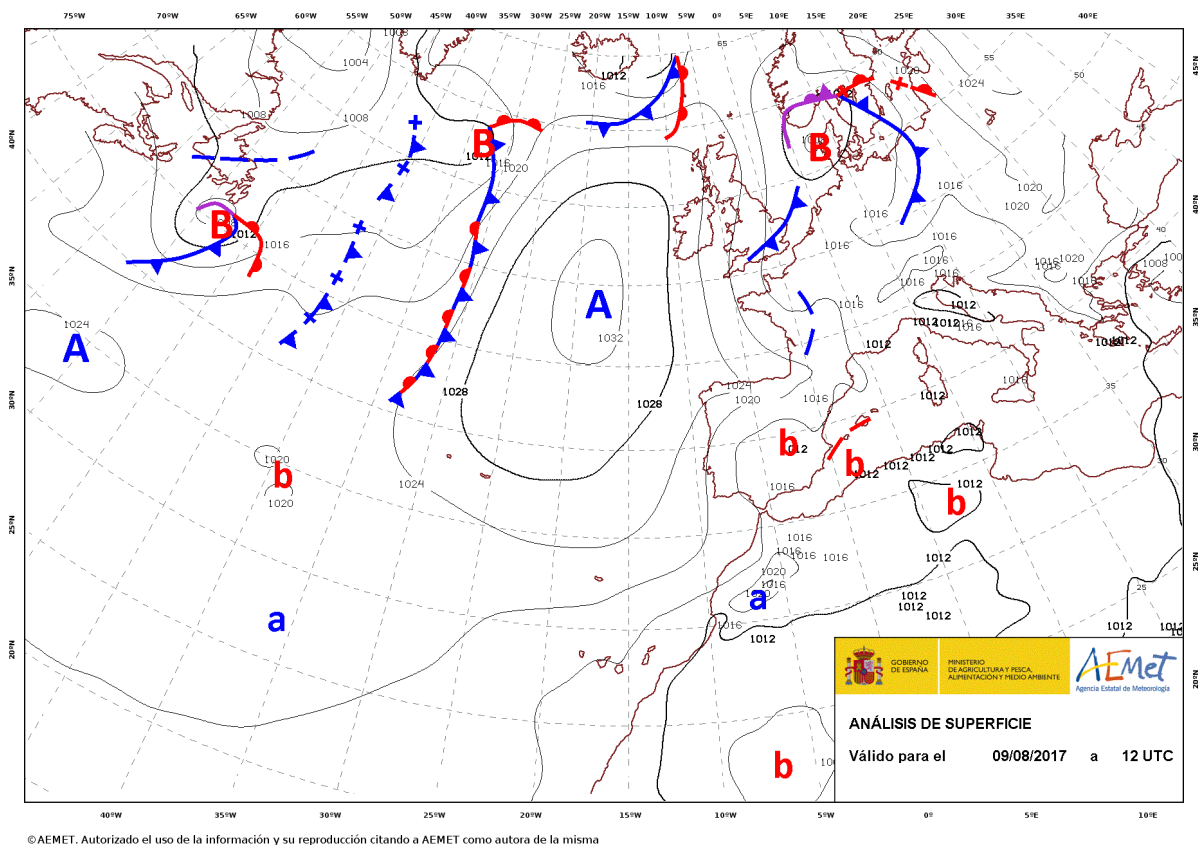


Figura 16.5: Análisis sinóptico de superficie para el día 9 de agosto de 2017 a las 12 UTC (AEMET).

### 16.3 Verificación de las técnicas disponibles en AEMET

En el taller *Predicción probabilista de fenómenos de tiempo severo*, que se celebró en la sede central de AEMET en Madrid en octubre de 2017, el grupo de predecibilidad de AEMET presentó un estudio donde se comparan los conjuntos de predicciones que se pueden generar en AEMET (que no operativos) usando 4 de las técnicas presentadas anteriormente. Los conjuntos se forman siempre con modelos de alta resolución, a 2.5 km en la horizontal. En este apartado se presentan los resultados más significativos de dicho estudio.

#### 16.3.1 Técnicas de generación de conjuntos disponibles en AEMET

El grupo de predecibilidad de AEMET está integrando operativamente en los ordenadores del ECMWF el AEMET- $\gamma$ SREPS (cap. 22 en la página 333) desde abril del 2016. Además de este sistema, tenemos las

herramientas para generar otros 3 sistemas de conjuntos mediante 3 de las técnicas presentadas en el apartado anterior: el LETKF, el método NMC y el lagged ensemble. Estos tres métodos están únicamente basados en el modelo HARMONIE-AROME, de cuyo consorcio AEMET es socio. En este ejercicio de verificación se van a comparar estas 4 técnicas para un mismo periodo temporal en un dominio que cubre la península ibérica. En este apartado sólo se presentarán resultados de lo que llamamos *verificación objetiva*. Se remite al lector al capítulo 22.7 en la página 351 para ver un par de casos de estudio de este mismo periodo utilizando lo que llamamos *verificación subjetiva*.

#### 16.3.2 Descripción de los experimentos

Durante una semana del mes de agosto de 2017, desde el 8 al 15, se integraron los 4 conjuntos de predicciones, tanto a las 00 como a las 12 UTC, hasta 24 horas de predicción. En esa semana un potente anticiclón situado sobre el Atlántico Norte, abarcando latitudes desde las islas británicas hasta las Canarias, marcaba un predominio de las altas presiones en toda la zona,



con la baja térmica estival sobre la península ibérica. La situación sinóptica sobre la Península para el día 9 a las 12 UTC se muestra en la Figura 16.5 en la página anterior. Como fenómenos significativos del periodo cabe destacar la convección diurna en algunos lugares de la Meseta Central, así como la convección asociada a algún pequeño frente en el Mediterráneo.

Tal y como se ha presentado con anterioridad, todos los conjuntos salvo AEMET-γSREPS asimilan observaciones en un ciclo de 3 horas. Las observaciones asimiladas son las llamadas convencionales (estaciones automáticas, manuales, barcos, boyas y radiosondeos). El número de miembros para los conjuntos es 20, salvo para el *lagged ensemble* que por construcción sólo puede tener 8 miembros (Figura 16.4 en la página 248). Las predicciones están hechas en un dominio centrado sobre la península ibérica con 2.5 km de resolución horizontal y  $576 \times 480 \times 65$  puntos de rejilla. En la Figura 16.6 se muestra el dominio de integración.

### 16.3.3 Verificación

Es conveniente comparar las predicciones con las observaciones a lo largo de un periodo suficientemente largo para que la muestra estadística genere datos significativos. Aunque dicha significación puede estimarse estadísticamente, aquí no la mostramos por simplicidad. Asumimos valores altos de significación

para las variables en superficie, donde tenemos alrededor de 13 000 parejas predicciones-observación. En cambio, en la vertical los valores de significación serán más pequeños, puesto que los datos de radiosondeos son muy inferiores en número. En total tenemos alrededor de 100 parejas predicciones-observación en la vertical.

Los parámetros del modelo cuya verificación presentamos son: velocidad (módulo) del viento a 10 metros (S10m), temperatura a 2 metros (T2m), humedad relativa a 2 metros (HR2m), presión al nivel del mar (PNM) y temperatura (T), humedad específica (Q) y módulo del viento (S) en la vertical. Como *verdad* de lo que sucedió tomamos las observaciones convencionales.

Como estamos en el marco de la predicción probabilista, para cada punto de observación tendremos  $n$  valores de predicción ( $n = 20, 20, 20$  y  $8$ ). Esto cambia el paradigma de verificación. Se remite al lector al capítulo 15 en la página 207 para más información sobre este marco de trabajo. Presentamos resultados de verificación usando tres métodos comunes de verificación probabilista: *dispersión-error*, *histogramas de rango* y *Continuous Ranked Probability Score* (CRPS, [8]). Estos métodos ofrecen un diagnóstico de rasgos fundamentales en predicción probabilista como la *amplitud*, la *consistencia* y la *respuesta frente a eventos binarios* en la estimación de la incertidumbre (sec. 15.9 en la página 224).



Figura 16.6: Dominio de integración de los distintos experimentos.

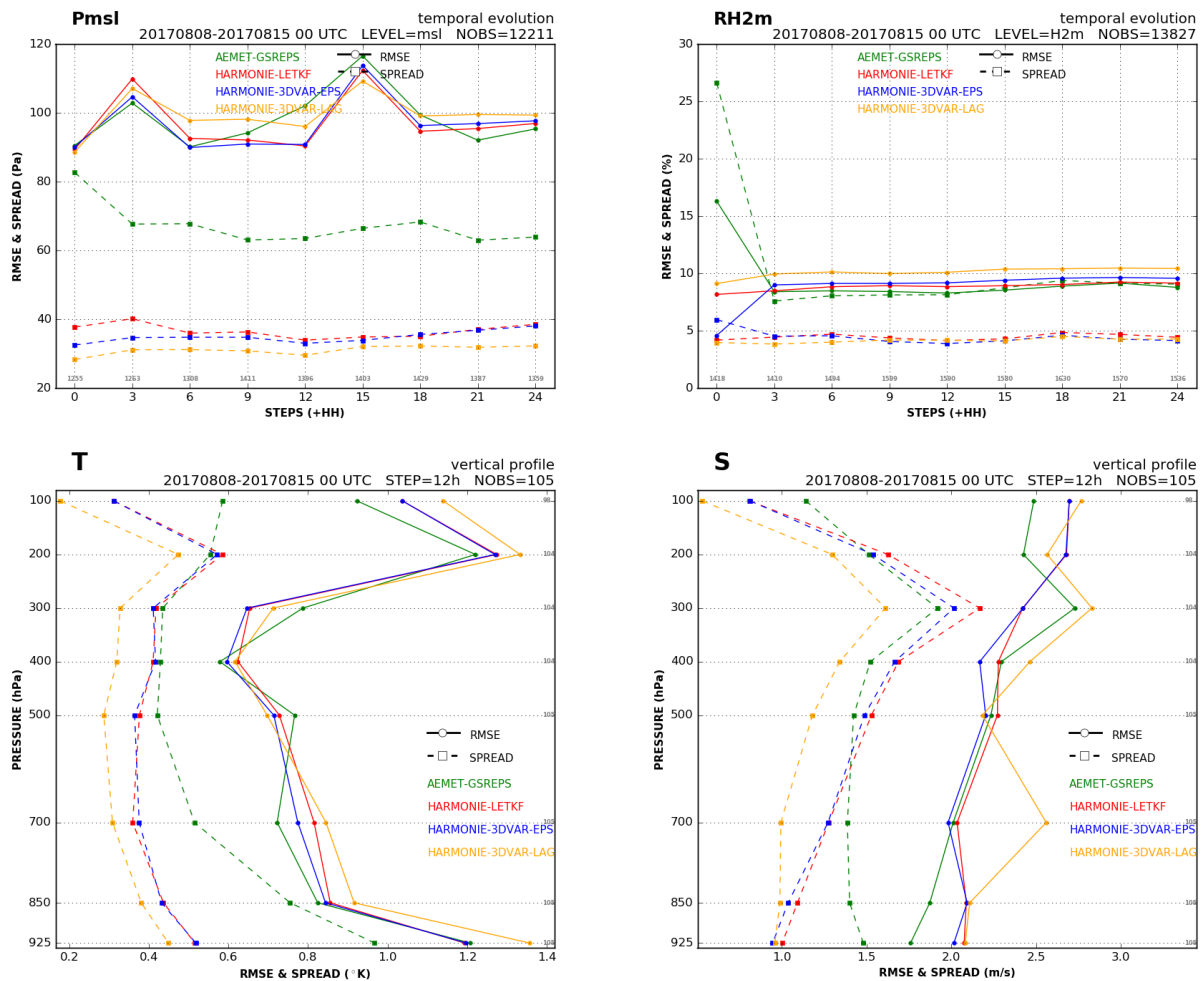


Figura 16.7: Relaciones dispersión-error para presión reducida al nivel del mar y humedad relativa a 2 m (arriba) y temperatura y módulo del viento en el alcance H+12 en la vertical (abajo) (ver texto).

En la Figura 16.7 se muestra claramente como el AEMET- $\gamma$ SREPS es el conjunto que tiene una mejor relación *dispersión-error* en general, en tanto que las líneas de RMSE (raíz del error cuadrático medio, sec. 15.3 en la página 213) y *spread* (*dispersión*, sec. 15.8.2 en la página 223) están más juntas. Si las líneas coincidieran exactamente tendríamos una relación perfecta o lo que es lo mismo, una estimación de la *amplitud de la incertidumbre* (*dispersión*), perfecta. Como decíamos, el AEMET- $\gamma$ SREPS mejora al resto de conjuntos, claramente en los parámetros humedad relativa a 2 metros (HR2m) y presión reducida al nivel del mar (PNM). Cabe señalar que en la gráfica de HR2m, los valores de dispersión-error en el tiempo inicial son incorrectos debido a que uno de los modelos locales del AEMET- $\gamma$ SREPS no diagnóstica HR2m en este instante, es decir, que no hay que tomarlos en cuenta en la comparación. En los perfiles de temperatura y módulo del viento también se ve claramente este resultado, sobre todo en niveles bajos.

Este primer resultado se va a repetir en toda la verificación. AEMET- $\gamma$ SREPS tiene una capacidad de generar dispersión de la que el resto de conjuntos carecen. Este resultado coincide con los resultados obtenidos con el antiguo AEMET-SREPS [6]. Que la mejora de la relación dispersión-error sea principalmente alrededor de la superficie se explica porque es en la superficie donde los modelos locales tienen más diferencias entre ellos, diferencias susceptibles de generar dispersión.

Otro resultado importante en la Figura 16.7 es que el lagged ensemble carece de suficiente dispersión, resultado probablemente relacionado con el hecho de que solamente contiene 8 miembros. De este resultado se podrían extraer dos conclusiones: la primera, 8 miembros para un conjunto son pocos y la segunda, el lagged ensemble a partir de un único modelo determinista con análisis en ciclos de 3 horas y 48 de predicción no es una herramienta adecuada para predicción probabilista.

La Figura 16.8 muestra un ejemplo de histogramas de rango (sec. 15.8.1 en la página 222) para las variables T2m, PNM y HR2m en el alcance de predicción H+06. Recordemos que un diagrama relativamente plano se corresponde con un comportamiento *consistente* del SPC con las observaciones, generando una dispersión adecuada para capturar las observaciones en cada caso. De otro modo, los diagramas, relativamente habituales, en forma de U corresponden a SPC subdispersivos o *poco dispersivos*. AEMET- $\gamma$ SREPS

es el que presenta el diagrama más plano (con clara diferencia), por lo que este conjunto es mucho más *consistente* respecto a las observaciones que el resto. El resto de conjuntos son claramente *poco dispersivos*, aunque LETKF genera un poco más de dispersión que los demás. Lo que sí se ve claro es la *simetría* poco deseable en el caso del conjunto NMC, debido al método para construir las perturbaciones iniciales 16.5 en la página 247.

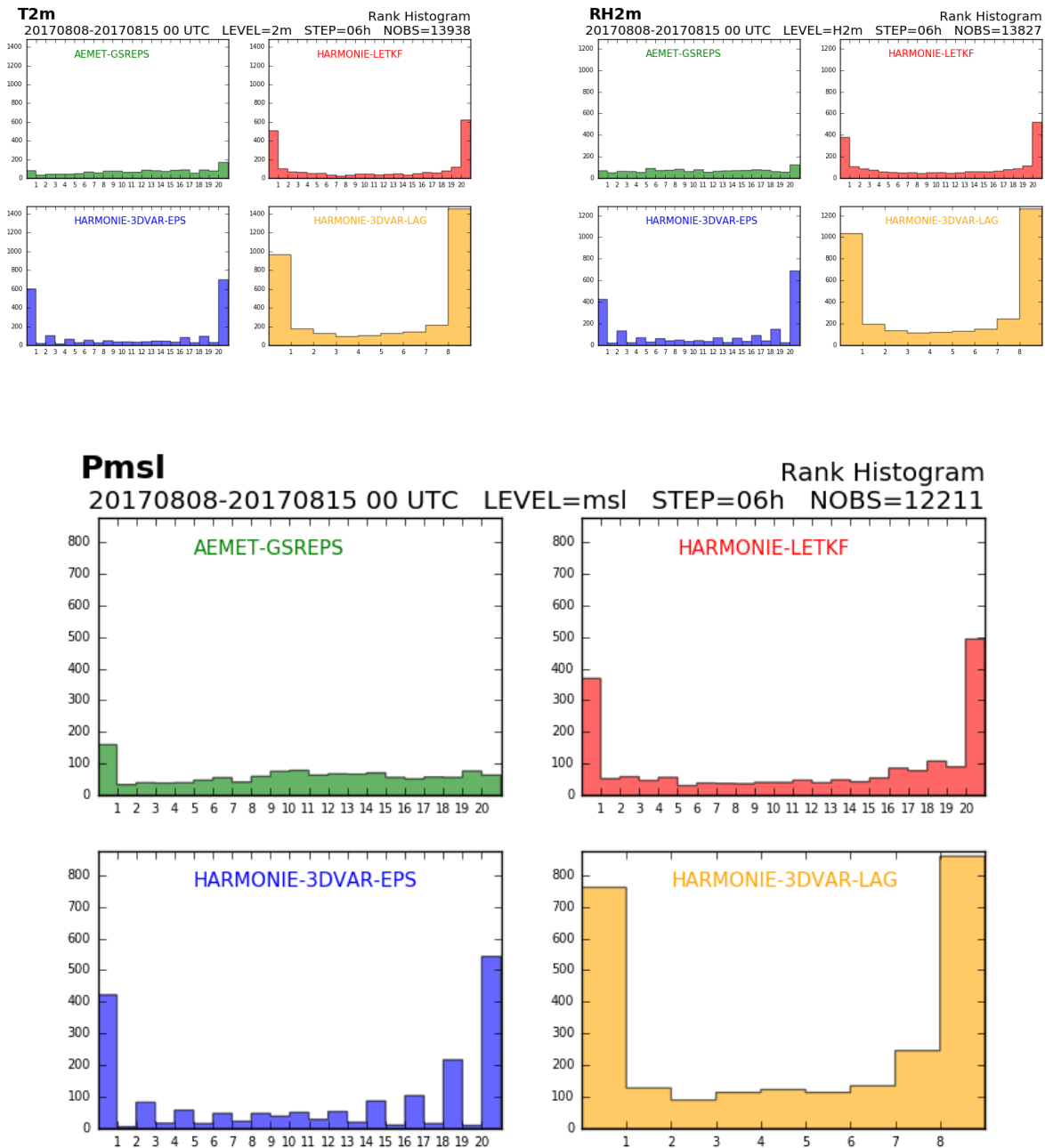


Figura 16.8: Histogramas de rango para T2m, HR2m (arriba) y PNM (abajo) en el alcance H+06 (ver texto).



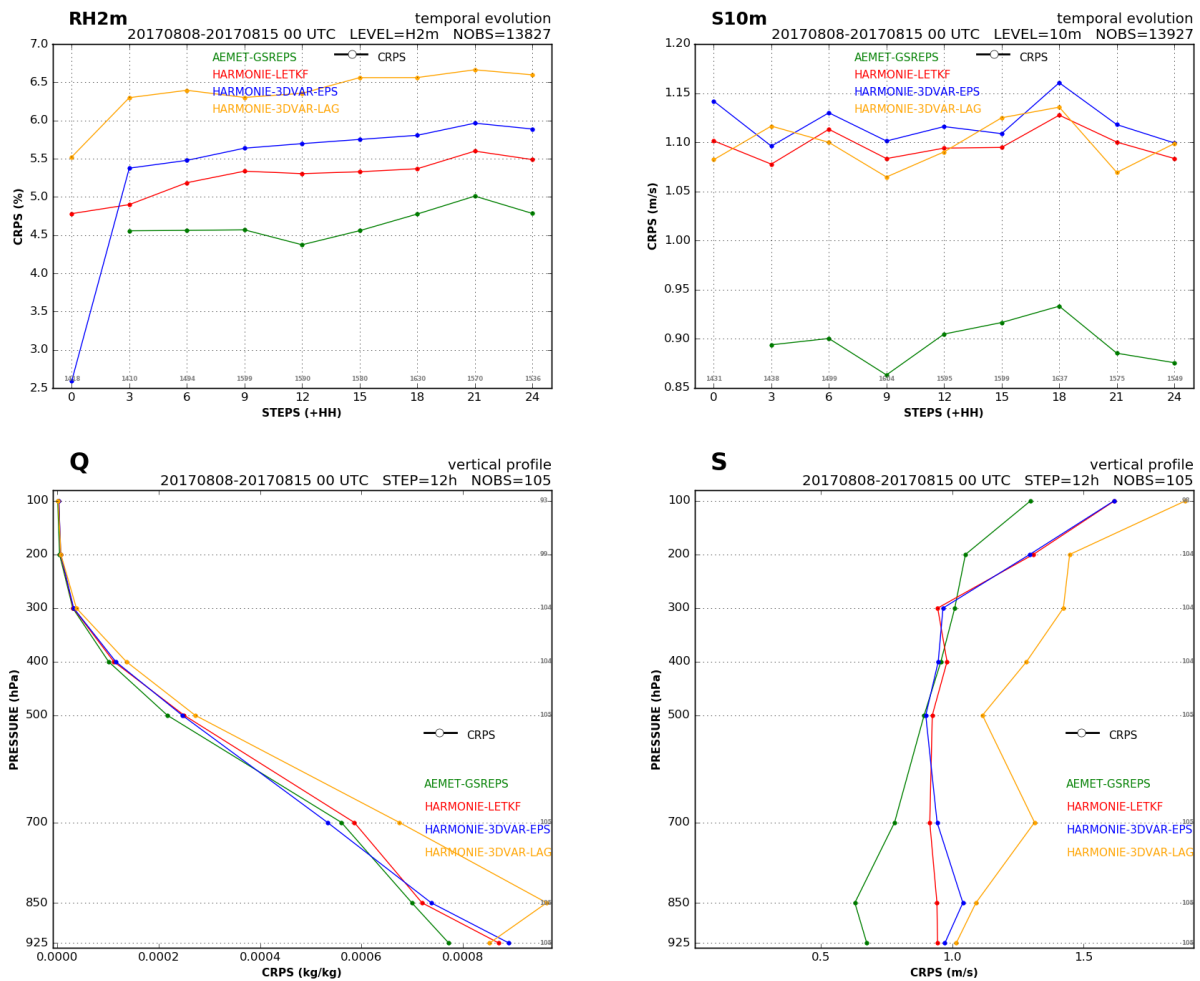


Figura 16.9: CRPS para RH2m, S10m y módulo del viento y humedad en la vertical (ver texto).

Finalmente, en la Figura 16.9 se presentan varios gráficos para el *Continuous Ranked Probability Score* (CRPS, [8]). Esta medida es el equivalente al RMSE en predicción determinista de modo que, al tratarse de un error, cuanto más pequeño, mejor el resultado. Estos gráficos confirman la calidad del AEMET- $\gamma$ SREPS por lo que respecta a la *respuesta frente a eventos binarios* en la predicción probabilista. Otra vez, la diferencia con los demás conjuntos es muy grande en cualquier parámetro en torno a la superficie. Por otro lado, si nos fijamos en los parámetros RH2m y S10m, podemos concluir que LETKF obtiene mejores resultados que 3DVAR-EPS basado en NMC. Este resultado indica que, de las técnicas disponibles a día de hoy en AEMET, usar la técnica LETKF dentro del propio AEMET- $\gamma$ SREPS para la asimilación de datos es a priori lo mas conveniente (cap. 22 en la página 333).

## 16.4 Resumen y conclusiones

Este capítulo trata uno de los fundamentos en la construcción de un sistema de predicción por conjuntos, la generación de estados iniciales. En el primer apartado se hace una introducción al concepto de *análisis* atmosférico y de los componentes que requieren su construcción, que son la predicción previa y las observaciones. Así mismo se introduce el *error del análisis*, que constituye la base conceptual para comprender la incertidumbre del estado inicial atmosférico y la necesidad de generar distintos estados iniciales para tenerla en cuenta en la predicción.

Seguidamente, se describen algunas de las principales técnicas utilizadas en centros de predicción operativa (principalmente AEMET y ECMWF), para la generación de estados iniciales en un sistema de predicción por conjuntos. En particular, se explica el multimodelo, el LETKF, el método NMC basado en 3DVAR, el lagged ensemble, el ensemble data assimilation (EDA) y finalmente, los vectores singulares. Para cada una de

estas técnicas se presentan ventajas e inconvenientes, tanto desde el punto de vista teórico como en su aplicación práctica en un entorno operativo. Se señala que, de estos métodos, AEMET ha optado históricamente por aplicar el multimodelo.

A continuación se presenta un ejercicio de verificación objetiva de 4 de las técnicas presentadas arriba, con los modelos de alta resolución (2.5 km) disponibles en AEMET. Este ejercicio consiste en integrar 2 veces al día durante la segunda semana de agosto de 2017 cuatro sistemas de predicción por conjuntos y comparar la bondad de sus predicciones probabilistas. Como métodos de verificación se utilizan la relación dispersión-error, el histograma de rango y el CRPS. De la interpretación de los resultados se deduce que el AEMET- $\gamma$ SREPS, basado en la técnica

multimodelo, es el que da, con diferencia, los mejores resultados en todos los métodos de verificación. De este resultado se puede inferir que un conjunto basado en un único modelo local parece ser menos capaz de estimar posibles estados atmosféricos, cosa que limita claramente su potencial como sistema de predicción probabilista. Una segunda conclusión es que LETKF funciona algo mejor que el método NMC basado en 3DVAR, principalmente en la generación de dispersión de buena calidad y en la predicción de algunas variables de superficie como HR2m y T2m. Este resultado apoya los planes de introducir LETKF en el AEMET- $\gamma$ SREPS para los ciclos de análisis. Finalmente, los resultados muestran claramente cómo un conjunto de pocos miembros como es el lagged ensemble en nuestro caso, no da resultados de suficiente calidad en predicción probabilista.

## 16.5 Referencias

- [1] BONAVIDA, Massimo y col. “EnKF and Hybrid Gain Ensemble Data Assimilation. Part II: EnKF and Hybrid Gain Results”. En: *Monthly Weather Review* 143.12 (dic. de 2015), páginas 4865-4882. ISSN: 0027-0644. DOI: [10.1175/MWR-D-15-0071.1](https://doi.org/10.1175/MWR-D-15-0071.1) (citado en página 246).
- [2] BUIZZA, Roberto y PALMER, Tim N. “The Singular-Vector Structure of the Atmospheric Global Circulation”. En: *Journal of the Atmospheric Sciences* 52.9 (mayo de 1995), páginas 1434-1456. ISSN: 0022-4928. DOI: [10.1175/1520-0469\(1995\)052<1434:TSVSOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1995)052<1434:TSVSOT>2.0.CO;2) (citado en página 249).
- [3] EBERT, Elizabeth E. y EBERT, Elizabeth E. “Ability of a Poor Man’s Ensemble to Predict the Probability and Distribution of Precipitation”. En: *Monthly Weather Review* 129.10 (oct. de 2001), páginas 2461-2480. ISSN: 0027-0644. DOI: [10.1175/1520-0493\(2001\)129<2461:AOPMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2461:AOPMS>2.0.CO;2) (citado en página 244).
- [4] EBISUZAKI, W y KALNAY, Eugenia. “Ensemble experiments with a new lagged average forecasting scheme”. En: *WMO Research Activities in Atmospheric and Oceanic Modeling Rep* 15 (1991), página 308 (citado en página 248).
- [5] FISHER, M. “Assimilation Algorithms”. 2002. URL: <https://www.ecmwf.int/sites/default/files/elibrary/2002/16932-assimilation-techniques-3-3dvar.pdf> (citado en página 247).
- [6] GARCÍA-MOYA, José Antonio y col. “Predictability of short-range forecasting: A multimodel approach”. En: *Tellus, Series A: Dynamic Meteorology and Oceanography* 63.3 (mayo de 2011), páginas 550-563. ISSN: 02806495. DOI: [10.1111/j.1600-0870.2010.00506.x](https://doi.org/10.1111/j.1600-0870.2010.00506.x) (citado en páginas 245, 246, 252).
- [7] HAMRUD, Mats y col. “EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation”. En: *Monthly Weather Review* 143.12 (dic. de 2015), páginas 4847-4864. ISSN: 0027-0644. DOI: [10.1175/MWR-D-14-00333.1](https://doi.org/10.1175/MWR-D-14-00333.1) (citado en página 246).
- [8] HERBACH, Hans. “Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems”. En: *Weather and Forecasting* 15.5 (oct. de 2000), páginas 559-570. ISSN: 0882-8156. DOI: [10.1175/1520-0434\(2000\)015<0559:DOTCRP>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2) (citado en páginas 251, 254).
- [9] HOFFMAN, Ross N y KALNAY, Eugenia. “Lagged average forecasting, an alternative to Monte Carlo forecasting”. En: *Tellus A* 35.2 (1983), páginas 100-118. DOI: [10.3402/tellusa.v35i2.11425](https://doi.org/10.3402/tellusa.v35i2.11425) (citado en página 248).
- [10] HUNT, Brian R., KOSTELICH, Eric J. y SZUNYOGH, Istvan. “Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter”. En: *Physica D: Nonlinear Phenomena* 230.1-2 (jun. de 2007), páginas 112-126. ISSN: 0167-2789. DOI: [10.1016/J.PHYSD.2006.11.008](https://doi.org/10.1016/J.PHYSD.2006.11.008) (citado en página 246).
- [11] ISAKSEN, LARS y col. “The new ensemble of data assimilations”. En: *ECMWF Newsletter* 123 (2010), páginas 17-21 (citado en página 248).
- [12] ISAKSEN, Lars y col. *Ensemble of data assimilations at ECMWF*. European Centre for Medium-Range Weather Forecasts, 2010 (citado en página 248).
- [13] LU, Chungu y col. “Short-Range Numerical Weather Prediction Using Time-Lagged Ensembles”. En: *Weather and Forecasting* 22.3 (jun. de 2007), páginas 580-595. ISSN: 0882-8156. DOI: [10.1175/WAF999.1](https://doi.org/10.1175/WAF999.1) (citado en página 248).
- [14] PALMER, T N y col. “Stochastic parametrization and model uncertainty”. En: *ECMWF Tech. Memo* 598 (2009), páginas 1-42 (citado en página 249).



- [15] PENDERGRASS, Angeline G y ELMORE, Kimberly L. “Ensemble Forecast Bias Correction”. En: *4th Annual Student Conference. Amer. Meteor. Soc., San Diego, CA P.* 2004, páginas 1-19 (citado en página 248).
- [16] QUIBY, J y DENHARD, M. “SRNWP-DWD poor-man ensemble prediction system: the PEPS project”. En: *Eumetnet Newsletter* 8 (2003), páginas 9-12 (citado en página 244).
- [17] SHUTTS, G y col. “Representing model uncertainty: Stochastic parametrizations at ECMWF”. En: *ECMWF Newsletter* 129 (2011), páginas 19-24 (citado en página 249).

