

# **Diseño y ajuste de un modelo de pronóstico de convección profunda para la Provincia de Mendoza**

Joás Santiago Grossi

Presentado ante la Facultad de Ciencias Exactas y Naturales como parte de los requerimientos para la obtención del grado de Licenciado en Ciencias Básicas con Orientación en Física de la Universidad Nacional de Cuyo.

Noviembre 2014

Director de Tesis: Dr. Diego Araneo

## Resumen

Mediante el Análisis de Componentes Principales, en el presente trabajo se determinan los principales patrones que caracterizan los perfiles verticales de *Temperatura* y *Temperatura del punto de rocío* para los niveles estándar de 925 a 300 hPa en la provincia de Mendoza. Las series temporales de peso asociadas a las componentes se correlacionan por regresión múltiple logística con las observaciones de presencia de nubes convectivas, a fin de ajustar un modelo estadístico de pronóstico de inestabilidad a corto plazo, mediante el cual sea posible calcular un índice de previsión de ocurrencia de convección a partir de los datos de radio-sondeo para los niveles estándar, válido para las 24 horas siguientes a la medición.

A partir de los análisis efectuados se ha logrado ajustar un modelo estadístico capaz de pronosticar la ocurrencia de convección con alrededor de un 80% de efectividad, con errores en torno al 8 y 11% por imprevisto y falsa alarma respectivamente.

Se observa que las variables dinámicas y energéticas como la helicidad, la cortante vertical del viento en distintos niveles y el CAPE no son factores determinantes en las previsiones del inicio de la convección para la región del oasis norte de la provincia de Mendoza. Además se observa que ni la implementación del método de rotación VARIMAX ni la variación en el rango temporal de validez del pronóstico (definición del rango temporal de ocurrencia de convección) mejoran la eficiencia del modelo estadístico.

Adicionalmente se combina este modelo estadístico con el modelo numérico de pronóstico Eta-PRM para evaluar su eficiencia en la predicción de ocurrencia de tormentas severas. A pesar de que la utilización de sondeos simulados como entrada en el modelo estadístico implica una disminución en la eficiencia del pronóstico resultante, esta efectividad se mantiene en alrededor del 70 a 73% para cuatro días posteriores al día de la corrida del modelo numérico.

## Índice

1. Introducción.....	4
2. Hipótesis y Objetivos.....	5
3. Datos y Metodología.....	9
4. Resultados.....	13
4.1. Análisis de Componentes Principales para los perfiles verticales de T y Td.....	13
4.2. Ajuste del modelo de pronóstico a partir de los patrones de perfiles verticales de T y Td.....	18
4.3. Ajuste del modelo de pronóstico a partir de los patrones rotados y sin rotar de perfiles de T y Td, la cortante del viento, la helicidad y el CAPE.....	26
4.4. Combinación del modelo 12 UTC con las salidas del modelo numérico Eta-PRM.....	28
5. Conclusiones.....	34
6. Apéndices.....	36
7. Bibliografía.....	43

# 1. Introducción

La Provincia de Mendoza se encuentra en el centro-oeste de la República Argentina, situada al pie de la Cordillera de los Andes, extendiéndose de  $32^{\circ}$  a  $37^{\circ} 35'$  S, y de  $66^{\circ} 30'$  a  $70^{\circ} 35'$  O. Su superficie es de  $148.827 \text{ km}^2$  y cuenta con una población estimada de 1.700.000 habitantes. En ella se distinguen tres regiones morfológicas naturales bien contrastadas: las montañas andinas al oeste; las planicies en el centro y el este, conformadas por los pedemontes y llanuras; y las mesetas y volcanes de la Payunia en el sur (Informe Ambiental, 2009; de aquí en más IA). Mendoza se presenta como una provincia mediterránea y continental, de clima árido a semiárido, con temperaturas medias de  $9^{\circ}\text{C}$  en invierno y  $25^{\circ}\text{C}$  en verano, y una amplitud térmica media de  $13,5^{\circ}\text{C}$  para ambas estaciones. La precipitación anual total es de alrededor de 220 mm, de la cual casi el 50% se acumula en los meses de verano; mayormente producida por tormentas convectivas, con frecuente caída de granizo (Servicio Meteorológico Nacional, 2009). La combinación de rasgos geomorfológicos, climáticos, edáficos y bióticos diferencia unidades homogéneas con distintas posibilidades desde el punto de vista de los asentamientos humanos y las actividades productivas. En este marco, el aprovechamiento del agua de los ríos y del agua subterránea, ha generado importantes oasis de riego que concentran más del 90% de la población y representan sólo el 3% de la superficie mendocina, con 273.000 ha cultivadas distribuidas en tres zonas: Norte, Centro-Oeste, y Sur. Del total de áreas cultivadas, el 60% corresponde al oasis Norte. La existencia de fenómenos tales como las sequías, el granizo, las heladas y el viento Zonda, así como las frecuentes tormentas de nieve en la montaña, constituyen fuertes condicionantes climáticos para el desarrollo productivo, obstaculizando los asentamientos humanos en la Provincia (Martínez Carretero y Dalmasso, 1995).

Existen investigaciones en las cuales no sólo se estudió el origen y desarrollo de la actividad convectiva severa en Argentina, sino también su relación con el riesgo que implica y los daños que ocasiona (Nicolini y Brizuela, 1997). Particularmente en el Oasis Norte, el porcentaje anual de daños por granizo supera el 30%; mientras que en ciertos lugares de la Zona Sur llega hasta el 40% (IA, 2009). La circulación local en la zona norte, por su ubicación a sotavento de la Cordillera de los Andes, está condicionada tanto por el forzante mecánico representado por la orografía, como por el calentamiento diferencial que genera brisas valle-montaña. El período de mayor frecuencia de ocurrencia de tormentas se encuentra entre los meses de octubre y marzo, pudiendo registrarse algunos eventos en setiembre y abril (Saluzzi et al, 1979). La convección es generalmente vespertina, presentando el máximo en las primeras horas de la noche y un segundo período de actividad en horas de la madrugada, observándose excepcionalmente ocurrencia de convección en horas de la mañana (Saluzzi, 1983). Asimismo, los efectos locales relacionados con la orografía, pueden influir en la ubicación espacial y en la iniciación de la convección en la región. La intensidad y la forma en que se desarrolla y organiza la convección, están directamente relacionados con la estructura termodinámica del entorno y la cortante vertical del viento.

El mejoramiento de las técnicas de pronóstico de convección profunda, con predicciones espacial y temporalmente más precisas, ayudaría a desarrollar sistemas de alerta temprana de eventos severos, que permitan tomar medidas tendientes a mitigar eventuales daños socio-económicos. El conocimiento de la convección profunda requiere, en su fase inicial, integrar los efectos atmosféricos de escalas mayores a la convectiva capaces de forzar el ascenso y la convergencia en capas bajas, o desestabilizar la atmósfera dando lugar a movimientos de ascenso en niveles medios.

Nicolini y Norte (1978) y Norte (1980) encontraron que la ocurrencia de convección en la zona norte de Mendoza se asocia a situaciones sinópticas frontales, a líneas de inestabilidad, o bien puede desarrollarse en una masa de aire localmente inestable. La elaboración operativa de pronósticos de convección requiere de los análisis sinópticos de superficie y altura (complementados con modelos

termodinámicos regionales), y principalmente del análisis termodinámico de la estructura vertical de la atmósfera (es decir, a partir de datos de radio-sondeo), para lo cual se calculan habitualmente distintos índices indicadores de la inestabilidad estática del entorno. A pesar del gran número de índices utilizados, sólo existen algunos estudios sistemáticos en Argentina que permiten determinar el desempeño comparativo de los mismos como predictores de ocurrencia de convección y precipitaciones (Moyano et al, 1972; Rosso et al, 1987). Norte (1982) analizó el desempeño de los índices SWEAT modificado, Total-Total y SPOT en la zona norte de Mendoza, encontrando que el primero da una mejor idea de la intensidad de la actividad convectiva en la región. Posteriormente, Simonelli (2000) realizó un estudio similar sobre la misma zona evaluando el desempeño de los índices K de Whitting, Showalter, Lifted, Total-Total, SWEAT, Fawbush-Miller, Cross-Total y Vertical-Total, mediante un análisis discriminante escalonado. Los resultados finales de su trabajo mostraron que el Fawbush-Miller es el único índice con un desempeño aceptable como predictor de la convección para un período de 24 horas posteriores a la hora de la observación, con una efectividad (expresada como el porcentaje de aciertos) de un 76%; sin embargo, la dificultad en el cálculo de este índice hace que el mismo no sea tenido en cuenta en forma operativa.

Las características espacio-temporales de los perfiles de temperatura y humedad en la troposfera y baja estratosfera fueron analizadas mediante el estudio de Funciones Ortogonales Empíricas (EOF) primeramente por Smith y Woolf (1976), recientemente por Wang et al (2001) y Deming et al (2008); y en Argentina por Duarte (1993). Mediante estas técnicas los autores pudieron identificar los principales patrones temporales de temperatura en cada nivel de la atmósfera, caracterizando perfiles verticales con variabilidad temporal homogénea. La aplicación de técnicas estadísticas como el estudio de EOF o el Análisis de Componentes Principales (ACP) permiten en general la caracterización estadística objetiva de variables vectoriales, tales como los perfiles verticales de temperatura de los ejemplos citados o cualquier otra variable física; y además haciendo un ajuste de regresión múltiple logística entre la serie de componentes de carga producidas por el ACP y un indicador de ocurrencia de convección convenientemente definido; es posible ajustar un modelo estadístico que permita calcular un índice de inestabilidad de forma mucho más simple y con un grado de efectividad semejante al de Fawbush & Miller. Los ajustes realizados hasta el momento permitieron definir un índice con un nivel de efectividad del 73% según Araneo et al (2011).

## 2. Hipótesis y Objetivos

Dentro de la atmósfera, entre una parcela de aire<sup>1</sup> y su entorno se producen tres tipos de estados de equilibrio: inestable, estable e indiferente. El equilibrio inestable se debe a que la parcela de aire es más ligera, menos densa o más cálida que el entorno, de modo que cualquier pequeño impulso, por débil que sea, la hará elevarse hasta un nivel donde adquiera una densidad igual o mayor a la del aire del entorno. El equilibrio estable se debe a que la parcela está más fría que su entorno, es decir más densa que el aire que la rodea, de modo que la parcela de aire retornará a su estado original frente a cualquier impulso que tienda a elevarla. Y el equilibrio indiferente se da cuando la parcela al ser desplazada adquiere una densidad igual a la de su entorno con lo cual permanecerá en el lugar al cual se desplazó.

Debido a lo mencionado en el párrafo anterior se tiene que la intensidad de las corrientes verticales

---

<sup>1</sup> Volumen de aire lo suficientemente pequeño como para que la temperatura, la humedad y la presión puedan considerarse uniformes en toda su extensión.

que se producen, hacia arriba o hacia abajo, dependen principalmente del tipo de equilibrio. A su vez el tipo de equilibrio está determinado fundamentalmente por dos factores, la variación de la temperatura con la altura (gradiente vertical de temperatura) del aire que rodea a la parcela que asciende o desciende y la variación de temperatura que experimenta internamente por sí misma la parcela de aire que se mueve verticalmente. Por lo tanto, los fenómenos convectivos están estrechamente ligados a la variación de los gradientes de temperatura y humedad lo cual motiva el análisis de los perfiles de temperatura ( $T$ ) y temperatura del punto de rocío ( $Td$ ) a través del ACP.

Los ajustes efectuados por Araneo et al (2011) hasta el momento comprenden sólo el análisis de 326 perfiles verticales de temperatura ( $T$ ) y temperatura de rocío ( $Td$ ) en cinco niveles de la tropósfera, desde los 850 hasta los 300 hPa, y que definen como se mencionó anteriormente un índice con un nivel de efectividad del 73%. Sin embargo, la incorporación de un mayor número de datos, el aumento en la precisión de las observaciones mediante el incremento del número de capas consideradas, la incorporación de predictores dependientes de variables dinámicas y termodinámicas no consideradas en los trabajos previos, la redefinición de los criterios a partir de los cuales se discriminan eventos convectivos y no convectivos, entre otras modificaciones, podría implicar mejoras en los ajustes con el consecuente aumento en la efectividad del modelo obtenido.

En efecto, es posible que la efectividad del índice mejore si la base de datos analizada incluye valores correspondientes a niveles de capas inferiores de la tropósfera, donde se encuentran los mayores contenidos de humedad y consecuentemente gran parte de la energía (calor latente) necesaria para el inicio de la convección. Adicionalmente, el aumento en la cantidad de niveles de observación constituye un aumento en la precisión con la cual se observa y representa el estado de la atmósfera en un determinado momento, lo cual implicaría una mejora en la precisión de los ajustes con un consecuente aumento en la efectividad del modelo.

Por otra parte, el modelo definido por Araneo et al (2011) fue obtenido en base al ajuste de las observaciones de radiosondeo de las 12UTC (9 AM local) con la convección observada en las 24 horas siguientes (indicador de evento Convección – No convección, C-NC). Sin embargo, parte de los errores por falsa alarma o imprevisto (sorpresa) podrían deberse a la forma de definición del indicador de C-NC debido a que, por ejemplo, el índice podría estar indicando un ambiente inestable, con actividad convectiva producida en horas previas a la del radiosondeo, que luego se estabiliza con la desaparición de la convección en las siguientes 24 horas (falsa alarma); o bien podría indicar un ambiente estable pero que luego se inestabiliza produciendo actividad convectiva en horas de la madrugada siguiente, lo cual no es advertido a la hora del radiosondeo (sorpresa). La redefinición del indicador de C-NC, por ejemplo cambiando el rango horario de observación de eventos, podría ayudar a disminuir estos errores y aumentar la precisión del modelo definido.

Adicionalmente, la efectividad del índice podría mejorar realizando una rotación VARIMAX a las componentes principales que surgen del ACP. Esto se fundamenta en el hecho de que la rotación podría permitir una interpretación más “real” los factores de carga (i.e. más fácilmente identificables con las variables originales), al indicar una asociación positiva o negativa más clara entre la variable (en nuestro caso el sondeo estandarizado) y la componente principal rotada asociada al nuevo factor de carga (o una ausencia de asociación si el valor es cercano a 0). Por lo tanto, es probable que la eficiencia del modelo mejore si la rotación permite definir patrones que discriminan mejor los casos de estabilidad e inestabilidad atmosférica, y consecuentemente una mejor asociación (mayores correlaciones) entre predictores (componentes rotadas) y la variable respuesta (indicador de C-NC).

Otro aspecto a tener en cuenta para el desarrollo de la convección profunda es la disponibilidad de

energía termodinámica en el entorno. En efecto, la efectividad del índice podría mejorar si la Energía Potencial Convectiva Disponible (CAPE, por sus siglas en inglés) es incluida entre los predictores del modelo. La Energía Potencial Convectiva Disponible se define como:

$$\text{CAPE} = \int_{\text{LFC}}^{\text{EL}} B \, dz \approx g \int_{\text{LFC}}^{\text{EL}} \frac{T'_v}{\bar{T}_v} \, dz \quad (2.1)$$

(ver Markowski y Richardson, 2010) donde  $z$  representa la altura vertical y varía entre el nivel de libre convección (LFC en inglés) y el nivel de equilibrio (EL en inglés); y la flotabilidad o fuerza de empuje  $B$  se expresa en términos de la perturbación de temperatura virtual de la parcela elevada ( $T'_v$ ) relativa a la temperatura virtual del ambiente ( $\bar{T}_v$ ), siendo  $T_v = \bar{T}_v + T'_v$  la temperatura virtual de la parcela elevada. La temperatura virtual  $T_v$  se expresa como:

$$T_v = T \frac{1 + r_v/\epsilon}{1 + r_v} \approx T(1 + 0.61r_v) \quad (2.2)$$

la cual se puede interpretar como la temperatura que tendría una parcela de aire seco con una densidad equivalente a la de una parcela de aire húmedo a la misma presión, y donde la relación de mezcla de vapor de agua  $r_v$  se define como:

$$r_v = \frac{\rho_v}{\rho_d} = \frac{\epsilon e}{p_d} = \frac{\epsilon e}{p - e} \quad (2.3)$$

donde  $\epsilon = R_d / R_v \approx 0.622$ ,  $\rho_v$  y  $\rho_d$  expresan la densidad de vapor y aire seco, y  $p$  y  $e$  las presiones de aire seco y vapor respectivamente, tomando  $p$  como la presión total.

El CAPE es proporcional a la energía cinética que una parcela ganaría del entorno como resultado de la contribución de la flotabilidad a la aceleración vertical. En ambientes típicos de tormentas severas los valores de  $\text{CAPE} \leq 1000 \text{ J kg}^{-1}$  son usualmente considerados como pequeños, mientras que valores  $\geq 2500 \text{ J kg}^{-1}$  son considerados como valores grandes. La hipótesis de que el CAPE pueda utilizarse como predictor en el modelo estadístico se fundamenta en el hecho de que mientras mayor sea su valor hay una mayor disponibilidad de energía potencial disponible en el ambiente y por lo tanto una mayor posibilidad de desarrollo de convección. En consecuencia, es posible que la eficiencia del modelo mejore mediante la incorporación de esta variable como predictor en el ajuste.

Por otra parte, es probable que la efectividad del índice también mejore si se incluyen dentro de los predictores del modelo, variables representativas de las condiciones dinámicas del entorno, tales como la Helicidad Relativa a la Tormenta y la Cortante Vertical del Viento (SRH y VWS por sus siglas en inglés respectivamente).

La Cortante Vertical es el factor más influyente en la determinación del tipo, organización y longevidad de una tormenta (aunque en exceso puede reducir el ascenso en ambientes poco inestables ya que favorece el “entrainment”<sup>2</sup>). Un parámetro que funciona bien para cuantificar la

---

<sup>2</sup> Incorporación de aire seco del entorno hacia el interior de la nube convectiva, que favorece la evaporación con un consecuente debilitamiento de la convección.

cortante y predecir el tipo de tormenta es la diferencia vectorial de viento en 0-6 km. La cortante vertical tiende a favorecer la organización, longevidad y severidad de la convección debido a 2 razones:

1. *El grado con el cual la precipitación y el flujo saliente interfieren con la ascendente se reduce a medida que la cortante se incrementa en las capas altas.* Hay una tendencia a que la precipitación caiga más lejos de la ascendente cuanto mayores son: a. la cortante profunda y b. los vientos en altura relativos a la tormenta.

2. *La cortante en capas bajas influye en el ascenso a lo largo del frente de ráfagas y por lo tanto en la habilidad del frente para disparar nueva convección.* En ambientes con *cortante débil*, el levantamiento de aire del entorno por el frente de ráfagas es débil y tiende a no iniciarse la convección. Con *cortante moderada*, el levantamiento en el frente de ráfagas es intensificado a lo largo de un flanco preferido, induciendo repetidamente la iniciación de nuevas celdas. Una *cortante fuerte*, induce tormentas severas y bien organizadas en nubes con rotación.

La Helicidad es una medida de que tan alineados están la dirección del movimiento del fluido con su vorticidad horizontal. Por lo tanto la helicidad está estrechamente relacionada con la vorticidad en la dirección del viento horizontal en una capa ("streamwise vorticity" en inglés). Las corrientes que son altamente helicoidales están dominadas por la streamwise vorticity. Matemáticamente la helicidad  $H$  es el producto punto de los vectores de velocidad y vorticidad, es decir,

$$\mathcal{H} = \mathbf{v} \cdot \boldsymbol{\omega} = \mathbf{v} \cdot \nabla \times \mathbf{v}. \quad (2.4)$$

El concepto de helicidad se ha aplicado a tormentas supercelulares debido a la relación estrecha que esta tiene con la vorticidad en la dirección del viento horizontal en una capa. En la aplicación a tormentas convectivas se suele integrar la helicidad asociada al viento del entorno ( $\bar{\mathbf{v}}$ ) y la vorticidad ( $\bar{\boldsymbol{\omega}}_h$ ) sobre cierta profundidad  $d$  que es aproximadamente la profundidad del flujo de entrada de la tormenta (1-3 km), esto es,

$$\mathcal{H} = \int_0^d \bar{\mathbf{v}} \cdot \bar{\boldsymbol{\omega}}_h \, dz \quad (2.5)$$

donde

$$\bar{\boldsymbol{\omega}}_h = \left( -\frac{\partial \bar{v}}{\partial z}, \frac{\partial \bar{u}}{\partial z} \right) = \mathbf{k} \times \mathbf{S}. \quad (2.6)$$

A partir de (5) y (6) se tiene que la helicidad relativa a la tormenta (SRH en inglés) es,

$$\begin{aligned} \text{SRH} &= \int_0^d (\bar{\mathbf{v}} - \mathbf{c}) \cdot \bar{\boldsymbol{\omega}}_h \, dz = \int_0^d |\bar{\mathbf{v}} - \mathbf{c}| \omega_s \, dz \\ &= - \int_0^d \mathbf{k} \cdot (\bar{\mathbf{v}} - \mathbf{c}) \times \mathbf{S} \, dz. \end{aligned} \quad (2.7)$$

donde  $\mathbf{c}$  es el vector medio de movimiento de la tormenta.

Sumado a la cortante vertical (0-6 km), la SRH se usa a menudo para pronosticar si es probable la formación de tormentas supercelulares. Valores de SRH  $>150 \text{ m}^2 \text{ s}^{-2}$  en 0-3 km indican que hay un



potencial que permitiría la formación de superceldas (estrictamente hablando la SRH probablemente debería ser considerado como predictor de un mesociclón, a su vez un mesociclón es la característica que define una supercelda). Si bien la cortante vertical del viento en 0-6 km y la SRH parecen ser independientes en realidad tienden a estar correlacionadas; valores de cortante vertical  $>15-20 \text{ m s}^{-1}$  en 0-6 km sugieren la formación de superceldas que a su vez corresponden a valores de SRH que favorecen los mesociclones. Al igual que el valor de CAPE, los valores de la cortante vertical del viento y la helicidad relativa a la tormenta no garantizan el inicio de la convección, pero son variables fuertemente asociadas a procesos convectivos.

A pesar del gran desarrollo que los modelos numéricos meteorológicos regionales han tenido en los últimos años debido a la mejora en la capacidad de cómputos a raíz de la disponibilidad de procesadores más potentes, la representación de la convección profunda a partir de las salidas de estos modelos aún no resulta con la precisión requerida por los pronosticadores, haciendo difícil su predicción a partir de las variables pronosticadas por los mismo. La modelación de variables como la precipitación o indicadores de inestabilidad requiere de modelos numéricos con gran resolución espacial y temporal, y de parametrizaciones de procesos y condiciones locales específicas con las que estos modelos generalmente no cuentan, debido a la mayor demanda computacional requerida para aumentar su resolución y la falta de datos climatológicos locales para ajustar las parametrizaciones. Sin embargo, algunos modelos regionales locales, como el Eta-PRM que actualmente se corre operativamente en el Centro Científico Tecnológico – CONICET de Mendoza, ha demostrado reproducir con exactitud aceptable los perfiles verticales de  $T$  y  $Td$  y las componentes horizontales de viento en altura (Simonelli et al, 2010). Por lo tanto, es de esperar que el modelo estadístico ajustado a las condiciones climatológicas locales, en combinación con las salidas del Eta-PRM de estas variables (i.e. empleando estas salidas como predictores en el modelo estadístico), constituya una herramienta que ayude a mejorar los pronósticos de convección en la provincia con al menos uno a cuatro días de antelación.

A partir de estas perspectivas, el objetivo de este trabajo es investigar las respuestas del modelo estadístico descrito por Araneo et al (2011), a la incorporación de estas modificaciones y su combinación con el modelo dinámico local Eta-PRM, a fin de obtener un índice de inestabilidad con mayor nivel de efectividad que los obtenidos hasta el momento, que sirva como herramienta sencilla en la elaboración de pronósticos de tormentas severas, principalmente en el área del oasis norte de la provincia de Mendoza para ayudar a desarrollar sistemas de alerta temprana que permitan tomar medidas tendientes a mitigar eventuales daños socio-económicos.

### **3. Datos y Metodología**

Se tomaron datos de temperatura ( $T$ ), temperatura del punto de rocío ( $Td$ ) y viento horizontal en los niveles estándar de 925, 850, 700, 500, 400 y 300 hPa a partir de 659 radio-sondeos correspondientes a las 12 UTC realizados en la Estación Meteorológica Mendoza-Aero perteneciente al Servicio Meteorológico Nacional (SMN), para los períodos octubre-marzo de los años 2005/2006 a 2012/2013.

Del total de 659 radio-sondeos se destinó una muestra de 320 correspondientes a los periodos 2005/2006 a 2008/2009 para en el cálculo del ACP y el ajuste del modelo siguiendo los procedimientos descritos en Araneo et al (2011), reservando los restantes 339 radio-sondeos correspondientes a las temporadas 2009/2010 a 2012/2013 a fin de testear la efectividad del modelo

ajustado en el pronóstico de eventos de convección profunda para estos últimos años.

Por otra parte se consideraron datos de provenientes de los reportes horarios del código METAR de la estación Mendoza-Aero a fin de determinar los días con observación de fenómenos convectivos. Se definió “Ocurrencia de Convección para las 12 UTC” cuando en la observación meteorológica se reporta al menos en alguna hora la presencia de nubes bajas de los grupos 3 y/o 9, de acuerdo con el código de clasificación de nubes de la Organización Meteorológica Mundial, en el período de 24 horas comprendido entre las 12 UTC del día y las 12 UTC del día siguiente. Además se definió “Ocurrencia de Convección para las 9 UTC” cuando se reportan estos mismos fenómenos en el período de 24 horas comprendido entre las 9 UTC del día y las 9 UTC del día siguiente. Recíprocamente, la “No Ocurrencia” de convección se definió cuando no se reportan nubes bajas de los grupos 3 y/o 9 en ninguna de las 24 horas consideradas. Cabe destacar que, debido a ausencia de datos en algunas horas, los días con faltantes para ciertas horas y sin reporte de nubes convectivas en las restantes no pudieron considerarse como “no convectivos”, debiendo ser descartados. Por tal motivo, la muestra de días considerados no convectivos es considerablemente menor a la de días convectivos.

Es importante mencionar que se tiene información técnica real de la ocurrencia o no convección para muchos días pero no para la totalidad de los 659 días con datos de radiosondeo. Por tal motivo, en los siguientes análisis que se presentan en este trabajo se especifica expresamente en cada caso la cantidad de sondeos empleados.

### ***Método ACP***

Para el cálculo de las Componentes Principales (CPs) disponemos los datos de radiosondeo formando un arreglo matricial  $\mathbf{X}_{[12 \times 320]}$  en el cual cada columna representa el sondeo de un día determinado, las primeras seis filas contienen los valores correspondientes a  $T$  (para cada nivel de presión) y las restantes los valores de  $Td$  (también para cada uno de los seis niveles). Dado que el sondeo constituye una variable en la cual el perfil vertical medio climatológico (es decir, el sondeo medio) explica casi la totalidad de la varianza, se procedió a la remoción del mismo, obteniendo la

matriz de desvíos  $\tilde{\mathbf{X}}$ , con coeficientes  $\tilde{x}_{ij} = x_{ij} - \bar{x}_i$ , donde  $\bar{x}_i = \sum_{j=1}^{320} x_{ij}$ . Posteriormente, esta matriz

de desvíos es estandarizada por columnas obteniendo la matriz  $\tilde{\mathbf{X}}_s$  de coeficientes  $\tilde{x}_{s\ ij} = \frac{\tilde{x}_{ij} - \bar{\tilde{x}}_j}{s_j}$ ,

donde  $\bar{\tilde{x}}_j = \sum_{i=1}^{12} \tilde{x}_{ij}$  y  $s_j = \sqrt{\frac{1}{11} \sum_{i=1}^{12} (\tilde{x}_{ij} - \bar{\tilde{x}}_j)^2}$ .

Con la matriz de desvíos estandarizada se efectúan los cálculos para la obtención de las CPs de acuerdo a las ecuaciones matriciales:

$$\mathbf{Z} = \tilde{\mathbf{X}}_s \mathbf{Q} \quad (3.1)$$

$$\mathbf{F} = \mathbf{Q} \mathbf{D}^{1/2} \quad (3.2)$$

donde  $\mathbf{Z}_{[12 \times 320]}$  es la “matriz de componentes principales” (o componentes de puntaje),  $\mathbf{Q}_{[320 \times 320]}$  y  $\mathbf{D}_{[320 \times 320]}$  son respectivamente las matrices de autovectores y autovalores de la matriz de

correlaciones entre las columnas de  $\tilde{\mathbf{X}}$ ; y  $\mathbf{F}_{[320 \times 320]}$  es la “matriz de componentes de carga” (o loadings, es decir, la matriz que contiene las correlaciones entre cada columna de  $\tilde{\mathbf{X}}$  y cada columnas de  $\mathbf{Z}$ ).

Además resulta que  $\tilde{\mathbf{X}}_s = \mathbf{Z}_s \mathbf{F}'$  (donde la prima significa “transpuesta”) y por lo tanto cada sondeo estandarizado (cada columna de  $\tilde{\mathbf{X}}_s$ ) puede escribirse como combinación lineal de las componentes principales estandarizadas (columnas de  $\mathbf{Z}_s$  estandarizadas) en la forma:

$$\tilde{\mathbf{x}}_{s,j} = f_{j1} \mathbf{z}_{s1} + f_{j2} \mathbf{z}_{s2} + \dots + f_{j10} \mathbf{z}_{s12} \quad (3.3)$$

donde el coeficiente  $f_{ji}$  es el elemento de la  $j$ -ésima columna y la  $i$ -ésima fila de  $\mathbf{F}$ , y teniendo en cuenta además, que el rango de  $\mathbf{X}$  no puede ser mayor a 12. Información adicional sobre esta metodología y sus propiedades puede consultarse en Lattin y otros (2003).

La ecuación (3.3) permite identificar cada sondeo estandarizado con un número no mayor a 12 factores de carga  $f_{ji}$ , los cuales constituyen el peso de cada componente para representar el sondeo en cuestión.

En el apartado **Resultados 4.1** y en el apéndice A se muestran las componentes principales que resultan de aplicar el método del ACP a los 320 sondeos y también para los 659 sondeos con la intención de mostrar que tanto la muestra que se ha tomado de 320 sondeos como la población de 659 sondeos tienen las mismas componentes principales.

### **Regresión Logística**

Una vez identificada la ocurrencia o no de convección para cada período de 24 horas (a partir de las 12UTC y a partir de las 9UTC, se procede para ambos casos a asignar a cada día correspondiente el valor “1” en caso de registrarse convección (en al menos una de las 24 horas, en adelante caso C) y “0” en caso de no registrarse (en ninguna de las 24 horas, en adelante caso NC), a fin de definir un vector de respuesta  $\mathbf{c}_{[row \times 1]}$  (donde *row* es una variable que especifica la cantidad de sondeos que se toman, del total de 320 que se utilizaron para el ajuste, para los cuales se tiene información técnica real de la ocurrencia o no convección) que contenga la información de los días C y NC (i.e. con valores 1 para casos C y 0 para casos NC).

A continuación, se procede al ajuste de un modelo de regresión múltiple logística entre el vector respuesta  $\mathbf{c}$  y las componentes de carga  $\mathbf{f}_j$ , de la forma:

$$\hat{c}_i = P(c_i = 1) = \frac{e^{w_i}}{1 + e^{w_i}} \quad (3.4)$$

donde  $\hat{c}_i$  es un estimador del coeficiente de  $\mathbf{c}$  correspondiente al  $i$ -ésimo día y representa la probabilidad de un evento convectivo, y  $w_i$  es el  $i$ -ésimo elemento del vector:

$$\mathbf{w} = b_0 + b_1 \mathbf{f}_1 + b_2 \mathbf{f}_2 + \dots + b_n \mathbf{f}_n \quad (3.5)$$

donde los coeficientes  $b_0, b_1, \dots, b_n$  se ajustan por máxima verosimilitud y  $n$  es el número de componentes significativas retenidas.

Dado que  $\mathbf{F} = \frac{\tilde{\mathbf{X}}_s' \mathbf{Z}_s}{m-1}$ , donde  $m$  es el número de filas de  $\mathbf{X}$  (en nuestro caso  $m=12$ ), entonces resulta:

$$\mathbf{w} = \tilde{\mathbf{X}}_s' \frac{\mathbf{Z}_s^* \mathbf{b}^*}{m-1} + b_0 [\mathbf{1}]_{row} \quad (3.6)$$

donde  $\mathbf{Z}_s^*$  es la matriz que contiene las componentes de puntaje estandarizadas que corresponden a los predictores usados en la ecuación de regresión (3.4)-(3.5) (es decir, todas o aquellas componentes con coeficientes  $b_i$  significativos),  $\mathbf{b}^*$  es la matriz vector que contiene los coeficientes de ajuste y  $[\mathbf{1}]_{row}$  es un vector columna de  $row$  elementos iguales a 1.

La matriz  $\mathbf{A} = \frac{\mathbf{Z}_s^* \mathbf{b}^*}{m-1}$  es un vector columna que sólo depende de los resultados del ACP y del análisis de regresión (3.4)-(3.5). Debido a esto, una vez determinados  $\mathbf{A}$  y  $b_0$  del análisis estadístico descrito, dado cualquier sondeo anómalo estandarizado  $\tilde{\mathbf{x}}_s$  (no necesariamente perteneciente al presente análisis), el índice de C (1) o NC (0) puede estimarse mediante la ecuación:

$$\hat{c} = \frac{e^{\tilde{\mathbf{x}}_s \cdot \mathbf{A} + b_0}}{1 + e^{\tilde{\mathbf{x}}_s \cdot \mathbf{A} + b_0}} \quad (3.7)$$

lo cual serviría para futuros pronósticos.

Si el índice definido por la ecuación (3.7) está asociado al vector  $c$  para la Ocurrencia de Convección a partir de las 12 UTC entonces en adelante nos referiremos a él como “índice 12 UTC” y recíprocamente, si está asociado al vector  $c$  para la Ocurrencia de Convección a partir de las 9 UTC entonces lo denominaremos “índice 9 UTC”.

A fin de comprobar la efectividad del índice 12 UTC y del índice 9 UTC para predecir eventos de convección profunda, se aplicó la ecuación (3.7) correspondientemente para observaciones tomadas de los períodos octubre-marzo de 2009/2010 a 2012/2013. En el apartado **Resultados 4.2** se muestran los valores de efectividad para estos índices.

Además se procedió a comprobar si la efectividad del índice mejora al aplicar el método de rotación VARIMAX y al incorporar al ajuste de regresión dado por la ecuación (3.5), otros predictores como las variables dinámico-energéticas SRH, VWS y el CAPE mencionadas. Los valores de la helicidad y la cortante se obtienen a partir de los valores de velocidad y dirección del viento incluidos en los datos radio-sondeo. Debido a la gran cantidad de datos faltantes en los registros a niveles completos de radio-sondeo, y dado el buen grado de exactitud con que el modelo Eta-PRM reproduce perfiles de T y Td en su pronóstico a 12 hs (Simonelli et al, 2010), el valor del CAPE se estimó integrando de las salidas de este modelo dinámico en sus 20 niveles. En el apartado **Resultados 4.3** se presentan los resultados correspondientes a cada análisis propuesto.

Posteriormente se evaluó la efectividad del índice desarrollado en combinación con salidas del

modelo numérico Eta-PRM. Para ello las variables predictoras del modelo estadístico ajustado fueron extraídas de las salidas del modelo Eta-PRM en sus previsiones a cuatro días a partir del día de la corrida, para casos correspondientes a los períodos octubre-marzo de 2009/2010 a 2012/2013 como antes. El índice de inestabilidad calculado mediante el modelo estadístico alimentado por cada previsión del Eta-PRM para los distintos casos, se contrastó con las observaciones meteorológicas a fin de obtener los porcentajes de acierto y error bajo distintos criterios; en el apartado **Resultados 4.4** se muestran estos resultados.

## 4. Resultados

### 4.1. Análisis de Componentes Principales para los perfiles verticales de T y Td

La Tabla 4.1.I muestra la distribución de varianzas explicadas para las componentes obtenidas a partir del ACP para los 320 sondeos del ajuste. A pesar de tener una cantidad total de 12 observaciones por sondeo (6 niveles para  $T$  y 6 para  $Td$ ), las varianzas acumuladas del análisis de CPs revelan que cada sondeo de entrada puede escribirse como combinación lineal de sólo 11 patrones principales (rango de la matriz de correlaciones). Esto puede deberse a que en el nivel superior, la humedad relativa es lo suficientemente baja como para que la  $Td$  resulte dependiente exclusivamente de la  $T$ , lo cual reduce un grado de libertad en el sistema.

CP	Autovalor	Varianza Explicada [%]	Varianza Acumulada [%]
1	106,3693	33,24	33,24
2	61,6824	19,28	52,52
3	43,7762	13,68	66,20
4	38,5872	12,06	78,25
5	33,0374	10,32	88,58
6	17,2436	5,39	93,97
7	6,0275	1,88	95,85
8	5,2990	1,66	97,51
9	3,7385	1,17	98,68
10	2,7499	0,86	99,53
11	1,4890	0,47	100,00

Tabla 4.1.I: Varianzas correspondientes a las 11 CPs, explicadas porcentuales y acumuladas.

Por otra parte, la Figura 4.1.1 muestra el logaritmo de los autovalores en función del número de componente principal (diagrama de LEV).

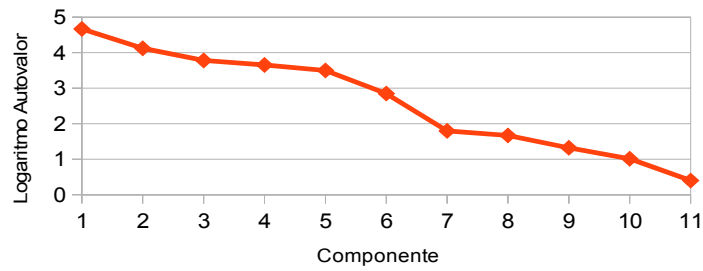
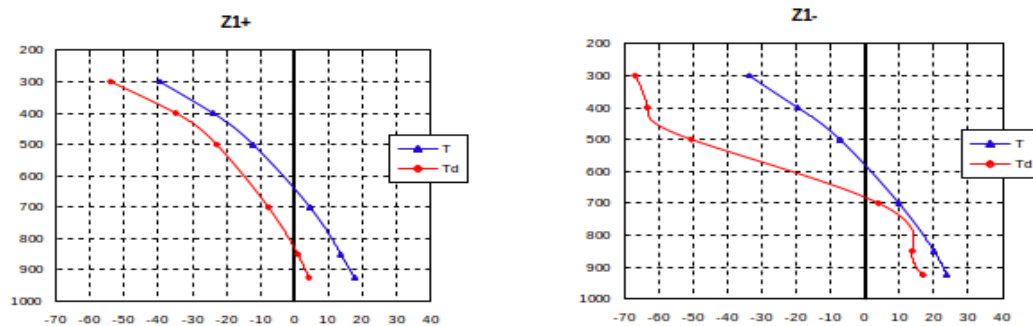
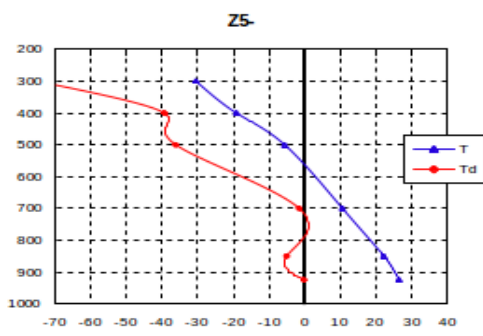
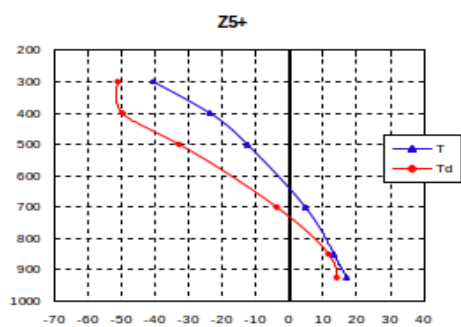
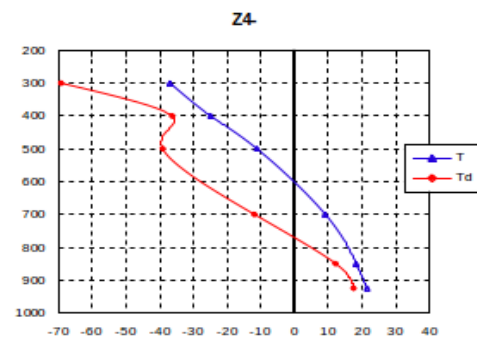
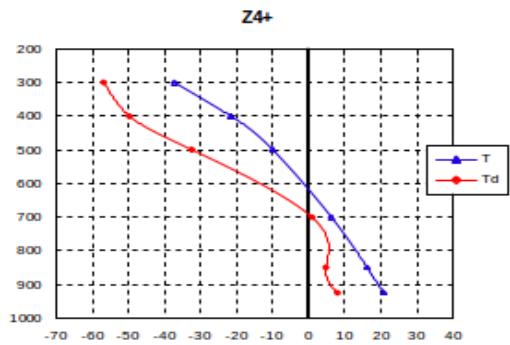
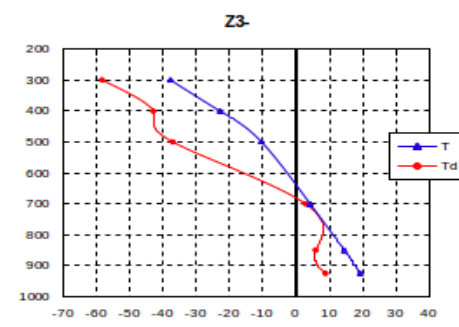
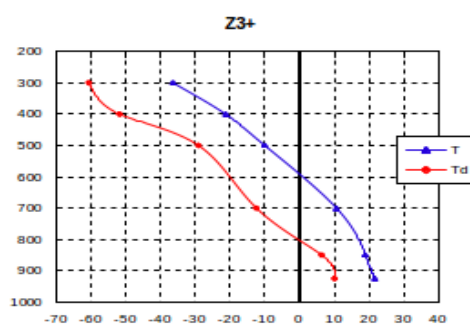
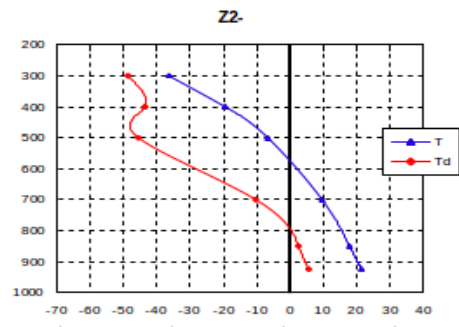
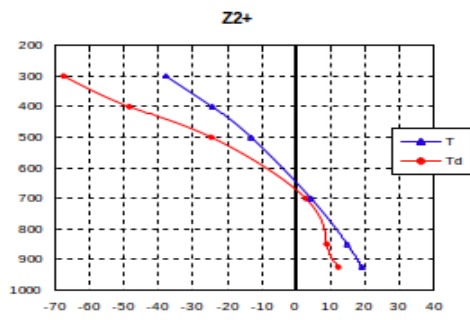


Figura 4.1.1: Diagrama de LEV para las componentes principales

Los valores del logaritmo decrecen casi linealmente para las primeras 6 componentes, observándose una marcada caída entre las componentes 6 y 7. Si bien se podrían tomar como significativas las primeras 6 CPs, las cuales acumulan el 94% de la varianza, en este trabajo y debido al objetivo propuesto se decide retener las 11 componentes para poder disponer del 100% de la información que las componentes explican.

Cada CP  $z$  puede interpretarse en su sentido directo  $[z(+)]$  o inverso  $[z(-)]$  de acuerdo lleven respectivamente componentes de carga  $f$  asociadas positivas o negativas en la ecuación [3]. En la Figura 4.1.2 se muestran las primeras 6 componentes principales, en sus modos directo e inverso, correspondientes al ajuste con la sub-muestra de 320 sondeos. Estos patrones, que explican casi la totalidad de la varianza del sistema, representan las mismas componentes que para el ajuste efectuado con la muestra completa de 659 sondeos, las cuales se muestran en el apéndice A junto a las restantes componentes correspondientes a cada ajuste. Puede observarse además que todos los patrones principales son identificables en ambas muestras, aunque el orden de acuerdo a varianza explicada para las componentes de orden mayor (componentes 7 a 11, que explican el 4% de varianza restante), pueda alterarse en la sub-muestra de 320 sondeos con respecto a la muestra completa de 659. Esto significa que los patrones obtenidos no dependen del tamaño de la muestra y que los definidos a partir de la sub-muestra 320 sondeos y empleados para el ajuste del modelo estadístico son representativos de la muestra completa.





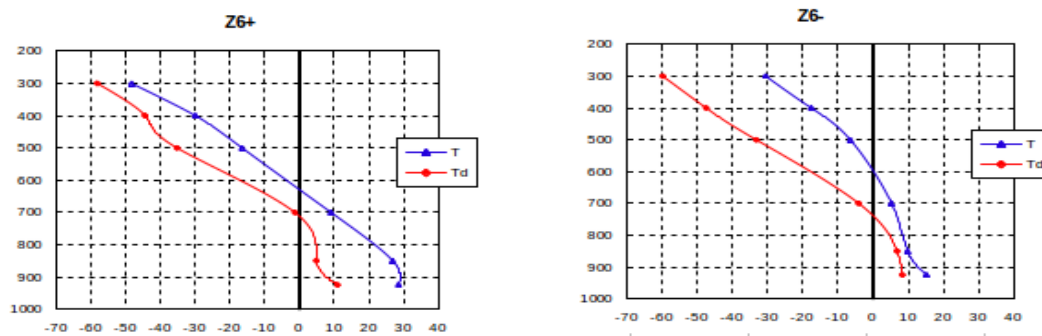


Figura 4.1.2: Modos directo e inverso de las primeras 6 CPs para el análisis con 320 sondeos.

Debido a que resulta difícil interpretar las desviaciones de los sondeos respecto del sondeo medio, en la Figura 4.1.2 se sumó el sondeo medio a las CPs no estandarizadas para facilitar su interpretación. Consecuentemente, y debido a que las CPs se calcularon en base a desvíos estandarizados de los sondeos originales, para los cuales las desviaciones de  $Td$  pueden superar en algún nivel a las de  $T$ ; en algunos patrones y para algunos niveles, los valores representados de  $Td$  pueden superar a los de  $T$  indicando un estado de saturación o próximo a la saturación. En otras palabras, las CPs son patrones de sondeos anómalos (desviaciones del perfil vertical medio de  $T$  y  $Td$ ), y por lo tanto modelos en los que eventualmente puede aparecer en algún nivel una desviación de  $Td$  mayor a la diferencia entre las medias climatológicas de  $T$  y  $Td$  para ese nivel. En ese caso, sumar el sondeo medio no bastará para compensar esa desviación de  $Td$  presente en el patrón y la misma se presentará mayor que  $T$  en el gráfico; para corregir este efecto se multiplicó la componente correspondiente por un factor de corrección que permita separar los perfiles de  $T$  y  $Td$  para lograr una mejor comprensión visual, este factor no influye sobre los cálculos pero permite evitar confusiones en la interpretación.

La CP 1 en su modo directo (inverso) muestra un sondeo mayormente húmedo (seco) entre los niveles de 500 y 300 hPa y medianamente seco (húmedo o saturado) entre los niveles de 700 y 920 hPa; mientras la CP 2 en modo directo (inverso) representa un sondeo húmedo o saturado (seco) en la capa 920-500 hPa, con un gradiente vertical de  $T$  más (menos) acentuado en niveles inferiores y una capa relativamente seca (húmeda) entre 400 y 300 hPa. Por su parte, la CP 3 en modo directo (inverso) exhibe un entorno seco (saturado) en 700 hPa y relativamente húmedo (seco) en 500 hPa con un débil (marcado) gradiente de  $Td$  y marcado (débil) gradiente de  $T$  entre estos dos niveles. La CP 4 en modo directo (inverso) muestra un entorno seco (húmedo) en 400 hPa y un débil (marcado) gradiente de  $Td$  en niveles bajos. La CP 5 en su modo directo (inverso) presenta un entorno saturado (seco) en el nivel de 850 hPa y relativamente más seco (húmedo) en 700 hPa (marcado gradiente (inversión) de  $Td$  en la capa 850-700 hPa). La CP 6 en modo directo (inverso) representa un sondeo seco (húmedo) en niveles bajos, que se humedece (seca) gradualmente con la altura.

A modo de ejemplo, en la Figura 4.1.3 se muestran sondeos de situaciones reales observadas que exhiben las mismas características que los patrones respectivos. Estos sondeos corresponden a los casos con el mayor factor de carga asociado al patrón correspondiente. Dichos factores de carga, junto a los restantes asociados a estos días particulares, se muestran en la Tabla 4.1.II.



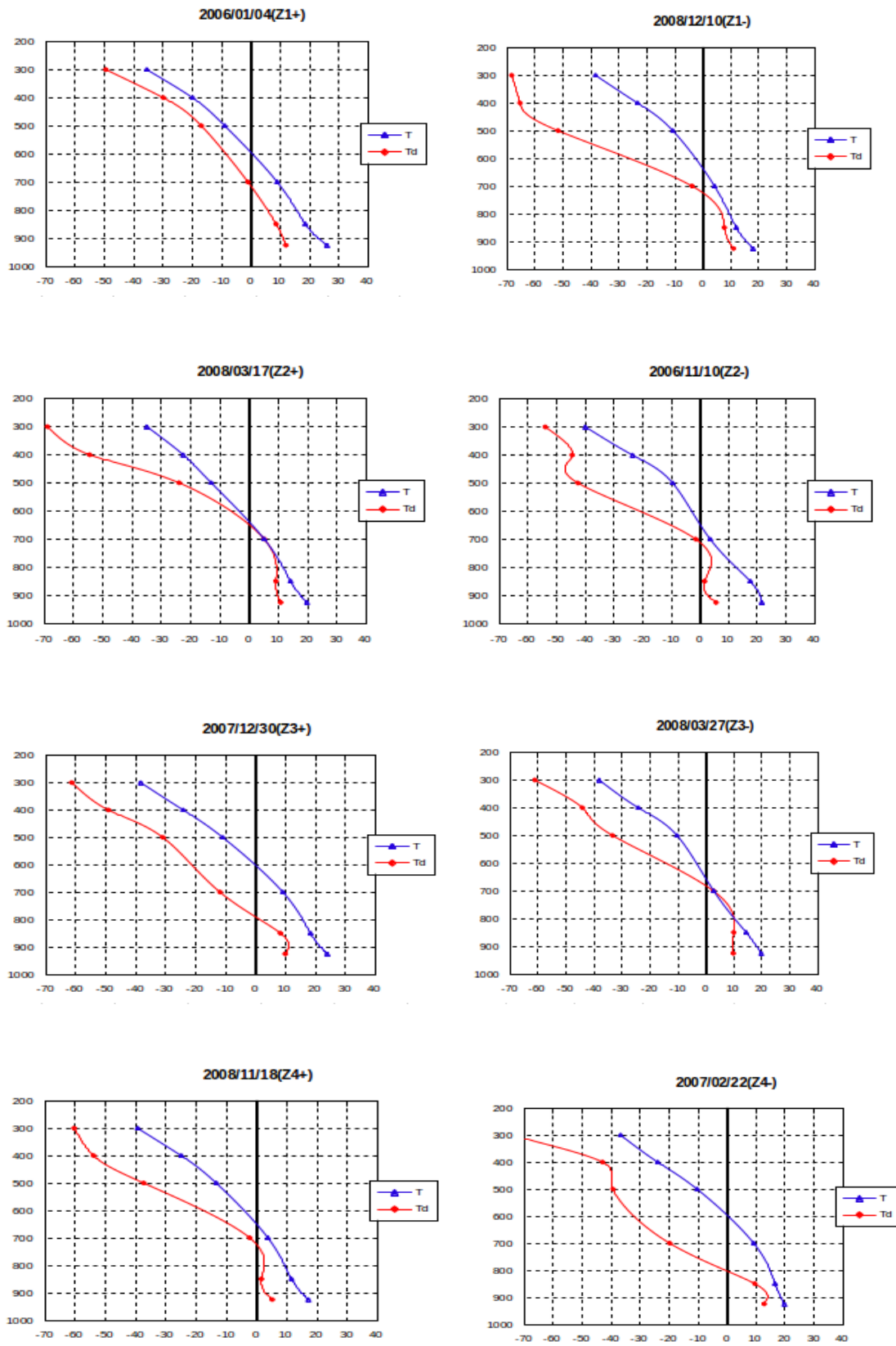


Figura 4.1.3: Sondeos reales asociados a los modos de las primeras cuatro CPs

FechaLoadings	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
2006/01/04	0,9790	0,0370	0,0150	-0,0136	0,0857	0,1185	0,1212	-0,0467	-0,0233	-0,0298	-0,0114
2008/12/10	-0,9790	-0,0253	0,0171	0,0033	0,1056	-0,1592	0,0168	0,0148	-0,0421	-0,0437	0,0040
2008/03/17	-0,1296	0,9284	0,0266	0,2711	-0,1277	-0,1478	0,0732	-0,0209	0,0262	0,0141	-0,0471
2006/11/10	0,0111	-0,8382	-0,3852	0,2358	-0,0610	0,2382	-0,0326	-0,0661	-0,1289	0,0666	0,0800
2007/12/30	0,0773	0,0473	0,8548	-0,3908	0,0427	0,1657	0,2460	0,0458	-0,0105	0,0765	0,1025
2008/03/27	0,1431	0,4494	-0,7783	0,0146	0,2049	-0,0238	0,1965	0,0795	-0,1820	0,1918	0,1193
2008/11/18	-0,2590	-0,0582	-0,0593	0,9098	0,0656	-0,2023	0,0256	-0,0386	0,1810	-0,1318	0,0269
2007/02/22	-0,0781	0,0108	0,3427	-0,8998	-0,1322	-0,2005	0,0458	0,0609	-0,0506	0,0452	-0,0084

Tabla 4.1.II: Factores de carga asociados a los casos con el mayor valor correspondiente a las primeras cuatro CPs, en azul se muestran los casos asociados con  $z$  [+] y en rojo con  $z$  [-].

La similitud entre los sondeos correspondientes a los días mostrados en la Figura 4.1.3 y los patrones correspondientes presentados en la Figura 4.1.2, valida la interpretación de las componentes descritas como patrones de sondeos reales, asociados a casos de situaciones atmosféricas efectivamente existentes en la muestra.

## 4.2. Ajuste del modelo de pronóstico a partir de los patrones de perfiles verticales de $T$ y $Td$

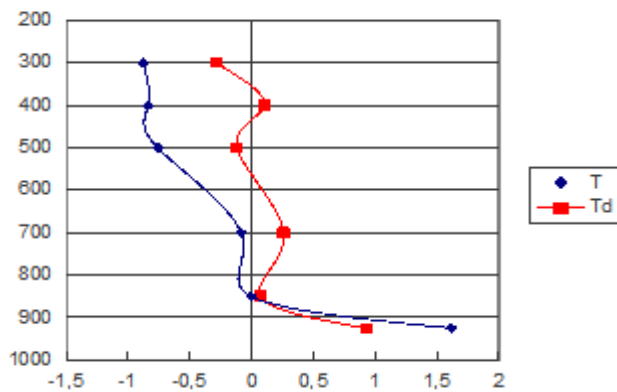
### Índice 12 UTC

Los coeficientes de ajuste  $b$  de la ecuación de regresión múltiple (3.5) para 278 sondeos se muestran en la Tabla 4.2.I. Los coeficientes significativos al 5% de confianza se encuentran resaltados en negrita. De acuerdo a los valores mostrados, las componentes 1, 2, 4 y 6 se encuentran significativamente correlacionadas con el índice convectivo  $c_{[278,x]}$ , las componentes 1 y 4 en modo inverso y las componentes 2 y 6 en modo directo. Esto implica, de acuerdo a los patrones mostrados en la Figura 4.1.2, que la actividad convectiva se encuentra favorecida en ambientes húmedos o saturados en las capas más bajas de la atmósfera, con un gradual secamiento y aumento del gradiente de  $T$  en las capas superiores. En estas condiciones, la inestabilidad y la energía potencial convectiva disponible tiende naturalmente a aumentar.

$b_0$	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
1,36336	-0,910869	1,61758	-0,829485	-2,01862	0,076142	5,88182	2,48787	-1,57754	-0,571692	-3,30254	0,384927

Tabla 4.2.I: Coeficientes de ajuste  $b$  de la ecuación de regresión múltiple logística. Los valores en negrita indican coeficientes distintos de cero al 5% de significancia.

Los valores estimados del coeficiente de C-NC pueden obtenerse a partir de la ecuación (3.7). Como se mencionó, el vector  $\mathbf{A}$  y el coeficiente  $\mathbf{b}_0$  sólo dependen de los resultados del ACP y del análisis de regresión (3.4)-(3.5), y el producto escalar en la ecuación (3.7) entre este vector y el vector de sondeo anómalo estandarizado  $\tilde{x}_s$  indica que el índice  $\hat{c}$  representa una medida de similitud entre ambos, ya que la ecuación logística (3.7) es una función monótona creciente dependiente dicho producto escalar y acotada entre 0 y 1. En otras palabras, el vector  $\mathbf{A}$  representa la combinación lineal de componentes principales que mejor ajusta los casos de C-NC, y en consecuencia, los perfiles verticales de  $T$  y  $Td$  mostrados por el vector  $\mathbf{A}$  representan una situación ideal para la ocurrencia de convección, los cuales pueden verse en la Figura 4.2.1 (junto con los coeficientes correspondientes al vector  $\mathbf{A}$  y el  $\mathbf{b}_0$ ).



	T925	T850	T700	T500	T400	T300	Td925	Td850	Td700	Td500	Td400	Td300
<b>A</b>	1,617	-0,0046	-0,086	-0,761	-0,837	-0,877	0,926	0,072	0,258	-0,119	0,104	-0,292

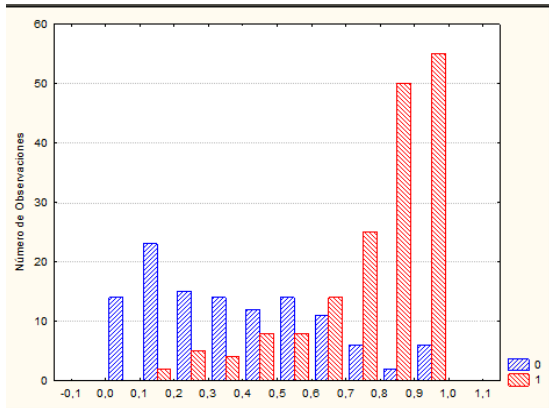
<b>bo</b>	1,363
-----------	-------

Figura 4.2.1: Gráfico del vector **A**, definido a partir de las CPs y los coeficientes de la ecuación de regresión múltiple logística. Además valores del vector **A** y de **bo**.

En 925 hPa se observa una fuerte anomalía positiva de  $T$  que pasa abruptamente a una marcada anomalía negativa en 500 hPa, manteniéndose así en niveles superiores, lo cual determina un fuerte gradiente térmico entre estas capas. Conjuntamente,  $Td$  presenta también anomalías marcadas positivas en capas bajas, disminuyendo levemente con la altura hasta el nivel de 500 hPa, a partir de donde se mantiene esta anomalía negativa salvo una pequeña anomalía positiva que se observa en el nivel de 400 hPa.

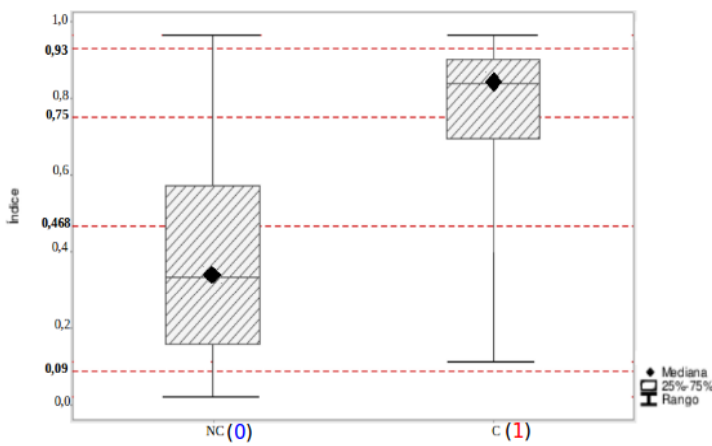
Esta situación caracterizaría una atmósfera muy caliente en capas bajas y muy fría y húmeda (probablemente saturada) en capas medias y altas, favoreciendo la ocurrencia de convección. La situación inversa caracterizaría una atmósfera estable en la cual la convección se vería inhibida.

Mediante la ecuación (3.7) se calcularon los valores estimados del índice de C-NC para 288 sondeos (de una sub-muestra distinta a la utilizada para el ajuste del modelo), separándoles luego de acuerdo a los casos de ocurrencia real de C (identificados con el valor 1) o NC (identificados con el valor 0). La Figura 4.2.2 muestra un histograma de las correspondientes distribuciones de frecuencias de los valores de índice estimado, junto con la tabla de frecuencias para cada caso y sus gráficas box-plot en las que se representan sus valores extremos y cuartiles. Estos últimos resultaron de 0,16, 0,331 y 0,559 para el caso NC y 0,692, 0,837 y 0,9 para el caso C.



Número de Observaciones	0	1
0 a 0,1	14	0
0,1 a 0,2	23	2
0,2 a 0,3	15	5
0,3 a 0,4	14	4
0,4 a 0,5	12	8
0,5 a 0,6	14	8
0,6 a 0,7	11	14
0,7 a 0,8	6	25
0,8 a 0,9	2	50
0,9 a 1	6	55

Figura 4.2.2: Panel superior) Histograma de frecuencias de los valores del índice de C-NC(1-0) estimado a partir de la ecuación (7) para 288 sondeos separados en 171 eventos en los que efectivamente se observó convección(conjunto C) y 117 en los que no se observó (conjunto NC).



Panel inferior) Box-plot asociado a los datos del histograma anterior. Los valores resaltados de 0,93 y

Los índices superiores a 0,75 corresponden a valores que se encuentran en la cola derecha del 10% de la distribución de casos NC, mientras los inferiores a 0,468 se encuentran en la cola izquierda del 10% de los casos C. De esta manera, dado un sondeo cualquiera sobre el cual se estima el índice a partir de la ecuación (3.7), un valor superior a 0,75 indicaría una previsión de ocurrencia de convección con una probabilidad de error menor al 10%. Análogamente, un valor inferior a 0,468 indicaría un pronóstico de no ocurrencia de convección con igual probabilidad de error. El rango comprendido entre 0,468 y 0,75 representa un intervalo de incertidumbre para el cual el índice no discrimina entre casos de C-NC con estos niveles de error. Por tal motivo, se procede a definir un criterio a partir del cual se establezca un pronóstico de ocurrencia de convección si el índice estimado es mayor a cierto valor y la no ocurrencia cuando es menor a dicho valor (probabilidad o valor de corte). Para ello se ordenan en forma creciente todos los valores del índice obtenidos mediante la ecuación (3.7) para la cantidad de sondeos de la muestra, es decir, 288 valores del índice estimado, resultando:

Orden	1	2	3	4	5	...	284	285	286	287	288
Criterio de C-NC	0,0198	0,0264	0,0308	0,0361	0,0428	...	0,9808	0,9840	0,9876	0,9886	0,9893

A continuación se establece un valor de corte, a partir del cual se define el *criterio de ocurrencia o no de convección*, de entre los valores muestrales del índice ya ordenados en forma creciente y se define como *orden* a los números naturales que identifican el carácter creciente del valor del índice. A fin de determinar la efectividad del modelo de pronóstico bajo el criterio establecido, se calcula el porcentaje de aciertos sumando la cantidad de casos en los cuales se verifica el pronóstico

(observación de convección con índice mayor al valor de corte y no ocurrencia de convección con índice menor a dicho valor), y dividiendo por el total de casos de la muestra (en particular 288 para el presente análisis). Conjuntamente, se determinan los porcentajes de error por falsa alarma (es decir, se pronosticó convección y no la hubo), y sorpresa o imprevisto (es decir, no se pronosticó convección y la hubo), contando los casos correspondientes y dividiendo por el total de casos de la muestra.

A modo ejemplo, tomando como Criterio de C-NC un valor de corte igual a 0,5, se encuentra que la efectividad total del índice es de un 79,8611%, mientras que el error se reparte en 13,5416% de probabilidad de falsa alarma y 6,5972% de probabilidad de sorpresa.

Dado que el porcentaje de efectividad depende del Criterio de C-NC establecido, o sea del valor de corte definido, dicho criterio (o valor de corte) se puede variar a fin de maximizar la efectividad y al mismo tiempo disminuir los errores. En la Tabla 4.2.II se muestran los valores de efectividad, sorpresa y falsa alarma para el Criterio de C-NC según el valor de 0,5 y para el valor de 0,6613 que corresponde a la máxima efectividad que se puede obtener para los 288 casos analizados.

	Efectividad [%]	Sorpresa [%]	Falsa Alarma [%]
<b>0,5</b>	79,86	6,59	13,54
<b>Máximo Muestral (0,6613)</b>	83,33	11,11	5,55
<b>Máximo Ajustado (0,5768)</b>	80,01	8,27	11,7

Tabla 4.2.II: Valores de efectividad, sorpresa y falsa alarma para el valor de 0,5 a modo de ejemplo, para el máximo muestral (0,6613) que corresponde al máximo porcentaje de acierto que se puede obtener para el pronóstico de C-NC y para el máximo ajustado (0,5768) que corresponde al máximo porcentaje de acierto que se puede obtener para el pronóstico de C-NC según la curva de efectividad ajustada a la muestra.

Las distribuciones de frecuencias observadas para los casos de C y NC mostradas en la Figura 4.2.2, pueden ajustarse a una distribución de densidad de probabilidad Gamma, con un 5% de confianza de acuerdo al test  $\chi^2$ , haciendo previamente un cambio de variable  $t = 1-x$  para la distribución de los casos convectivos C, lo cual se muestra en la Figura 4.2.3.

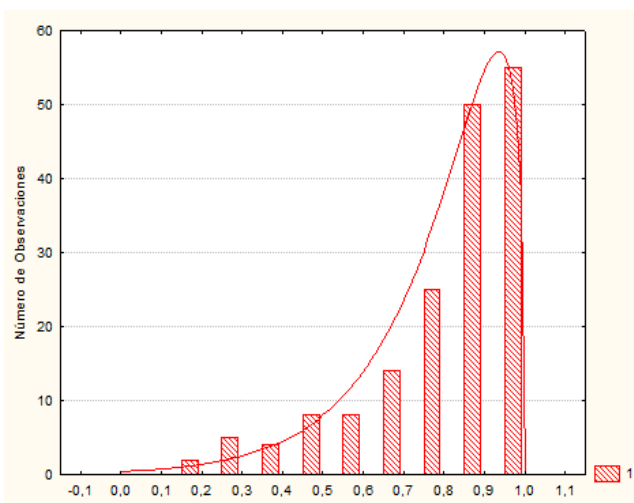


Figura 4.2.3: Distribución de frecuencias para el caso C junto con la curva ajustada.

La función teórica ajustada correspondiente a los casos de C (o casos 1) resulta:

$$f(x) = 270,245 (1-x)^{0,414} e^{-\frac{(1-x)}{0,1546}} \quad (4.2.a)$$

Luego se ajusta la segunda distribución Gamma para los casos de NC (o casos 0) que se muestra en la Figura 4.2.4.

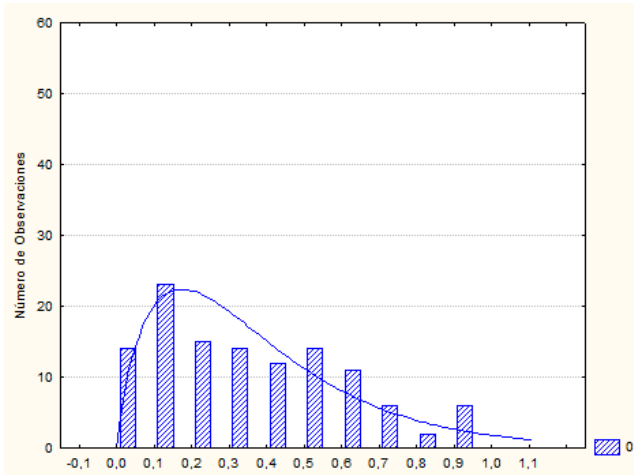


Figura 4.2.4: Distribución de frecuencias para el caso NC junto con la curva ajustada.

La función teórica ajustada correspondiente a los casos de NC resulta:

$$f(x) = 226,725 x^{0,842} e^{-\frac{x}{0,2065}} \quad (4.2.b)$$

Para la función (4.2.a) se tiene el 90% del área bajo la curva entre los valores  $x=0,546$  y  $x=1,0$  y para la función (4.2.b) se tiene el 90% del área bajo la curva entre los valores  $x=0,0$  y  $x=0,77$ .

De esta manera, dado un sondeo particular sobre el cual se estima el índice a partir de la ecuación (3.7), un valor superior a 0,77 indicaría la ocurrencia futura de convección con una probabilidad de error menor al 10%. Análogamente, un valor inferior a 0,546 indicaría la no ocurrencia de convección con igual probabilidad de error. El rango comprendido entre 0,546 y 0,77 representa un intervalo de incertidumbre en el cual el índice no discrimina para estos niveles de error. Este intervalo es semejante al encontrado previamente mediante el análisis de los percentiles correspondientes, a partir de los datos de la muestra examinada.

Además, para la función (4.2.a) se tiene el 91,97% del área bajo la curva entre los valores  $x=0,5$  y  $x=1,0$ ; y para la función (4.2.b) se tiene el 73,61% del área bajo la curva entre los valores  $x=0,0$  y  $x=0,5$ ; con lo cual si se toma como criterio *natural* de ocurrencia de convección y no convección el valor de corte de 0,5, la *efectividad teórica* basada en el ajuste de los histogramas resulta del 82,79%, con un error por *falsa alarma* del 13,19% y por *sorpresa* del 4,02%.

Por otra parte, como se ha mencionado, tanto el porcentaje de efectividad del modelo como los de sus errores dependen del valor de corte elegido para definir el criterio de C-NC, el cual se puede

variar a fin de maximizar la efectividad y minimizar los errores. Si bien el máximo porcentaje de efectividad resulta del 83,3% para un valor de corte igual a 0,66 (Tabla 4.2.II), la función de efectividad puede ajustarse mediante un polinomio de grado 2 a fin de obtener un valor de corte que maximice la efectividad ajustada que sea más representativo de la muestra en su conjunto y no tan dependiente de un valor particular de la misma. A su vez, una función del tipo  $f(x) = a e^{bx} x^c$  ajusta muy bien los valores de error por sorpresa. En la Figura 4.2.5 se muestran las gráficas de los conjuntos de valores de efectividad, sorpresa y falsa alarma junto con las curvas que mejor se ajustan correspondientemente (las curvas se han ajustado en función del orden de los diferentes conjuntos de valores y explican una varianza mayor al 96%). La curva que mejor ajusta los valores de la efectividad está dada por:

$$f(x) = -0,00148x^2 + 0,33740x + 60,77958 \quad (4.2.c)$$

y la curva que mejor ajusta los valores de la sorpresa está dada por:

$$f(x) = 0,000006 x^{3,120914} e^{-0,005470 x} \quad (4.2.d)$$

La curva de ajuste de los valores de falsa alarma resulta como consecuencia de estas últimas dos curvas ajustadas ya que cada valor de falsa alarma es la diferencia entre 100% y la suma de los dos valores de efectividad y sorpresa correspondientes.

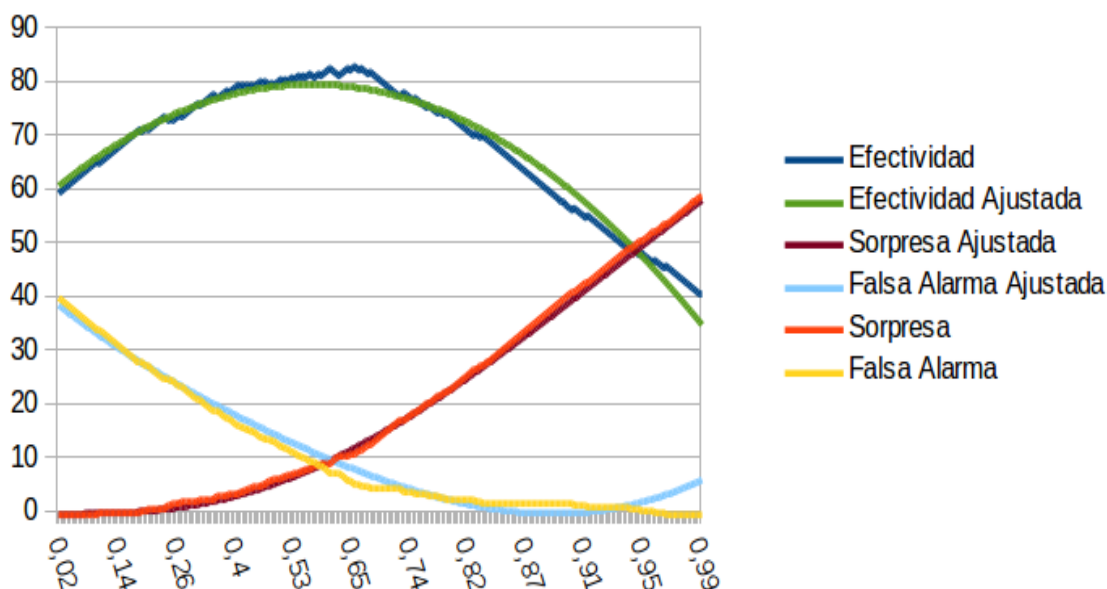


Figura 4.2.5: Gráficas de los conjuntos de valores de efectividad, sorpresa y falsa alarma junto con las curvas ajustadas correspondientemente en función de los Criterios de C-NC.

De aquí se obtiene, como se muestra en la Tabla 4.2.II, que el máximo valor de efectividad es de un 80,01%, correspondiente al Criterio de C-NC con valor 0,5768 y donde el error se reparte en 11,71% de probabilidad de falsa alarma y 8,28% de probabilidad de sorpresa.

En síntesis, para el modelo ajustado a partir de las observaciones efectuados en las 12 horas posteriores al radio-sondeo (índice 12UTC), se obtiene que, dado un sondeo sobre el cual se estima el índice a partir de la ecuación (3.7), si el valor estimado del índice es superior a 0,77 indicaría la ocurrencia futura de convección (dentro de las 12 horas siguientes) con una probabilidad de error menor al 10%, y si el valor es inferior a 0,55 indicaría la no ocurrencia de convección con igual probabilidad de error. Adicionalmente, si el criterio de ocurrencia de convección o no convección se establece con un valor de corte de alrededor de 0,577, la efectividad del modelo se eleva a alrededor del 80%, con un 11,70% de probabilidad de error por falsa alarma y un 8,28% de probabilidad de error por sorpresa.

El resultado que se utilice como opción para hacer un pronóstico de convección para una dada situación depende justamente del valor que estima el índice según (3.7) para el sondeo en cuestión.

### Índice 9 UTC

Como se ha mencionado, con las observaciones de convección realizadas en las 12 horas siguientes al sondeo (índice 12 UTC), es posible que parte del error por falsa alarma se deba a inestabilidad producida en las horas previas al sondeo, que desaparece en las horas posteriores. En este caso, el sondeo podría estar representando un ambiente inestable, arrojando un valor alto del índice, pero la convección no se produciría en las horas posteriores al sondeo resultando un pronóstico fallido. Análogamente, parte del error por sorpresa podría deberse a actividad convectiva producida en las horas previas al sondeo, debido a inestabilidad producida en las primeras horas de la mañana. Esta inestabilidad puede resultar imposible de pronosticar si el sondeo de las 12 UTC del día anterior se desarrolló en un ambiente estable que luego se inestabilizó durante la madrugada del día siguiente, resultando un índice de bajo valor y consecuentemente un pronóstico fallido. A fin de verificar la posibilidad de que parte de los errores en el modelo se deban a la existencia de casos como los mencionados, se ajustó el modelo tomando las observaciones de C-NC en el lapso de 12 horas desde 3 horas antes de la realización de cada sondeo, esto es, desde las 9 UTC del día del sondeo, hasta las 9 UTC del día siguiente (Índice 9 UTC).

Los coeficientes de ajuste  $b$  de la ecuación de regresión múltiple (3.5) para 273 sondeos se muestran en la Tabla 4.2.III. Los coeficientes significativos al 5% de confianza se encuentran marcados en negrita. De acuerdo a los valores mostrados, las componentes 1, 2, 3, 4, 6 y 7 se encuentran significativamente correlacionadas con el índice convectivo  $c_{[273 \times 1]}$  (las componentes 1, 3 y 4 en modo inverso y las componentes 2, 6 y 7 en modo directo).

bo	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11
1.55854	<b>-0.759699</b>	1.49086	<b>-1.04277</b>	<b>-2.04859</b>	0.537259	<b>5.69828</b>	<b>3.12642</b>	-2.43473	-0.409536	-3.12234	-3.10474

Tabla 4.2.III: Coeficientes de ajuste  $b$  de la ecuación de regresión múltiple logística. Los valores en negrita indican coeficientes distintos de cero al 5% de significancia.



Al igual que antes, mediante la ecuación (3.7) se calcularon los valores estimados del índice de C-NC para 285 sondeos independientes de la sub-muestra empleada para el ajuste. En la Figura 4.2.6 se muestra un histograma de las correspondientes distribuciones de frecuencias para los casos de C y NC junto con la tabla de frecuencias para cada caso y en la Tabla 4.2.IV se muestran los valores de efectividad, sorpresa y falsa alarma para el criterio de C-NC según el valor de 0,5 y para el valor de 0,5916, valor de corte que corresponde a la máxima efectividad que se puede obtener considerando los 285 valores de efectividad calculados.

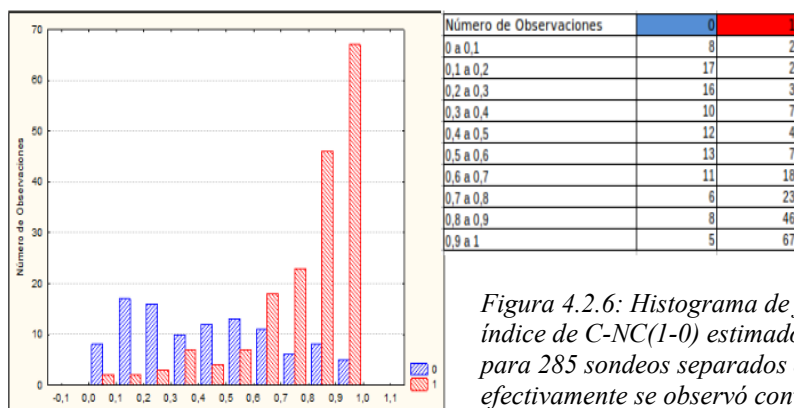


Figura 4.2.6: Histograma de frecuencias de los valores del índice de C-NC(1-0) estimado a partir de la ecuación (7) para 285 sondeos separados en 179 eventos en los que efectivamente se observó convección(conjunto C) y 106 en los que no se observó (conjunto NC).

	Efectividad [%]	Sorpresa [%]	Falsa Alarma [%]
<b>0,5</b>	78,59	6,31	15,08
<b>Máximo Muestral(0,5916)</b>	80,7	8,42	10,87

Tabla 4.2.IV: Valores de efectividad, sorpresa y falsa alarma para el valor de 0,5 a modo de ejemplo y para el máximo muestral (0,5916). Este último valor corresponde al máximo porcentaje de acierto que se puede obtener para el pronóstico de C-NC.

No se ha hecho un análisis en detalle para el índice 9 UTC ya que como se puede ver en las Tablas 4.2.II y 4.2.IV, el valor máximo muestral de la efectividad para el índice 12 UTC es mayor que para el índice 9 UTC con lo cual, dado objetivo propuesto en este trabajo, en adelante se decide continuar trabajando con el índice 12 UTC.

### **4.3. Ajustes del modelo pronóstico a partir de los patrones rotados y sin rotar de perfiles de T y Td, la cortante vertical de viento, la helicidad y el CAPE**

En este apartado se muestran los valores de máxima efectividad del índice 12 UTC al aplicar el método de rotación VARIMAX y al incorporar al análisis las variables SRH, VWS y CAPE como predictores.

Para ello se toman 195 sondeos del total de los 320 utilizados para el ajuste, para los cuales no solo se tiene información técnica real de la ocurrencia o no convección sino también información de viento (i.e. para calcular helicidad y cortante), y los valores de CAPE obtenidos a partir de las salidas del modelo Eta-PRM. Además se toman 246 sondeos del total de 339, los cuales se reservan para evaluar la efectividad del modelo al pronosticar eventos de convección profunda que cumplen con las mismas condiciones que los 195 mencionados con anterioridad. Por lo tanto, para cada análisis se dispondrá de los mismos sondeos para el proceso de ajuste del modelo estadístico y de los mismos sondeos para el proceso que nos permitirá definir la máxima efectividad para cada ajuste.

De acuerdo a los resultados obtenidos, en las Tablas 4.3.I y 4.3.II se ve que ni las variables dinámicas o energéticas, ni la rotación VARIMAX mejoran sustancialmente la efectividad del índice. Es importante resaltar que en este apartado lo que se quiere hacer es una comparación entre valores, es decir, que los valores que se obtengan de efectividad máxima para cada uno de los análisis no son relevantes porque lo que se quiere ver es si en igualdad de condiciones (misma muestra) la inclusión de nuevos predictores o la rotación VARIMAX mejoran el resultado de máxima efectividad del índice 12 UTC.

No se presentan en detalle los distintos análisis ya que los procedimientos que se siguen en cada uno son los mismos que los ya descritos en la metodología y en el apartado **Resultados 4.2** (*Índice 12 UTC*).

Análisis de Regresión Logística Múltiple				
Predictor	coeficientes b			
	Exp. 1	Exp. 2	Exp. 3	Exp. 4
b0	<b>1,1443</b>	<b>0,8821</b>	<b>1,0884</b>	1,7122
F1	<b>-0,8731</b>	<b>-0,7957</b>	<b>-0,8906</b>	<b>-0,8607</b>
F2	<b>1,6713</b>	<b>1,3252</b>	<b>1,6540</b>	<b>1,7410</b>
F3	-0,8972	-0,8103	-0,8463	-1,131
F4	<b>-2</b>	<b>-1,7437</b>	<b>-2,0789</b>	<b>-1,956</b>
F5	0,3228	0,1101	0,3759	0,6404
F6	<b>5,4288</b>	<b>5,0251</b>	<b>5,4330</b>	<b>6,1980</b>
F7	2,3531	2,0306	2,1147	<b>4,5451</b>
F8	<b>-3,7951</b>	-3,1978	<b>-3,8995</b>	-3,0427
F9	-0,3766	0,0349	-0,6497	-4,0079
F10	-1,3392	-0,6381	-1,3402	-3,3915
F11	0,3876	0,4843	0,4495	0,9646
sh_850-925	-----	-----	-----	0,0583
sh_700-850	-----	-----	-----	0,016
sh_500-700	-----	-----	-----	-0,0269
sh_400-500	-----	-----	-----	0,1145
sh_300-400	-----	-----	-----	0,0761
sh_700-925	-----	-----	-----	-0,0653
sh_500-850	-----	-----	-----	0,0432
sh_400-700	-----	-----	-----	0,0052
sh_300-500	-----	-----	-----	-0,1146
sh_500-925	-----	-----	-----	-0,0869
sh_400-850	-----	-----	-----	-0,0818
sh_300-700	-----	-----	-----	-0,0024
sh_400-925	-----	-----	-----	0,031
sh_300-850	-----	-----	-----	-0,0235
sh_300-925	-----	-----	-----	0,0949
Helicidad	-----	-----	-0,0003	-0,0007
CAPE	-----	0,0259	-----	-----
<b>0,5</b>				
Efectividad	80,89	80,89	80,89	78,86
Sorpresa	6,09	6,5	6,5	8,94
Falsa Alarma	13	12,6	12,6	12,19
<b>máximo muestral</b>				
<b>0,5706    0,5848    0,5406    0,479</b>				
Efectividad	82,92	82,11	82,92	79,67
Sorpresa	8,13	8,94	6,91	6,91
Falsa Alarma	8,94	8,94	10,16	13,41

Tabla 4.3.I: Coeficientes b del análisis de regresión logística múltiple (para cuatro experimentos usando diferentes combinaciones de predictores) correspondientes a los predictores: loadings ( $F_k$ ), cortante entre los niveles  $i$  y  $j$  ( $sh_{i-j}$ ), Helicidad y CAPE. Los coeficientes significativos al 5% de confianza se encuentran resaltados en negro. Para los diferentes experimentos (pronosticando “convección” cuando el  $c$  estimado es  $> 0.5$  y “no convección” cuando el  $c$  estimado es  $< 0.5$ ; o pronosticando convección y no convección según el valor del máximo muestral) se muestran los porcentajes de efectividad, sorpresa y falsa alarma.

Análisis de Regresión Logística Múltiple		
Predictor	coeficientes b	
	Exp. 1	Exp. 2
b0	<b>1,1443</b>	<b>1,2397</b>
F1	<b>-1,1554</b>	-0,5442
F2	<b>-0,9928</b>	-0,5788
F3	-0,0759	-0,6737
F4	<b>-1,94</b>	<b>-1,568</b>
F5	<b>-1,7244</b>	<b>-1,437</b>
F6	<b>-5,0014</b>	-----
F7	1,3491	-----
F8	<b>-3,6871</b>	-----
F9	0,0635	-----
F10	3,2335	-----
F11	0,113	-----
	<b>0,5</b>	
Efectividad	80,89	67,01
Sorpresa	6,09	3,47
Falsa Alarma	13	29,51
	<b>máximo muestral</b>	
	<b>0,5706</b>	<b>0,6224</b>
Efectividad	82,92	70,83
Sorpresa	8,13	8,33
Falsa Alarma	8,94	20,83

Tabla 4.3.II: Coeficientes b del análisis de regresión logística múltiple para dos experimentos correspondientes a tomar como predictores a los loadings ( $F_k$ ) que resultan de aplicar el método de rotación VARIMAX a las 11 componentes principales y a 5 de ellas elegidas según el criterio de Kaiser. Los coeficientes al 5% de significancia se encuentran resaltados en negro. Para los diferentes experimentos (pronosticando “convección” cuando el c estimado  $> 0.5$  y “no convección” cuando el c estimado  $< 0.5$ ; o pronosticando convección y no convección según el valor del máximo muestral) se muestran los porcentajes de efectividad, sorpresa y falsa alarma.

#### 4.4. Combinación del modelo 12 UTC con las salidas del modelo numérico Eta-PRM

En este apartado se evalúa la efectividad del índice 12 UTC desarrollado a partir de datos de radio-sondeo y observaciones de superficie, en combinación con salidas del modelo numérico Eta-PRM, esto quiere decir que en la ecuación (3.7) la matriz  $\tilde{X}_s$  de sondeos estandarizados no estará dada en función de información de radio-sondeos reales como se ha hecho hasta el momento sino a partir de sondeos pronosticados por el modelo Eta, no solo para el día de la corrida sino también en sus previsiones a cuatro días a partir de este. El objetivo de esta evaluación es verificar el máximo valor de efectividad ajustada para el día de la corrida, a fin de cotejar la confiabilidad de los sondeos modelados por el Eta respecto de los sondeos observados (por comparación respecto de la máxima efectividad ajustada obtenida para el índice 12 UTC en el apartado **Resultados 4.2**), determinar la efectividad máxima para los cuatro días siguientes y establecer los criterios de ocurrencia de convección a tomar en cuenta para cada día.

Mediante la ecuación (3.7) se calcularon los valores estimados del índice de C-NC para el día de la corrida (Día 0) del modelo Eta y para los cuatro días posteriores. En la Figura 4.4.1 se muestran los histogramas de las distribuciones de frecuencias asociadas a los casos de ocurrencia real de C o NC para cada día (en el apéndice B se especifica, para cada día, la cantidad de sondeos que se han utilizado para estos cálculos, cuántos corresponden a eventos en los que efectivamente se observó convección (conjunto C) y a eventos en los que no se observó (conjunto NC); y la tabla de frecuencias asociada a cada histograma).

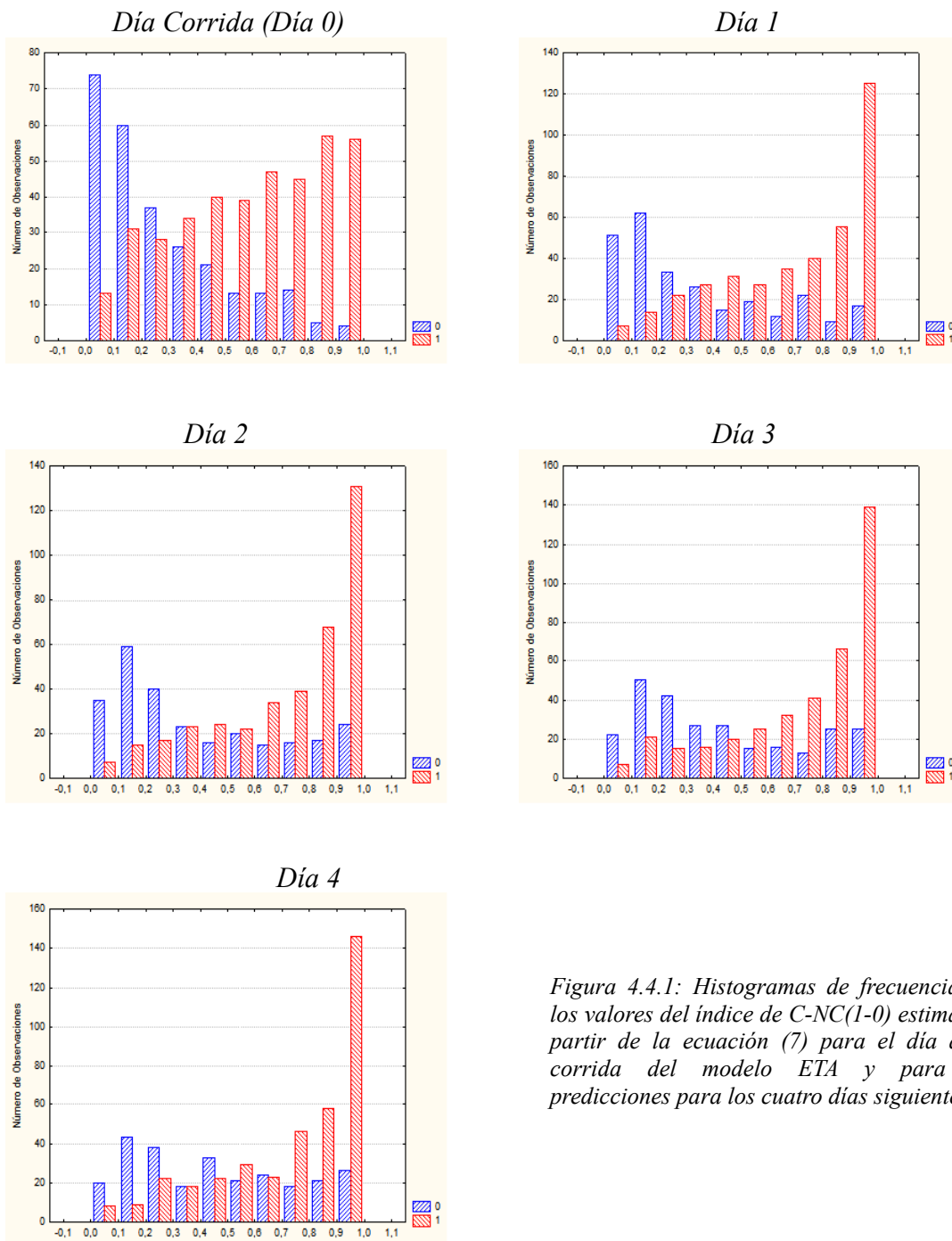


Figura 4.4.1: Histogramas de frecuencias de los valores del índice de C-NC(1-0) estimado a partir de la ecuación (7) para el día de la corrida del modelo ETA y para sus predicciones para los cuatro días siguientes.

Para el Día de la corrida (Día 0), el histograma correspondiente para los casos NC, muestra una mejor performance que para el caso de sondeos reales (índice 12 UTC mostrado en la sección Resultados 4.2), resultando una disminución del porcentaje de error por falsa alarma bajo los mismos criterios de ocurrencia (por ejemplo con un valor de corte de 0,5). Sin embargo, los pronósticos efectuados en días con observación real de convección (casos C) muestran una exactitud menor que en el caso de utilización de sondeos reales (no simulados), lo cual implica un aumento considerable en el porcentaje de error por sorpresa o imprevistos con respecto al error obtenido para el índice 12 UTC con iguales criterios de ocurrencia. Los Días 1 y 2 muestran distribuciones de índices similares a las obtenidas para sondeos reales, con muy buen desempeño en los casos de días con convección, aunque performances no tan buenas como las obtenidas para

sondeos reales en los casos de días sin observación de convección. Esta disminución en la habilidad del modelo de pronosticar casos de NC disminuye aún más para el caso de sondeos modelados correspondientes a los días 3 y 4 de la corrida, con un aumento en los porcentajes de error por falsa alarma para un mismo criterio de ocurrencia.

En la Tabla 4.4.I se muestran los valores de efectividad, sorpresa y falsa alarma para el Criterio de C-NC según el valor de 0,5 y para el valor que corresponde al máximo muestral, es decir, que corresponde a la máxima efectividad que se puede obtener para la cantidad de casos analizados (destacado entre paréntesis en cada caso).

		Efectividad [%]	Sorpresa [%]	Falsa Alarma [%]
<b>Día Corrida</b> <b>(Día 0)</b>	0,5	70,31	22,22	7,45
	Máximo Muestral(0,3164)	75,95	11,26	12,78
	Máximo Ajustado(0,3058)	73,19	11,68	15,12
<b>Día 1</b>	0,5	72,26	15,56	12,17
	Máximo Muestral(0,2719)	75,19	4,93	19,87
	Máximo Ajustado(0,3888)	73,13	10,61	16,25
<b>Día 2</b>	0,5	72,4	13,33	14,26
	Máximo Muestral(0,3372)	75,03	7,13	17,82
	Máximo Ajustado(0,4456)	72,13	11,19	16,67
<b>Día 3</b>	0,5	73,13	12,26	14,59
	Máximo Muestral(0,4507)	73,13	10,4	16,45
	Máximo Ajustado(0,4750)	70,74	12,54	16,7
<b>Día 4</b>	0,5	70,6	12,28	17,1
	Máximo Muestral(0,4665)	70,91	10,88	18,19
	Máximo Ajustado(0,5005)	69,91	12,61	17,46

*Tabla 4.1.I: valores de efectividad, sorpresa y falsa alarma para el día de la corrida del modelo ETA y para los cuatros días posteriores, para el valor de 0,5 a modo de ejemplo, para el máximo muestral que corresponde a la máxima efectividad que se puede obtener y para el máximo ajustado que corresponde a la máxima efectividad que se puede obtener según la curva de efectividad (polinomio de grado 2) ajustada a la muestra.*

El mejoramiento en la habilidad de pronóstico del modelo para los casos con observación real de convección en los días siguientes al de la corrida, se ve reflejado en la disminución de los errores por sorpresa bajo un mismo criterio de ocurrencia (por ejemplo tomando 0,5 como valor de corte en todos los casos, este error pasa de 22,2% para el Día 0 a 12,3% para los Días 3 y 4). Recíprocamente, el empeoramiento del pronóstico en los casos de NC se refleja en un aumento del error por falsa alarma para los días siguientes al de la corrida también bajo el mismo criterio de ocurrencia (tomando 0,5 como valor de corte, este error pasa de 7,5% para el Día 0 a 17,1% para el Día 4). Esto redundará en una efectividad mayor para los Días 1 a 3 que para el Día 0 (70% para el Día 0 y alrededor del 72% para Días siguientes).

Es importante mencionar que, a diferencia de lo hecho en el apartado 4.2 para el índice 12 UTC, para cada uno de los Días que se presentan en este apartado no ha sido posible ajustar una distribución de densidad de probabilidad Gamma, ni ninguna otra función usualmente empleada a tales fines, para las distribuciones de frecuencias observadas para los casos de C y NC de los histogramas mostrados, con niveles de significancia razonables. Por otra parte al igual que antes, la

función de efectividad puede ajustarse mediante un polinomio de grado 2 y la función del tipo  $f(x) = a e^{bx} x^c$  ajusta muy bien los valores de error por sorpresa. En la Figura 4.4.2 se muestran las gráficas de los conjuntos de valores de efectividad, sorpresa y falsa alarma junto con las curvas que mejor se ajustan correspondientemente, para cada uno de los Días (las curvas se han ajustado en función del orden de los diferentes conjuntos de valores y explican una varianza mayor al 97%; en el apéndice C se muestran las funciones obtenidas).

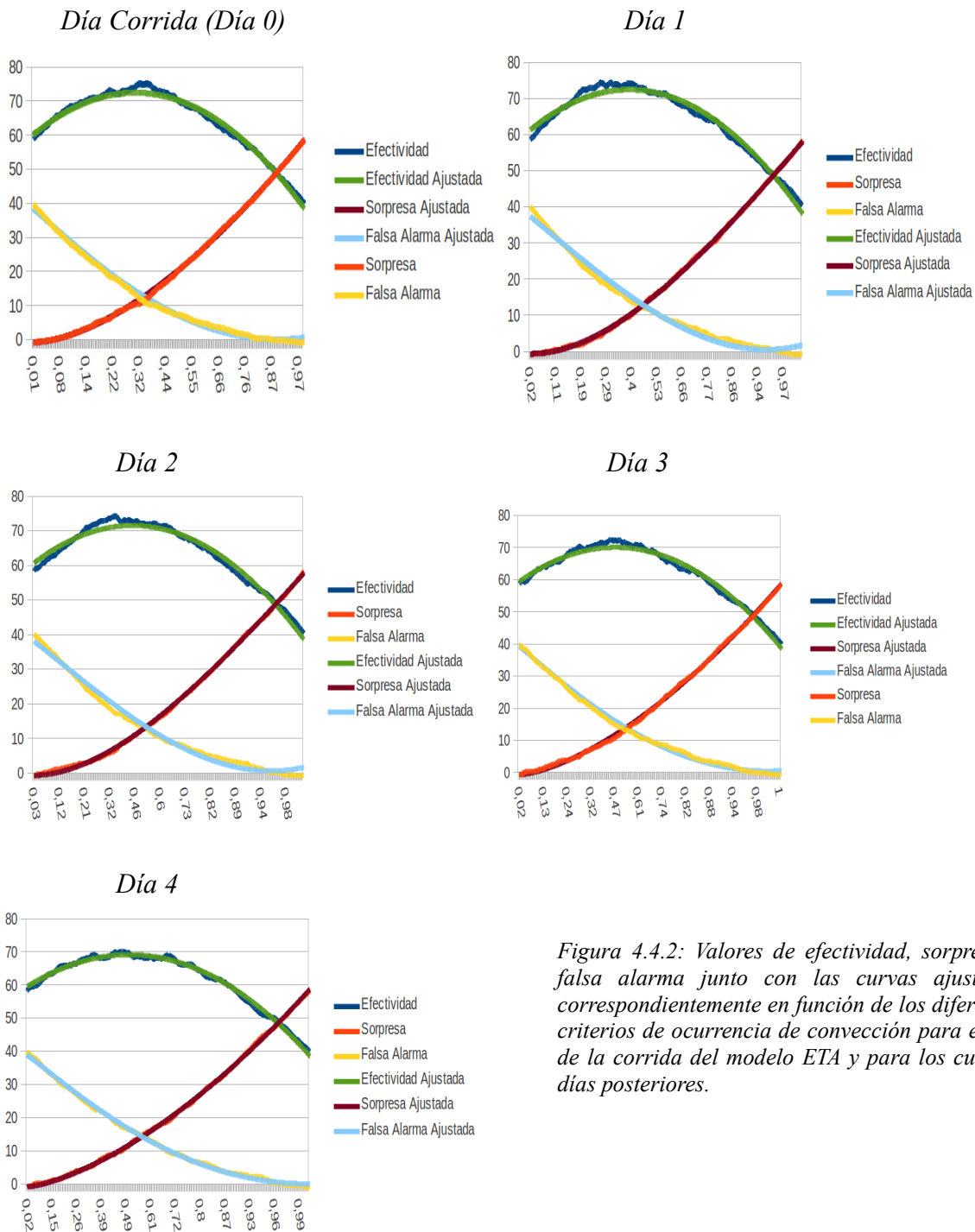
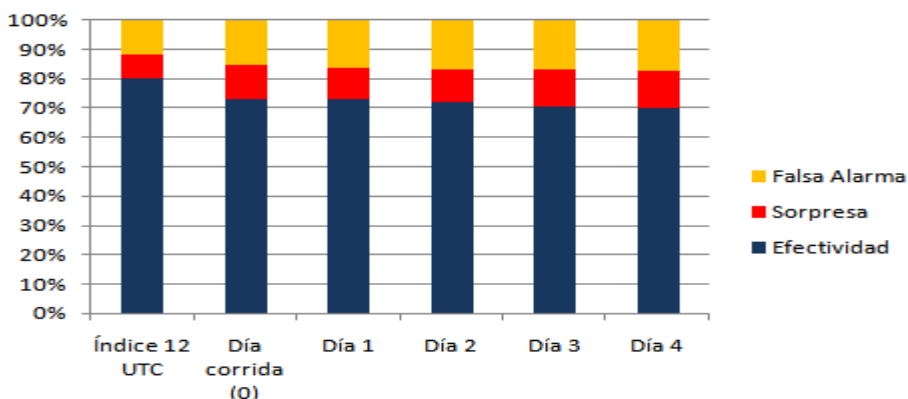


Figura 4.4.2: Valores de efectividad, sorpresa y falsa alarma junto con las curvas ajustadas correspondientemente en función de los diferentes criterios de ocurrencia de convección para el día de la corrida del modelo ETA y para los cuatros días posteriores.

De aquí se obtiene, como se muestra en la Tabla 4.1.I, que los máximos valores de efectividad para el día de la corrida y los 4 subsiguientes varían entre 73,19% y 69,91%. A su vez, dada la asimetría de las distribuciones de los valores del índice para los casos C y NC, los valores de efectividad pueden aumentar levemente disminuyendo el valor de corte a alrededor de 0,3-0,4. En este caso, las efectividades ajustadas rondan el 71 a 73%.

Por lo tanto, comparando las efectividades máximas ajustadas al tomar como variable de entrada el sondeo real y los simulados por el Eta-PRM, podemos concluir que la reducción en la eficiencia del pronóstico debido a la utilización de sondeos simulados varía desde alrededor de 8,5% ( $=|73,2-80|/80*100$ ) para el día de la corrida, a aproximadamente un 12,6% ( $=|69,91-80|/80*100$ ) para el Día 4; o sea, una reducción media de la efectividad ajustada de alrededor del 10% al reemplazar el sondeo real por un sondeo simulado.

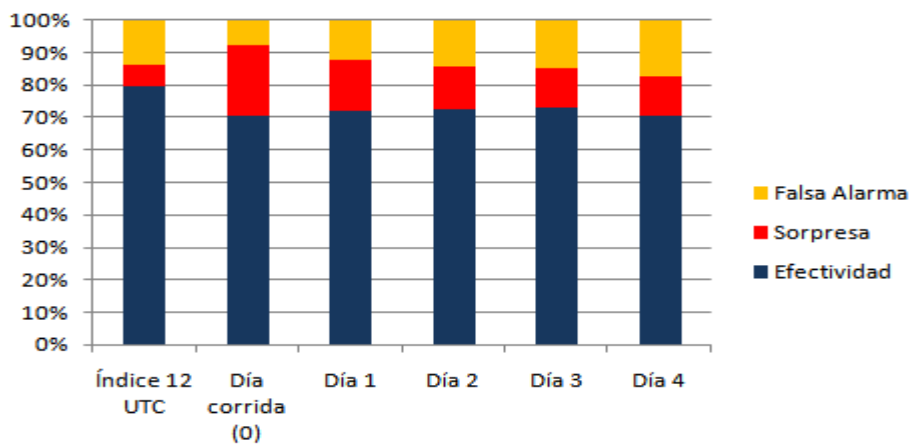
En la Figura 4.4.3 se resume gráficamente este análisis, mostrando los valores de efectividad máxima ajustada, sorpresa y falsa alarma para cada día junto con los criterios de ocurrencia de convección correspondientes y además en la Figura 4.4.4 se muestra un gráfico semejante pero para criterios de C-NC tomando 0,5 como valor de corte.



	Índice 12 UTC	Día corrida (0)	Día 1	Día 2	Día 3	Día 4
Efectividad	80	73,19	73,13	72,13	70,74	69,91
Sorpresa	8,27	11,68	10,61	11,19	12,54	12,61
Falsa Alarma	11,7	15,12	16,25	16,67	16,7	17,46
Criterio de Ocurrencia de Convección	0,5768	0,3058	0,3888	0,4456	0,475	0,5005

Figura 4.4.3: valores de efectividad máxima ajustada, sorpresa y falsa alarma para el índice 12 UTC(día del radio-sondeo), para el día de la corrida y para los cuatro días posteriores junto con los correspondientes criterios de ocurrencia de convección.





	Índice 12 UTC	Día corrida (0)	Día 1	Día 2	Día 3	Día 4
Efectividad	79,86	70,31	72,26	72,4	73,13	70,6
Sorpresa	6,59	22,22	15,56	13,33	12,26	12,28
Falsa Alarma	13,54	7,45	12,17	14,26	14,59	17,1
Criterio de Ocurrencia de Convección	0,5	0,5	0,5	0,5	0,5	0,5

Figura 4.4.4: valores de efectividad (Criterio de C-NC con valor 0,5), sorpresa y falsa alarma para el índice 12 UTC(día del radio-sondeo),para el día de la corrida y para los cuatro días posteriores junto con los correspondientes criterios de ocurrencia de convección.

## 5. Conclusiones

Mediante el Análisis de Componentes Principales se han determinado los principales patrones que caracterizan los perfiles verticales de  $T$  y  $Td$  en la provincia de Mendoza, para los niveles estándar de 925 a 300 hPa. Las primeras 6 componentes explican el 94% de la variabilidad, y los patrones determinados se pueden asociar a situaciones reales tanto en sus modos directos como inversos.

Las series temporales de peso asociadas a las componentes han sido correlacionadas por regresión múltiple logística con las observaciones de presencia de nubes convectivas, a fin de ajustar un modelo estadístico de pronóstico de inestabilidad a corto plazo, mediante el cual es posible calcular un índice de previsión de ocurrencia de convección a partir de los datos de radio-sondeo para los niveles estándar, válido para las 24 horas siguientes a la medición. Adicionalmente, este análisis permite determinar las condiciones atmosféricas más propicias para la ocurrencia de actividad convectiva, las cuales se caracterizarían por una atmósfera muy caliente en capas bajas y muy fría y húmeda (probablemente saturada) en capas medias y altas, favoreciendo la inestabilidad. La situación inversa caracterizaría una atmósfera estable en la cual la convección se vería inhibida. Si bien estas técnicas habían sido implementadas previamente en trabajos anteriores con resultados satisfactorios, la incorporación de más y mejores datos para los análisis y ajustes correspondientes ha permitido la obtención de un modelo con mejores niveles de eficacia y menores errores.

Debido al incremento del número de niveles de observación, sobre todo en capas bajas, respecto de los utilizados en trabajos previos (ver Araneo et al., 2011) y la calidad de los datos utilizados para el ajuste del modelo estadístico, se ha visto que la eficiencia del modelo aumenta, disminuyendo al mismo tiempo los niveles de error, pasando de un máximo de 73% de efectividad a un máximo ajustado de 80%. Las distribuciones de frecuencias de valores del índice para casos con observación de convección y no convección, se ajustan significativamente (al 5% de confianza) a una distribución de probabilidades Gamma, a partir de las cuales puede determinarse la efectividad teórica del modelo, que asciende a casi 83%, con probabilidades de error por falsa alarma de 13% y por sorpresa de 4%, para un criterio de pronóstico de ocurrencia basado en un valor de corte de 0,5 (i.e. con un valor de índice superior a 0,5 se pronostica ocurrencia y con uno inferior se pronostica no ocurrencia). Sin embargo, la asimetría de estas distribuciones, debido a la asimetría en la cantidad de datos con observaciones reales de convección respecto a la cantidad de datos con no convección, hace que, variando el valor de corte en el criterio de ocurrencia, la eficiencia del modelo pueda aumentarse a más del 83%.

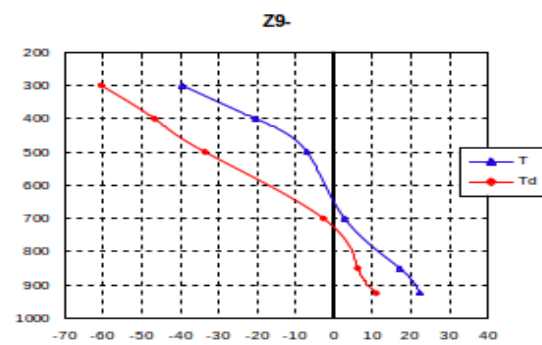
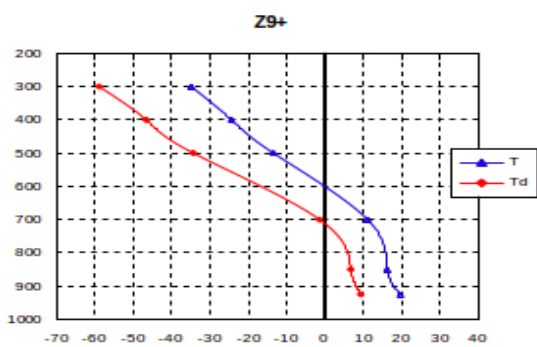
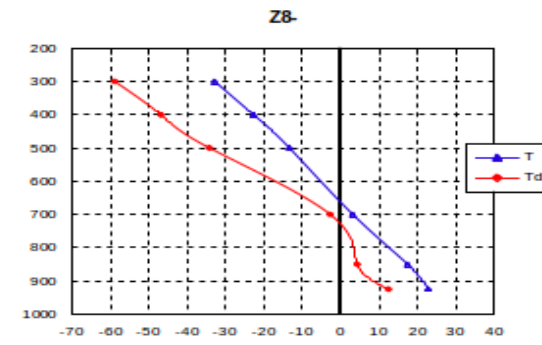
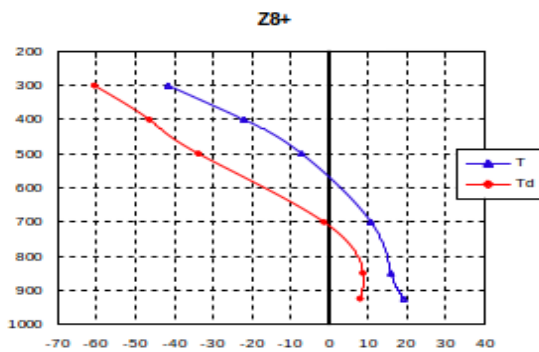
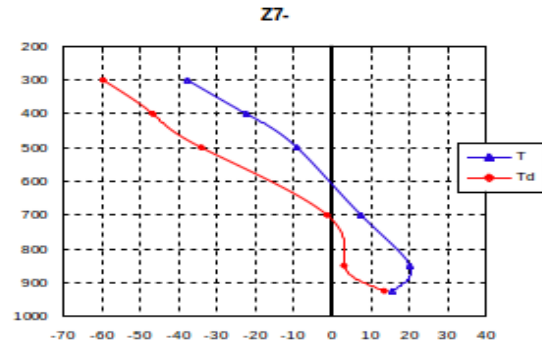
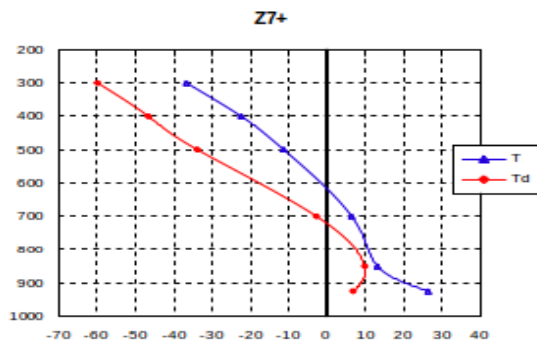
Por otro lado, se ha visto que las variables dinámicas y energéticas como la helicidad, la cortante vertical del viento en distintos niveles y el CAPE no son factores determinantes en las previsiones del inicio de la convección para la región del oasis norte de la provincia de Mendoza, ya que ni los factores dinámicos ni los energéticos mostraron inferencia significativa sobre la eficiencia del modelo. Además se ha visto que ni la implementación del método de rotación VARIMAX ni la variación en el rango temporal de validez del pronóstico (definición del rango temporal de ocurrencia de convección) mejoran la eficiencia del modelo estadístico. En consecuencia se observa que la ocurrencia de convección para la región estudiada depende primordialmente de las condiciones termodinámicas del entorno caracterizadas por los perfiles verticales de  $T$  y  $Td$ .

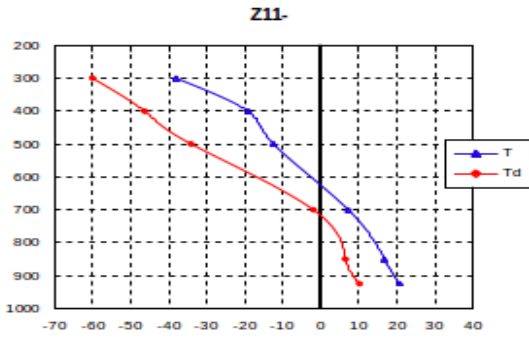
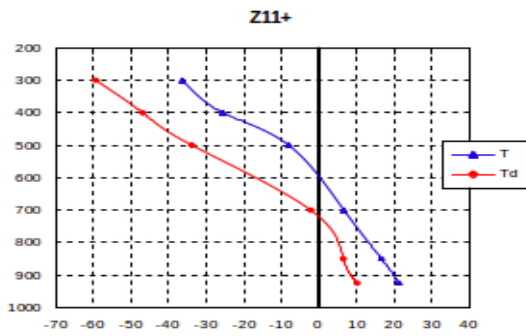
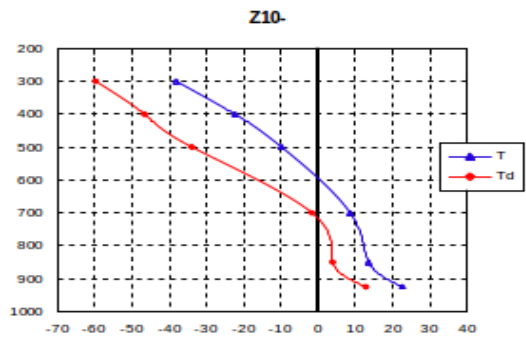
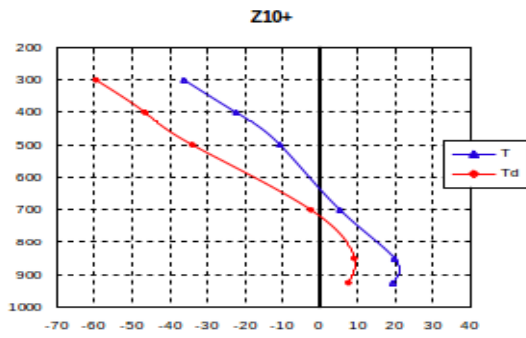
Por otra parte, a través de la combinación del modelo estadístico con el modelo numérico de pronóstico Eta-PRM, se ha visto que para el día de la corrida el modelo muestra una disminución del porcentaje de error por falsa alarma en comparación con el caso de sondeos reales (índice 12 UTC mostrado en la sección 4.2) bajo los mismos criterios de ocurrencia (por ejemplo con un valor de corte de 0,5). Sin embargo, los pronósticos efectuados en días con observación real de convección (casos C) muestran una exactitud menor que en el caso de utilización de sondeos reales (no simulados), lo cual implica un aumento considerable en el porcentaje de error por sorpresa o imprevistos con respecto al error obtenido para el índice 12 UTC con iguales criterios de ocurrencia. Para los primeros dos días con pronóstico basado en las salidas del Eta-PRM (i.e. previsiones a 48 y 72 hs.), se observa un muy buen desempeño en la habilidad del modelo de pronosticar casos de C, aunque no tan bueno para los casos de NC. Hay una disminución en la habilidad del modelo de pronosticar casos de NC y disminuye aún más para el caso de sondeos modelados correspondientes a los últimos dos días de la corrida (i. e. previsiones a 96 y 120 hs.), con un aumento en los porcentajes de error por falsa alarma para un mismo criterio de ocurrencia. Adicionalmente, a pesar de que la utilización de sondeos simulados por el Eta-PRM como entrada en modelo estadístico implica una disminución en la eficiencia del pronóstico resultante de alrededor del 10%, la efectividad del pronóstico no disminuye demasiado para las previsiones de 48 a 129 hs., manteniéndose en alrededor del 70 a 73% para los 4 días de la corrida.

## 6. Apéndices

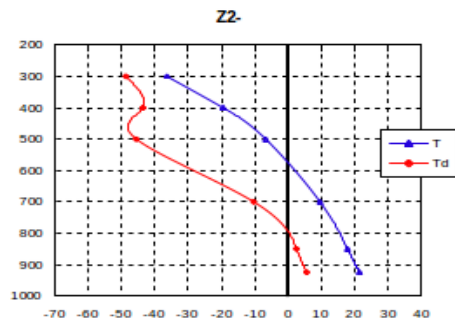
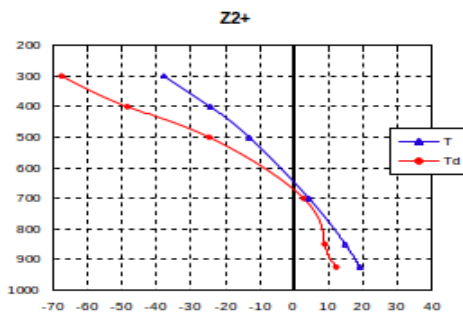
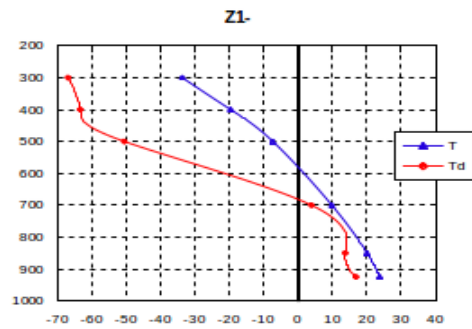
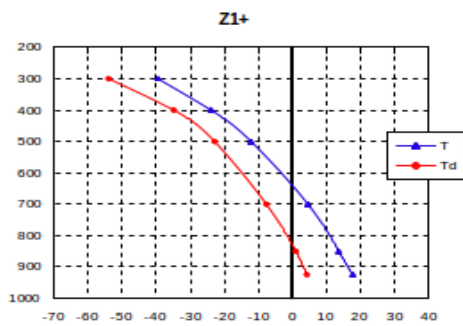
### Apéndice A

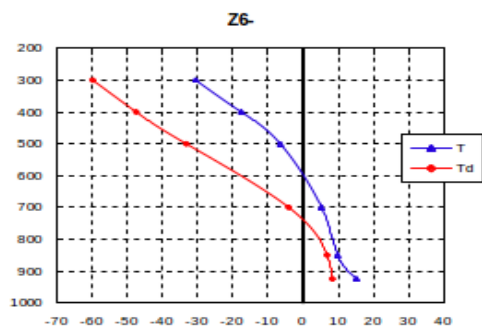
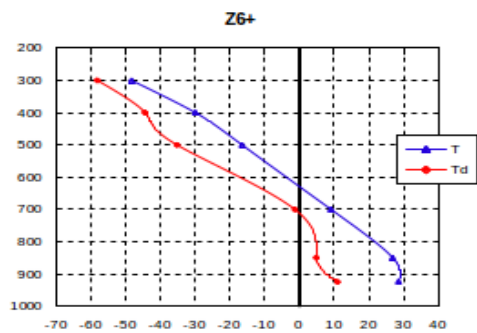
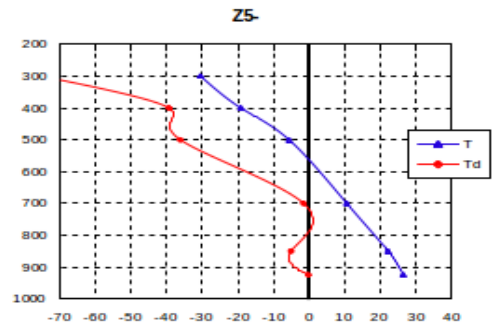
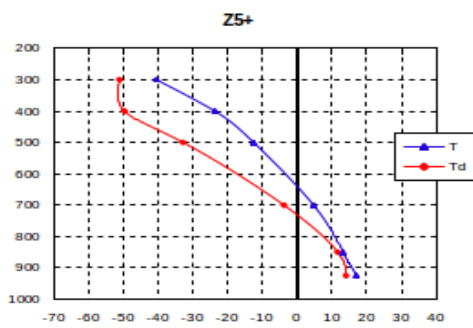
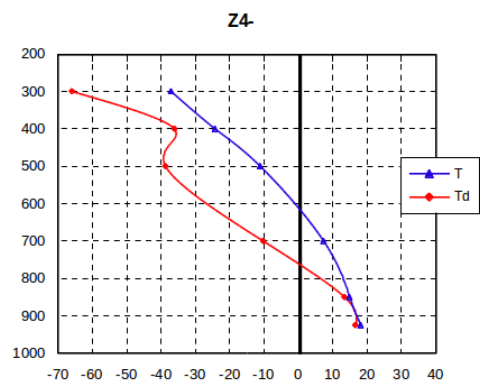
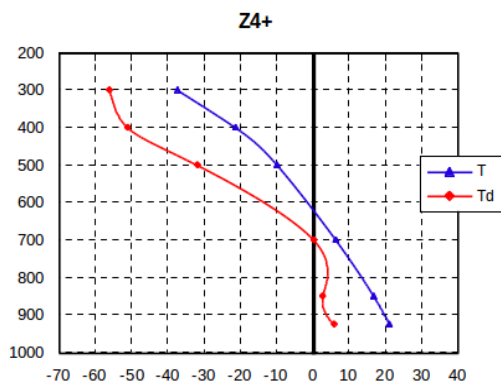
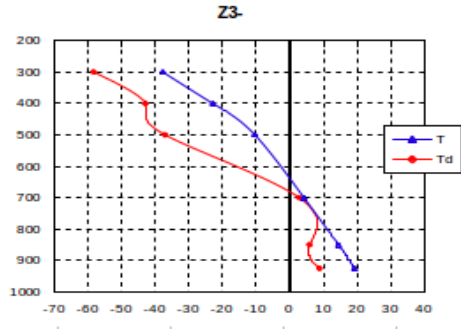
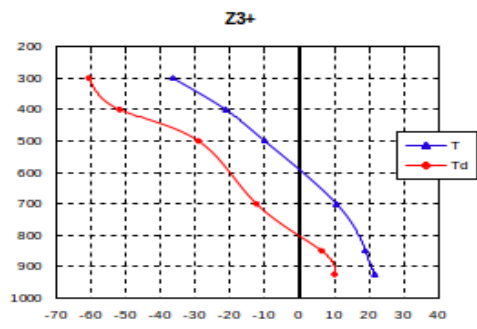
Componentes principales de la 7 a la 11 en sus modos directo e inverso correspondientes al ajuste con 320 sondeos.

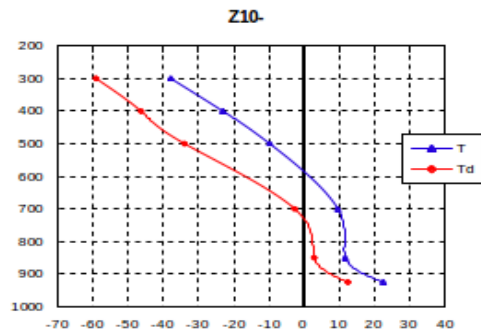
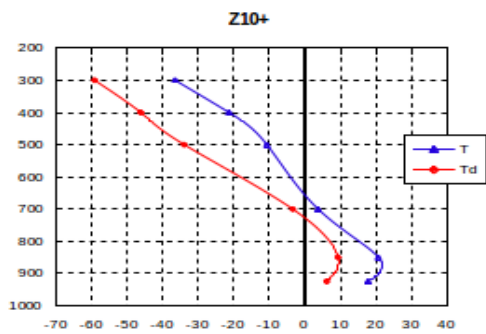
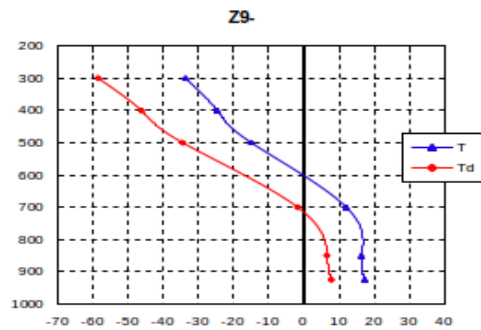
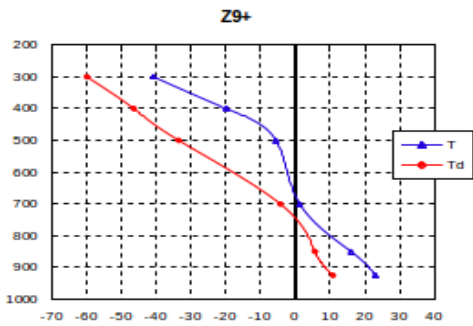
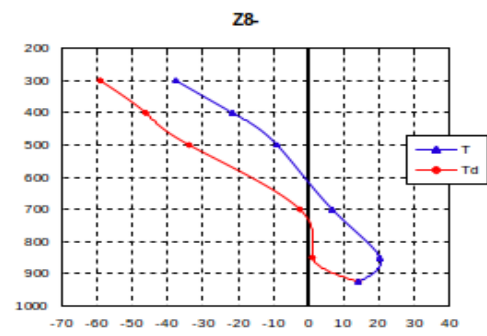
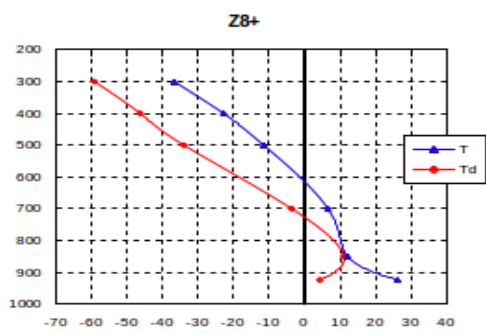
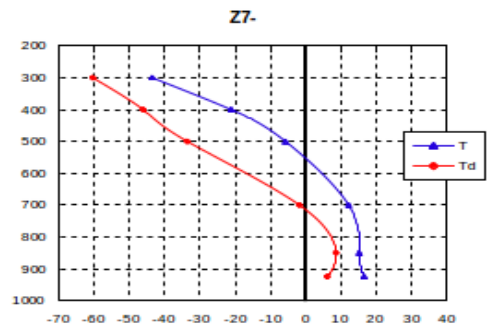
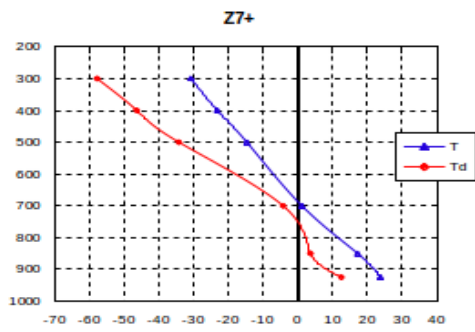


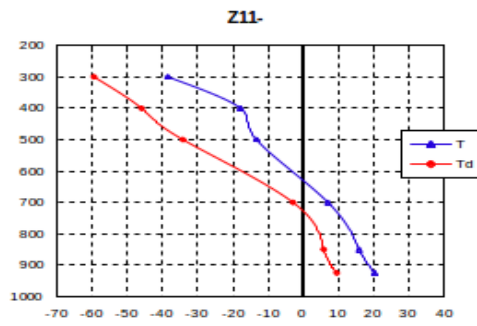
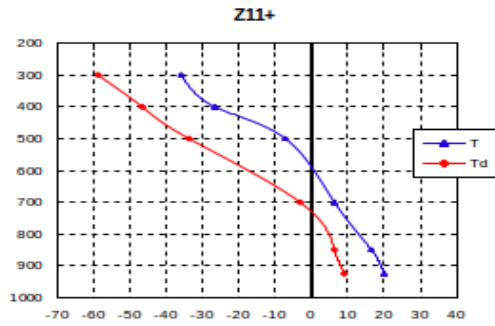


Se muestran a continuación las componentes principales de la 1 a la 11 en sus modos directo e inverso correspondientes al ajuste con 659 sondeos (las primeras 6 componentes se muestran para que se pueda corroborar que son las mismas que para el ajuste con 320 sondeos).









## Apéndice B

### *Día de Corrida (Día 0)*

Tabla de frecuencias correspondiente a 657 sondeos separados en 390 eventos en los que efectivamente se observó convección (conjunto C) y 267 en los que no se observó (conjunto NC).

Observaciones	0	1
0 a 0,1	74	13
0,1 a 0,2	60	31
0,2 a 0,3	37	28
0,3 a 0,4	26	34
0,4 a 0,5	21	40
0,5 a 0,6	13	39
0,6 a 0,7	13	47
0,7 a 0,8	14	45
0,8 a 0,9	5	57
0,9 a 1	4	56

### *Día 1*

Tabla de frecuencias correspondiente a 649 sondeos separados en 383 eventos en los que efectivamente se observó convección (conjunto C) y 266 en los que no se observó (conjunto NC).

Observaciones	0	1
0 a 0,1	51	7
0,1 a 0,2	62	14
0,2 a 0,3	33	22
0,3 a 0,4	26	27
0,4 a 0,5	15	31
0,5 a 0,6	19	27
0,6 a 0,7	12	35
0,7 a 0,8	22	40
0,8 a 0,9	9	55
0,9 a 1	17	125



### Día 2

Tabla de frecuencias correspondiente a 645 sondeos separados en 380 eventos en los que efectivamente se observó convección (conjunto C) y 265 en los que no se observó (conjunto NC).

Observaciones	0	1
0 a 0,1	35	7
0,1 a 0,2	59	15
0,2 a 0,3	40	17
0,3 a 0,4	23	23
0,4 a 0,5	16	24
0,5 a 0,6	20	22
0,6 a 0,7	15	34
0,7 a 0,8	16	39
0,8 a 0,9	17	68
0,9 a 1	24	131

### Día 3

Tabla de frecuencias correspondiente a 644 sondeos separados en 382 eventos en los que efectivamente se observó convección (conjunto C) y 262 en los que no se observó (conjunto NC).

Observaciones	0	1
0 a 0,1	22	7
0,1 a 0,2	50	21
0,2 a 0,3	42	15
0,3 a 0,4	27	16
0,4 a 0,5	27	20
0,5 a 0,6	15	25
0,6 a 0,7	16	32
0,7 a 0,8	13	41
0,8 a 0,9	25	66
0,9 a 1	25	139

### Día 4

Tabla de frecuencias correspondiente a 643 sondeos separados en 381 eventos en los que efectivamente se observó convección (conjunto C) y 262 en los que no se observó (conjunto NC).

Observaciones	0	1
0 a 0,1	20	8
0,1 a 0,2	43	9
0,2 a 0,3	38	22
0,3 a 0,4	18	18
0,4 a 0,5	33	22
0,5 a 0,6	21	29
0,6 a 0,7	24	23
0,7 a 0,8	18	46
0,8 a 0,9	21	58
0,9 a 1	26	146

## Apéndice C

### Día de Corrida (Día 0)

La curva que mejor ajusta los valores de la efectividad está dada por

$$f(x) = -0,00020x^2 + 0,10047x + 60,76098$$

y la curva que mejor ajusta los valores de la sorpresa está dada por

$$f(x) = 0,001034 x^{1,697640} e^{-0,000083 x}$$

### Día 1

La curva que mejor ajusta los valores de la efectividad está dada por

$$f(x) = -0,00020x^2 + 0,09537x + 61,85431$$

y la curva que mejor ajusta los valores de la sorpresa está dada por

$$f(x) = 0,000219 x^{2,008045} e^{-0,000776 x}$$

### Día 2

La curva que mejor ajusta los valores de la efectividad está dada por

$$f(x) = -0,00020x^2 + 0,09283x + 61,17287$$

y la curva que mejor ajusta los valores de la sorpresa está dada por

$$f(x) = 0,000350 x^{1,928605} e^{-0,000701 x}$$

### Día 3

La curva que mejor ajusta los valores de la efectividad está dada por

$$f(x) = -0,00019x^2 + 0,09055x + 60,00196$$

y la curva que mejor ajusta los valores de la sorpresa está dada por

$$f(x) = 0,003321 x^{1,500396} e^{0,000138 x}$$

#### Día 4

La curva que mejor ajusta los valores de la efectividad está dada por

$$f(x) = -0,00018x^2 + 0,08366x + 60,20415$$

y la curva que mejor ajusta los valores de la sorpresa está dada por

$$f(x) = 0,005486 x^{1,409244} e^{0,000280 x}$$

## 7. Bibliografía

Araneo D. C., Simonelli S. C., Norte F. A., Viale M. y Santos J. R. (2011): “Caracterización de sondeos estivales del norte de Mendoza mediante el Análisis de Componentes Principales y obtención de un índice de convección”. METEOROLÓGICA, ISSN: 0325-187X, Vol. 36, No 1, pp. 31-47.

Deming, J., Chaohua, D., Ruili, H. y Ronghui, C., 2008. Impact assessment of additional predictors to the retrieval of atmospheric profiles from infrared radiances. Proceedings of SPIE - The International Society for Optical Engineering, Art. no 71520D.

Duarte, M.L., 1993. Temporal variations of the vertical profile of temperatures of Argentina. Int. J. Climatol., 13, 437-445.

García L., F. (2009) Pronóstico de ocurrencia de lluvia usando Índices de Estabilidad Atmosférica y Cadenas de Markov. Electrificación del Caroní. 39 p.

Informe Ambiental, 2009. Secretaría de Medio Ambiente, Gobierno de Mendoza.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y. Leetmaa, A. y Reynolds, R., 1996. The NCEP / NCAR 40-year Reanalysis Project. Bull. Amer. Meteor. Soc., 77, 437-441.

Lattin, J., Carroll, J.D. y Green, P.E., 2003. Analyzing Multivariate Data. Duxbury Applied Series, Thompson Brooks/Cole. 556 págs.

Markowski P. and Richardson Y. 2010. Mesoscale Meteorology in Midlatitudes. Penn State University, University Park, PA, USA. 414 pp.

Martínez Carretero, E. y Dalmaso A.D. (Editores), 1995. Mendoza Ambiental. IADIZA – Ministerio de Medio Ambiente, Urbanismo y Vivienda. 292 págs.

Moyano, M.C., Blanco, V.A. y Casinelli, E.J., 1972. Análisis comparativo de índices de inestabilidad como predictores de fenómenos convectivos para las estaciones Resistencia, Córdoba y Ezeiza. Meteorologica, 3, No 1-2-3, 135-156.

Nicolini, M.C. y Torres Brizuela, M. 1997. Estadística de vientos fuertes asociados a convección en Ezeiza y Resistencia y estudio numérico de los forzantes en un caso real. *Meteorologica*, 22, No 2, 19-35.

Nicolini, M. y Norte, F., 1978. Análisis en tiempo diferido de las situaciones sinópticas correspondientes a los días de tormenta convectiva en la Provincia de Mendoza. Informe de la temporada granicera 1976-77, 26-43.

Norte, F.A., 1982. Evaluación de índices de inestabilidad y parámetros sinópticos como predictores de tiempo convectivo en el norte de Mendoza. *Geoacta*, 11, No 2, 193-206.

Norte F., 1980. Clasificación sinóptica de las tormentas convectivas y su relación con las tareas operativas en el área de Defensa. Informe de las primeras experiencias de ajuste de defensa. Temporada granicera 1978-1979. Tomo 2, 51-95. Programa Nacional de Lucha Antigranizo. Comisión Nacional de Investigaciones Espaciales.

North G. R. and Erukhimova T. L., 2009. Atmospheric thermodynamics. Cambridge University Press. The Edinburgh Building, Cambridge CB2 8RU, UK. 281 pp.

Rosso, L.C., Quinteros, C.Y., de Candurra, R. y de Lázaro S.Q., 1987. Índices de inestabilidad como predictores de ocurrencia de precipitaciones. Anales II Congreso Interamericano de Meteorología y V Congreso Argentino de Meteorología, 3.2.1-3.2.4

Saluzzi, M.E., 1983. Aspectos físicos de la convección severa en Mendoza. Tesis Doctoral, Universidad de Buenos Aires, 115 págs.

Saluzzi, M.E., Nuñez J.M. y Perez, L., 1979. Las grandes nubes convectivas a sotavento de los Andes, sus características y su comportamiento. *Geoacta*, 9, N° 1, 35-45.

Servicio Meteorológico Nacional, 2009. Estadísticas Climatológicas.

Simonelli, S.C., 2000. Modelo estadístico de pronóstico de convección para la zona norte de la provincia de Mendoza. Tesis Doctoral, FCEyN, UBA, 158 págs.

Simonelli S. C., Araneo D. C., Norte F. A., Viale M. y Santos J. R. (2010): "Validación de los perfiles verticales de Temperatura, Humedad y Viento obtenidos a partir del modelo ETA-PRM para las localidades de Mendoza (Argentina) y Santo Domingo (Chile)". *XXV Reunión Científica de la Asociación Argentina de Geofísicos y Geodestas, AAGG '10. 2 al 5 de Noviembre. Córdoba, Argentina.*

Smith, W.L. y Woolf, H.M., 1976. The use of Eigenvectors of Statistical Covariance Matrices for Interpreting Satellite Sounding Radiometer Observations. *J. Atmos. Sci.*, 33, 1127-1140.

Wallace J. M. and Hobbs P. V. 2006. Atmospheric Science. An Introductory Survey. Academic Press, Elsevier, Burlington, USA. 505 pp.

Wang, J., Cole, H.L. y Carlson, D.J., 2001. Water Vapor Variability in the Tropical Western Pacific from 20-year Radiosonde Data. *Advances in Atmospheric Sciences*, 18, No 5, 752-766.