

Informe Entregables Proyecto AlertAr

Grupo III/2017-10

Técnicas de Clasificación Supervisada para la Discriminación entre Ecos Meteorológicos y No Meteorológicos usando Información de un Radar Meteorológico de Banda C

Integrantes:

- Sofia Ruiz Suarez – Servicio Meteorológico Nacional
- Luciano Vidal – Servicio Meteorológico Nacional
- Paola Salio – CIMA –CONICET/UBA
- Yanina Garcia Skabar – Servicio Meteorológico Nacional/CONICET

Octubre, 2017

Contenido

Resumen	3
Abstract	3
Detalle del Informe	4
1. Introducción	5
2. Metodología	5
3. Conclusiones	5
4. Referencias	5

Resumen	Abstract
<p>Los datos provenientes de los radares meteorológicos son de suma importancia para el diagnóstico y monitoreo de los sistemas que producen precipitación y sus posibles fenómenos severos asociados. Los ecos causados por objetivos no meteorológicos introducen errores en la información por lo que es necesario detectar la presencia de los mismos previo a la utilización de los datos. Este trabajo presenta cuatro técnicas de clasificación supervisada basadas en diferentes modelos estadísticos que buscan dar una respuesta a este problema.</p> <p>Asimismo como parte importante de este trabajo, se aplicaron técnicas de remuestreo estadísticas sobre los datos de entrenamiento, las que permitieron hacer un análisis más completo sobre los resultados. En la actualidad, las técnicas de remuestreo son herramientas fundamentales en la estadística moderna. Las mismas, a partir de simulaciones sobre los datos, permiten obtener información adicional sobre los modelos planteados. Para este trabajo se realizó un estudio de caso con datos provenientes del radar meteorológico doppler banda C de doble polarización ubicado en la Estación Experimental Agropecuaria INTA Anguil (La Pampa). Partiendo de la clasificación manual de un experto, se aplicaron cuatro métodos de clasificación supervisada de diferentes grados de flexibilidad en su estructura: Modelo lineal, Modelo Cuadrático, Modelo Logístico y Modelo de Bayes Naive. Luego se compararon los resultados y se evaluó el desempeño de cada uno de ellos. Si bien se encontraron dificultades a la hora de clasificar las zonas de frontera entre clases, los resultados obtenidos fueron adecuados, mostrando el mejor desempeño el modelo menos flexible, el modelo lineal. Se considera necesario seguir avanzando en esta línea de investigación a fin de incorporar una mayor cantidad de casos y tener una mayor significancia de los resultados.</p>	<p>Data coming from meteorological radars is of the utmost importance for the diagnosis and monitoring of precipitation systems and their possible associated severe phenomena. The echoes caused by objectives that are not meteorological introduce errors in the information. Therefore, it is necessary to detect their presence before using this data. This paper presents four supervised classification techniques based on different models which seek to give an answer to this problem.</p> <p>In addition, as an important part of this work, resampling techniques were implemented on the training set in order to further assess the results. Resampling methods are an indispensable tool in modern statistics. Those techniques provide additional information about the model of interest by repeatedly drawing samples from the data.</p> <p>Based on data from the C-band doppler radar located in Anguil and from a previous expert's manual classification, four supervised classification methods with different degrees of flexibility in their structure were implemented: Linear Model, Quadratic Model, Logistic Model and Bayes Naive Model. Finally, the results of each of them were assessed and compared. Although difficulties were encountered in classifying boundary zones between classes, the results obtained were adequate, showing the best performance in the least flexible model, the linear one. It is considered necessary to keep working in this line of research in order to include more cases in the analysis and allow a better inference on the results.</p>

Detalle del Informe	
1. Introducción	...
2. Metodología	...
3. Conclusiones	...
4. Referencias	...

NOTA: A continuación se adjunta trabajo a publicar en la revista **METEOROLOGICA** del Centro Argentino de Meteorólogos.

Ruiz Suarez, S., M. Sued, L. Vidal, P. Salio, D. Rodriguez, S. Nesbitt y Y. Garcia Skabar, 2017: Técnicas de Clasificación Supervisada para la Discriminación entre Ecos Meteorológicos y No Meteorológicos usando Información de un Radar Meteorológico de Banda C. Meteorológica. En revisión.



Meteoro logica

31 Asimismo como parte importante de este trabajo, se aplicaron técnicas de remuestreo
32 estadísticas sobre los datos de entrenamiento, las que permitieron hacer un análisis más
33 completo sobre los resultados. En la actualidad, las técnicas de remuestreo son
34 herramientas fundamentales en la estadística moderna. Las mismas, a partir de
35 simulaciones sobre los datos, permiten obtener información adicional sobre los modelos
36 planteados.

37 Para este trabajo se realizó un estudio de caso con datos provenientes del radar
38 meteorológico Doppler banda C de doble polarización ubicado en la Estación
39 Experimental Agropecuaria INTA Anguil (La Pampa). Partiendo de la clasificación
40 manual de un experto, se aplicaron cuatro métodos de clasificación supervisada de
41 diferentes grados de flexibilidad en su estructura: Modelo lineal, Modelo Cuadrático,
42 Modelo Logístico y Modelo de Bayes Naive. Luego se compararon los resultados y se
43 evaluó el desempeño de cada uno de ellos. Si bien se encontraron dificultades a la hora
44 de clasificar las zonas de frontera entre clases, los resultados obtenidos fueron
45 adecuados, mostrando el mejor desempeño el modelo menos flexible, el modelo lineal.
46 Se considera necesario seguir avanzando en esta línea de investigación a fin de
47 incorporar una mayor cantidad de casos y tener una mayor significancia de los
48 resultados.

49

50 **Palabras clave:** clasificación supervisada, remuestreo, eco no meteorológico, radar
51 meteorológico.



Meteorología

82 Los radares meteorológicos son una herramienta fundamental para el diagnóstico y
83 monitoreo de los sistemas precipitantes y los posibles fenómenos meteorológicos
84 severos asociados como granizo, ráfagas destructivas de viento e incluso tornados.

85 Para poder utilizar la información obtenida a partir de los radares meteorológicos es
86 necesario en primera instancia aplicar metodologías de control de calidad tendientes a
87 minimizar las fuentes de error (Zawadzki, 1984) presentes en los datos con el objetivo
88 de poder emplear los mismos rutinariamente en aplicaciones relacionadas con
89 estimaciones cuantitativas de precipitación, detección de granizo, elaboración de
90 pronósticos a muy corto plazo, asimilación de datos en modelos numéricos de
91 pronóstico del tiempo, diagnóstico de tiempo severo, entre otras aplicaciones. Uno de
92 los problemas más recurrentes en los datos de radar meteorológico, es la presencia de
93 ecos no meteorológicos tales como ecos biológicos (insectos, pájaros), ecos de terreno
94 (edificios, montañas), o incluso presencia de propagación anómala (Rico-Ramirez y
95 Cluckie, 2008; Berenguer y otros, 2006).

96 Esta problemática ha sido abordada por varios autores a lo largo de las últimas décadas
97 de diferentes maneras. Se han desarrollado distintas técnicas, las cuales pueden dividirse
98 en las que implementan filtro directamente en el procesador de señal I\Q del radar
99 (Siggia y Passarelli, 2004) y aquellas que lo hacen sobre los momentos obtenidos
100 (reflectividad, doppler, coeficiente de correlación, entre otras variables). Sobre estas
101 últimas a su vez, se pueden distinguir tres tipos de abordajes: las que utilizan técnicas de
102 árboles de decisión (Stein y Smith, 2001), las que lo hacen en base a redes neuronales
103 (Lakshmanan y otros, 2010; Greku y Krajewski, 2000), y las que utilizan las llamadas
104 técnicas de lógica difusa (Cho y otros, 2006; Gourley y otros, 2006; Hubbert y otros,
105 2009; Bo Young Ye y otros, 2015; Berenguer y otros, 2006; Rico-Ramírez y Cluckie,
106 2008; entre otros). Es posible notar en los últimos años una leve tendencia a utilizar
107 preferentemente los modelos basados en lógica difusa. En su mayoría los trabajos
108 realizados hacen un análisis pixel por pixel y se basan, según la frecuencia
109 electromagnética de observación utilizada por el radar y las variables disponibles, en



Meteoro logica

110 distinguir comportamiento de la información en dos clases: meteorológica y no
111 meteorológica.

112 Numerosos autores, entre ellos Moszkowicz y otros (1993), Berenguer y otros (2006),
113 Rico-Ramirez y Cluckie (2008), Gourley y otros (2006) buscan estimar para cada pixel
114 la probabilidad de pertenecer a la clase “eco no meteorológico” y proponen
115 clasificadores basados en la regla de clasificación de Bayes. En particular, Moszkowicz
116 y otros (1993) a fin de identificar la presencia de ecos de terreno asociados a
117 propagación anómala propusieron un método de clasificación basado en la regla de
118 Bayes tomando como campos de entrada en su modelo la elevación de la estrategia de
119 escaneo con mayor valor de reflectividad, el valor de ese máximo y el gradiente
120 horizontal de la reflectividad, entre otros. Los autores suponen que las funciones de
121 probabilidad condicional de cada campo sujeto a cada tipo de eco son gaussianas y
122 luego estiman los parámetros de dichas distribuciones con datos previamente
123 seleccionados y clasificados de forma manual por un experto.

124 Más recientemente Berenguer y otros (2006) presentan un trabajo focalizado hacia la
125 eliminación de propagación anómala. Para cada pixel evalúan la posibilidad de que la
126 medición haya sido afectada por dicho fenómeno asociando un valor entre 0 y 1. Con
127 este objetivo, analizan las distribuciones de las frecuencias de cada campo estudiado
128 condicional al tipo de eco. A partir de la regla de clasificación de Bayes, derivan la
129 probabilidad condicional de cada pixel de estar afectado según el valor de los campos.
130 Propone dos configuraciones distintas del algoritmo dependiendo si se trata de zonas
131 cercanas al mar o no.

132 Estudios previos han demostrado la gran utilidad de disponer de información
133 polarimétrica para identificar áreas con presencia de ecos no meteorológicos. Trabajos
134 como Schur y otros (2003) y Cho y otros (2006) presentan clasificadores basados en
135 técnicas de lógica difusa para radares de doble polarización. Los autores definen las
136 variables predictoras a partir del comportamiento de las variables polarimétricas tanto a
137 nivel espacial como temporal en los distintos tipos de ecos.



Meteoro logica

138 En el estudio presentado por Gourley y otros (2006) también se trabaja sobre datos
139 provenientes de radares de doble polarización con técnicas de lógica difusa. En este
140 caso los autores también tienen en cuenta el comportamiento de la velocidad radial y de
141 la continuidad en el campo de la reflectividad horizontal. Sugieren considerar funciones
142 de pertenencia derivadas a partir de una estimación no paramétrica de la densidad de
143 cada campo en cada eco.

144 Rico-Ramirez y Cluckie (2008) presentan un clasificador en donde la Regla de Bayes y
145 la lógica difusa son protagonistas. Trabajan con radares de banda C de doble
146 polarización. Este modelo se diferencia de los algoritmos presentados por Gourley y
147 otros (2006) y por Schur y otros (2003) en cuanto a la forma de calcular la textura de las
148 variables.

149 A diferencia de muchos de los trabajos previamente mencionados, el presente trabajo
150 busca abordar la problemática de la diferenciación entre tipos de ecos a partir de
151 técnicas de estadística clásica que modelan el problema según las leyes de la
152 probabilidad. Según Kosko (1994) la lógica difusa y la probabilidad difieren tanto en lo
153 conceptual como en lo teórico, pero a la vez coinciden en varios puntos. Ambos
154 sistemas combinan conjuntos y proposiciones de manera asociativa, distributiva y
155 conmutativa, y a la vez describen incertidumbres a partir de una cierta cantidad
156 perteneciente al intervalo $[0,1]$. La diferencia entre ambos enfoques radica en cómo son
157 considerados los conjuntos y sus complementos. Para el enfoque clásico la intersección
158 entre un conjunto y su complemento es vacía, por lo tanto la probabilidad de que esto
159 ocurra es cero. En cambio, en la teoría de la lógica difusa esto no siempre es verdadero,
160 es decir podría pasar que existiera algún elemento en la intersección entre un conjunto y
161 su complemento.

162 Actualmente en Argentina se cuenta con radares Doppler de doble polarización banda C
163 (frecuencia de 5.6 GHz) que funcionan de forma operativa y son utilizados a diario por
164 el Servicio Meteorológico Nacional. Además, se está llevando a cabo el proyecto
165 SiNaRaMe (Sistema Nacional de Radares Meteorológicos) que tiene como fin expandir
166 la red actual de radares a partir de la incorporación de instrumentos de fabricación

167 nacional. Este crecimiento vertiginoso de la disponibilidad de datos de radar en nuestro
168 país implica la necesidad de avanzar en la implementación de metodologías de control
169 de calidad de la información generada para poder contar con datos más precisos que
170 puedan ser utilizados en las distintas aplicaciones.

171 El objetivo del presente trabajo es introducir y evaluar el desempeño de cuatro métodos
172 de clasificación supervisada a fin de mejorar la diferenciación entre ecos
173 meteorológicos y no meteorológicos como un primer paso dentro del desarrollo de un
174 sistema de control de calidad de la información de radar a ser implementada por el
175 Servicio Meteorológico Nacional. Los métodos presentados están basados en la Regla
176 de clasificación de Bayes. Este trabajo propone hacer un análisis del comportamiento de
177 dichos métodos, para luego compararlos e inferir sobre la capacidad discriminante de
178 cada uno de ellos. Para la validación y testeo de los diferentes procedimientos, se
179 utilizaron técnicas estadísticas de remuestreo de manera de poder hacer un análisis más
180 completo de los resultados obtenidos.

181 El trabajo se organiza de la siguiente forma. En la Sección 2 se describen los datos
182 utilizados, en la Sección 3 se presenta la metodología a seguir y se exponen los métodos
183 de clasificación utilizados. La discusión de los resultados obtenidos se expone en la
184 Sección 4. Por último en la Sección 5 se dan las conclusiones del trabajo y se proponen
185 las líneas de trabajo futuro.

186 2) DATOS

187 Se utilizaron datos del radar meteorológico Selex SI Gematronik Doppler de doble
188 polarización en banda C (5.6 GHz) instalado en el predio de la Estación Experimental
189 Agropecuaria INTA Anguil (La Pampa), ubicado en 36° 22' 22,9"S y 63° 58' 58"O
190 (Figura 1). Los datos analizados corresponden a la estrategia de escaneo que genera un
191 volumen de datos cada 10 minutos conformado por un total de 10 elevaciones de antena
192 que varían entre 0,5 y 19,2 grados, con una resolución en rango de 0,25 km y de 1° en
193 azimut (ancho del haz), y un alcance máximo de 120 kilómetros. Se seleccionaron los
194 volúmenes con el mayor tiempo de emisión entre pulsos a fin de tener una velocidad



Meteoro logica

195 nyquist del radar alta, en este caso de 40 metros por segundo para el escaneo
196 seleccionado. Esto se corresponde con un valor de frecuencia de repetición de pulso de
197 1000 y 750 Hz considerando un procesamiento de señal staggered.

198 Para este trabajo de las variables generadas por el radar se consideraron: *coeficiente de*
199 *correlación co-polar* (ρ_{HV}), *velocidad radial Doppler* (V), y *reflectividad diferencial*
200 (Z_{DR}), para la primera y segunda elevación de antena del radar (0,5 y 1,3 grados). Al
201 considerar la velocidad Doppler como variable de entrada es necesario tener en cuenta
202 la posible presencia de *aliasing* en los datos (Battan, 1973). Dado que en los casos
203 trabajados no se detectó este efecto, no se realizaron correcciones previas a los datos.
204 No obstante, si existieran errores por *aliasing* se deberían corregir con el fin de que los
205 resultados sean consistentes.

206 Tal como se explicará en la próxima sección, los algoritmos de clasificación
207 supervisada que aquí se presentan se construyen utilizando una muestra de
208 entrenamiento donde, para cada observación, se dispone del valor de las variables
209 observadas y la categoría a la cual la observación pertenece (en este caso entre eco
210 meteorológico y no meteorológico). Esta información es fundamental para la
211 construcción del procedimiento de clasificación que determinará la clase a la que
212 pertenece una nueva observación en función del valor de las variables disponibles
213 medidas en la nueva observación. Con este objetivo, se analizó un conjunto de casos y
214 se realizó una clasificación manual asistida por un meteorólogo experto en el área. El
215 experto meteorólogo realizó una inspección visual donde se consideró la variabilidad
216 temporal y la estructura espacial de los ecos de radar para determinar la pertenencia a
217 una u otra categoría. Para el caso de ecos meteorológicos, se buscó ecos con un
218 apreciable desarrollo vertical (observable al menos en las tres primeras elevaciones de la
219 antena) en la variable reflectividad horizontal, que al mismo tiempo presenten
220 velocidades Doppler distintas de cero y valores de coeficientes de correlación HV
221 cercanos a la unidad. En contraste, los ecos no meteorológicos no presentan un
222 desarrollo vertical apreciable, ya que mayormente se ubican en el primer kilómetro de la
223 atmósfera y también presentan una señal con una textura espacial no homogénea tanto



Meteoro logica

224 en el caso de ecos biológicos como en ecos originados por propagación anómala. De
225 estos casos, se seleccionaron algunos de ellos para ser utilizados como muestra de
226 entrenamiento, y otros para ser utilizados como datos de testeo.

227 A fin de determinar los casos de entrenamiento se tuvo en consideración que estuvieran
228 disponibles todas las variables antes mencionadas. Además se buscó que estos casos
229 representaran diferentes configuraciones en relación a la distribución espacial de los
230 ecos meteorológicos y no meteorológicos.

231 Se consideraron entonces las siguientes fechas para formar parte de la muestra de
232 entrenamiento:

- 233 ● 20 de junio de 2009
- 234 ● 22 de noviembre de 2009
- 235 ● 1 de enero de 2010
- 236 ● 6, 4, 9, 20, 21 y 27 de febrero 2010

237

238 El número total de píxeles con información fue de 1.187.967. De ellos, el 47,1 %
239 correspondiente a señal meteorológica y un 52,9 % a señal no meteorológica.

240 Todos los clasificadores se testearon en un mismo conjunto de imágenes provenientes
241 de cuatro escenarios correspondientes a cuatro eventos distintos (Tabla 1). Se contó con
242 tres imágenes de cada uno de estos escenarios, resultando así un total de 12 imágenes.
243 Al momento de la selección de estas imágenes se procuró que las mismas resulten
244 representativas de los escenarios típicos encontrados: situaciones donde los ecos se
245 encuentren separados, situaciones con los ecos mezclados, situaciones con eco no
246 meteorológico sobre el sitio del radar, situaciones con presencia de eco meteorológico
247 intenso, entre otros.

248 3) METODOLOGIA

249 En vista de los trabajos discutidos y según los resultados obtenidos por los mismos, para
250 el presente trabajo se consideró trabajar con métodos de clasificación netamente



Meteorología

251 estadísticos, intentando en principio comprender el funcionamiento de los mismos, para
252 luego poder comparar los resultados provenientes de otras metodologías propuestas.

253 Junto con la técnica de clasificación a utilizar, son las variables de entrada, las que
254 determinan la bondad de un método. Como es de esperar la selección de las mismas
255 depende en parte de su capacidad discriminante, pero a la vez es necesario considerar su
256 disponibilidad. Varios autores han remarcado la capacidad predictiva que posee la
257 variable ρ_{HV} por si sola en cuanto a identificación de ecos no meteorológicos frente a
258 los meteorológicos: valores bajos de ρ_{HV} se suelen relacionar con ecos no
259 meteorológicos (Ryzhkov y Zrnic, 1998; Gourley y otros, 2006). No obstante, objetivos
260 como el granizo grande podrían llegar a tener valores bajos de este factor haciendo que
261 una región con presencia de eco meteorológico sea catalogada como no meteorológica.

262 El presente trabajo propone considerar el desvío estándar en una ventana de 3 x 3 en
263 rango y azimut de la reflectividad diferencial (SZDR). Es decir, para cada punto se
264 calcula el desvío estándar de Z_{DR} sobre la región de 9 vecinos cuyo centro es el punto en
265 cuestión.

266 Finalmente, las variables empleadas como variables de entrada (variables predictoras)
267 para los algoritmos desarrollados son:

- 268 • Coeficiente de correlación co-polar (ρ_{HV})
- 269 • Velocidad Radial Doppler (V)
- 270 • Desvío de la reflectividad diferencial (SZDR)

271 Si bien la variable SZDR otorga una noción de la estructura espacial al análisis, se
272 consideró necesario incorporar un proceso final a la clasificación obtenida por los
273 modelos (Postproceso), de manera de darle mayor importancia a la caracterización
274 espacial de los datos. Partiendo de la clasificación inicial del modelo, se reasignaron
275 algunos puntos teniendo en cuenta la clase de sus vecinos. Más específicamente, sobre
276 cada píxel se analizó una ventana de 3 x 3 (en rango y azimuth), si la mayoría de sus

277 vecinos (la mitad +1) fueron asignados a su misma clase, entonces no se hizo ningún
278 cambio, sino, se cambió la clase previamente asignada.

279 Uno de los inconvenientes encontrados fue que muchas veces los datos con los que se
280 contaba tenían variables faltantes, es decir a algunos puntos les faltaba información de
281 alguna de las variables (ρ_{HV} , V, y/o SZDR). A fin de salvar este inconveniente se
282 procedió de la siguiente manera: si la falta de dato se daba en la muestra de
283 entrenamiento, simplemente aquellos puntos fueron eliminados. En cambio, cuando el
284 punto con una variable faltante era parte del conjunto de testeo, se clasificó dicho punto
285 considerando el modelo con las variables disponibles. Es decir, para cada metodología
286 se ajustaron cuatro modelos, uno con las tres variables y los otros tres con las variables
287 tomadas de a dos. Luego dependiendo de la disponibilidad de variables en cada pixel se
288 predijo su clase según alguno de estos cuatro modelos ajustados. Por ejemplo si para el
289 píxel p, se contara con las variables ρ_{HV} y V, y SZDR no estuviese disponible, se
290 clasificaría el punto a partir del modelo que utiliza como variables (ρ_{HV}, V). Si la falta de
291 dato fuera en dos o más variables, se clasificaría al punto como sin clase (SC). En la
292 Figura 2 esquematizamos de manera sencilla la metodología a seguir.

293 Según los métodos de clasificación que se desarrollaran en la sección 3.1 se clasificaron
294 las imágenes del conjunto de testeo (Tabla I). Se asignó a cada pixel en una de las tres
295 categorías: meteorológica (M), no meteorológica (NM) o sin clase (SC). Estos
296 resultados fueron utilizados para evaluar la performance de cada método, comparando la
297 verdadera clase de cada uno de ellos con la asignada por el método de clasificación.

298 Una vez desarrollados los métodos y obtenidos los resultados para los casos de testeo,
299 se buscó hacer un análisis más completo sobre la performance de los clasificadores. Fue
300 necesario entonces darle mayor variabilidad a los datos. Con este objetivo se
301 implementaron técnicas de remuestreo sobre los datos de entrenamiento. El detalle de
302 este procedimiento se explica en la Sección 3.2.

303 3.1 METODOS DE CLASIFICACION SUPERVISADA

Meteoro logica

304 Las técnicas de clasificación supervisada parten de muestras pre-clasificadas con las que
305 se aspira aprender cómo discriminar (o clasificar) nuevas observaciones. Los métodos
306 abordados en este trabajo toman como premisa *la Regla de Clasificación de Bayes*, la
307 cual se basa en asignar la nueva observación a la clase con mayor probabilidad
308 condicional. Más específicamente, si se denota con $x = (x_v, x_{rho}, x_{szdr})$ a los valores de
309 las variables V, ρ_{HV} y SZDR en cierto punto p (información disponible), *la Regla de*
310 *Bayes* asignará al punto p el tipo de eco con mayor probabilidad condicional. Es decir,
311 si

$$312 \quad P(M|x) > P(NM|x) \quad (1)$$

313 se asignará p a la clase meteorológica, de lo contrario se lo asignará a la clase no
314 meteorológica, siendo que $P(M|x)$ y $P(NM|x)$ denotan la probabilidad de Eco
315 Meteorológico (M) y de Eco No Meteorológico (NM) respectivamente, condicional a
316 observar x.

317 La expresión (1) puede ser reformulada en términos de las probabilidades $P(M)$ y
318 $P(NM)$ de cada clase, combinadas con las funciones de densidad ($f_M(x)$ y $f_{NM}(x)$) del
319 vector $x = (x_v, x_{rho}, x_{szdr})$ en las mismas, obteniéndose así que (1) resulta equivalente a

$$320 \quad P(M)f_M(x) > P(NM)f_{NM}(x) \quad (2)$$

321 Las representaciones (1) y (2) sugieren dos enfoques diferentes para construir reglas de
322 clasificación. Por un lado, existen aquellas que, basadas en (1), que modelan
323 probabilidades condicionales. Por otra parte, la representación (2) da origen a
324 clasificadores mediante la estimación de las funciones de densidad condicional en cada
325 categoría, siendo que en muchas aplicaciones las probabilidades *a priori*, es decir $P(M)$
326 y $P(NM)$, suelen suponerse iguales (ambas con valor 0,5).

327 Existe una gran variedad de técnicas, que según el problema y los datos disponibles son
328 más o menos adecuadas. En lo que resta de esta sección se describirán brevemente las
329 metodologías que fueron utilizadas en este trabajo.

330 **a) Modelo Lineal Discriminante (LDA)**

331 El modelo discriminante lineal parte de la base que las funciones de densidad $f_M(x)$ y
 332 $f_{NM}(x)$ son gaussianas, ambas con distintos vectores de medias (μ_M y μ_{NM}) pero con la
 333 misma matriz de varianzas (Σ). Es decir,

$$334 \quad f_C(x) = \frac{1}{(2\pi)^{\frac{3}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \mu_C)' \Sigma^{-1}(x - \mu_C)\right\} \quad (3)$$

335 $\text{con } C \in \{M, NM\}$

336 Los parámetros del modelo (μ_M, μ_{NM} y Σ) se estimaron como:

$$337 \quad \hat{\mu}_C = \frac{1}{n_C} \sum_{i=1}^{n_C} x_{Ci} \quad (4)$$

$$338 \quad \hat{\Sigma} = \frac{n_M - 1}{n - 2} \hat{\Sigma}_M + \frac{n_{NM} - 1}{n - 2} \hat{\Sigma}_{NM}$$

$$339 \quad \hat{\Sigma}_C = \frac{1}{n_C - 1} \sum_{i=1}^{n_C} (x_{iC} - \hat{\mu}_C)(x_{iC} - \hat{\mu}_C)' \quad (5)$$

340 $\text{con } C \in \{M, NM\}$

341 donde $x_{Mi}(x_{NMi})$ denota la observación i -ésima perteneciente a la clase Meteorológica
 342 (No Meteorológica) de la muestra de entrenamiento, mientras que $n_M(n_{NM})$ denota el
 343 número total de observaciones de esa clase.

344 **b) Modelo Cuadrático Discriminante (QDA)**

345 En el método anterior se supuso que las observaciones dentro de cada una de las clases
 346 se distribuían según una ley Gaussiana, con distintas medias pero con la misma matriz
 347 de varianzas y covarianzas. Es posible ahora “relajar” este supuesto considerando el
 348 caso en que la matriz de varianzas y covarianzas sea diferente para cada clase. El
 349 razonamiento es el mismo que antes, simplemente que ahora será necesario estimar por

350 separado la matriz de varianzas y covarianzas de cada clase, además de cada una de las
 351 medias. Es decir, en este nuevo escenario los parámetros desconocidos que debemos
 352 estimar son: un vector de medias (μ_M y μ_{NM}) y una matriz de varianzas (Σ_M y Σ_{NM}) para
 353 cada clase.

354 Las estimaciones para los vectores de medias son las mismas que para el caso anterior
 355 (4), se distinguen estimaciones diferentes de las matrices de covarianzas (5).

356 Uno se podría preguntar por qué se elegiría el método lineal si es posible utilizar un
 357 método más flexible como el cuadrático. La respuesta viene de la mano de la relación
 358 sesgo-varianza presente en cada uno de los métodos: Si bien se tiene que el clasificador
 359 cuadrático es más flexible, a la vez necesita de la estimación de una mayor cantidad de
 360 parámetros que el clasificador lineal. Por otro lado, el modelo lineal si bien es menos
 361 flexible, es más estable en lo que refiere a cambios en el resultado por modificaciones
 362 en la muestra de entrenamiento.

363 Para obtener más información sobre las técnicas de clasificación dadas por el análisis
 364 del discriminante (a) y (b), se puede consultar en Peña (2002).

365 c) Modelo Bayes Naive (BN):

366 Este modelo busca estimar las funciones de densidad $f_M(x)$ y $f_{NM}(x)$ con aún menos
 367 suposiciones, partiendo únicamente de la muestra de entrenamiento. Se propuso
 368 entonces utilizar lo que se denomina *Estimador Núcleo de la Densidad* (en este caso
 369 núcleo gaussiano):

$$370 \quad \hat{f}_C(x) = \frac{1}{n_C h^3 (2\pi)^{3/2}} \sum_{i=1}^{n_C} \exp\left\{\frac{-1}{2h^2} ((x - x_i)'(x - x_i))\right\} \quad (6)$$

$$371 \quad C \in \{M, NM\}$$

372 En donde x es un vector de tres coordenadas (correspondientes a V, ρ_{HV} y SZDR) y h el
 373 parámetro de suavizado. La elección del parámetro de suavizado es esencial para el

Meteoro logica

374 buen funcionamiento del estimador. Como h controla la concentración de peso
 375 alrededor de cada punto de la muestra de entrenamiento, se tiene que valores chicos de
 376 h darán lugar a que únicamente las observaciones más cercanas al punto donde se quiere
 377 estimar la función de densidad sean relevantes en la estimación. De lo contrario, al
 378 tomar valores grandes de h , observaciones más lejanas influirán también en la
 379 estimación. Se han desarrollado varias técnicas para la selección del parámetro de
 380 suavizado, que no sólo pueden ser utilizadas para el caso del estimador núcleo, sino que
 381 también pueden ser fácilmente aplicables para otros estimadores. Para este trabajo se
 382 tomó como h el dado por *la Regla de Referencia a la normal*. Para más información
 383 sobre esta técnica y sobre este método de clasificación en general se puede consultar
 384 Silverman (1986) y Delicado (2008).

385 En general, el estimador tal como está en la expresión (6) no resulta práctico de aplicar,
 386 debido al alto nivel de cómputo que este implica, sumado a que cuando la cantidad de
 387 variables explicativas del modelo aumenta, la precisión del estimador disminuye. Como
 388 el objetivo final está en la clasificación, conocer de forma precisa la densidad de los
 389 datos resulta innecesario. Una forma de eludir este problema es construir un estimador
 390 bajo la hipótesis de que las variables (en este caso: ρ_{HV} , V y $SZDR$) son
 391 independientes. Se consideró entonces como estimador:

$$392 \quad \hat{f}_C(x) = \prod_{k=1}^3 \hat{f}_{Ck}(x_k) \quad k \in \{\rho_{HV}, V, SZDR\} \quad (7)$$

393 donde ahora $\hat{f}_{Ck}(x)$ denota el estimador no paramétrico de la densidad de la variable k
 394 en la clase C , y se calcula como en (6), pero utilizando solamente la variable x_k . Este
 395 método se lo denomina “Método de Bayes Naive”. Si bien el mismo parte de una
 396 suposición que no suele ser correcta en la mayoría de los casos, no sólo simplifica de
 397 forma evidente el estimador sino que también en la práctica da muy buenos resultados.

398 Hasta aquí los tres modelos presentados modelan y estiman las funciones de densidad
 399 de cada una de las clases. Una vez obtenidos los valores de los estimadores, al suponer

400 que las probabilidades a priori son iguales para ambas clases, la regla de clasificación
401 para un cierto punto (x) estará dada por:

$$402 \quad r(x) = \begin{cases} M & \text{si } \hat{f}_M(x) \geq \hat{f}_{NM}(x) \\ NM & \text{sino} \end{cases} \quad (8)$$

403 **d) Modelo de Regresión Logística (LG)**

404 Este método sugiere modelar $P(M|x)$ y $P(NM|x)$, para luego clasificar según (1). El
405 modelo de regresión logística propone considerar:

$$406 \quad P(C|x) = \frac{1}{1 + e^{-\beta_{0C} + \beta_{1C}'x}} \quad (9)$$

407 En donde β_{0C} es un parámetro univariado, $\beta_{1C} = (\beta_{1Cv}, \beta_{1C\rho_{HV}}, \beta_{1CsZdr})$ es un parámetro
408 de tres coordenadas y $C = M$ ó NM . Supongamos que $C = M$, entonces $P(NM|x) = 1 -$
409 $P(C|x)$, análogamente se definiría $P(M|x)$, si se considerara $C = NM$.

410 Fue entonces necesaria la estimación de dichos parámetros, para la cual se utilizó el
411 método de máxima verosimilitud. La idea esencial es la de estimar los parámetros
412 desconocidos de forma tal que para todos los elementos de la muestra de entrenamiento
413 la probabilidad estimada de pertenecer a la verdadera clase de procedencia (entiéndase
414 la clase M o la clase NM) sea, los más posible, cercana a uno. Obtener este valor, es
415 decir obtener el valor que maximiza la verosimilitud no resulta ser una tarea trivial. De
416 todas formas, fue posible aproximar estos estimadores a partir de métodos iterativos
417 muchos de los cuales han sido implementados en varios paquetes estadísticos como **R**
418 (<http://cran.r-project.org/>).

419 Más información sobre el modelo de regresión logística se puede encontrar en Peña
420 (2002) y Hastie y Tibshirani (2009), entre otros.

421 **3.2 TECNICAS DE REMUESTREO**

422 Las técnicas de remuestreo son herramientas indispensables en la estadística moderna.
423 Las mismas consisten en tomar repetidas muestras sobre los datos de entrenamiento



Meteoro logica

424 para ajustar nuevamente los modelos de interés, obteniendo así información adicional
425 sobre los mismos.

426 En el caso de este trabajo, las muestras para entrenar y testear los modelos requirieron
427 de una clasificación manual de las imágenes y por ello resultó dificultoso contar con un
428 volumen de datos considerable para trabajar. Si bien los resultados obtenidos en las
429 imágenes testeadas fueron buenos, se consideró necesario hacer un análisis estadístico
430 más profundo de las performance de los distintos métodos aplicados. Para ello fue
431 necesario plantear una manera de darle variabilidad a los datos. Luego de evaluar varias
432 posibilidades se decidió aplicar *técnicas de remuestreo* sobre los pixeles de las
433 imágenes de entrenamiento. Todos los clasificadores construidos a partir de los
434 diferentes métodos y entrenados con distintas muestras (del remuestreo) fueron
435 evaluados utilizando los datos test considerados en la Tabla I.

436 Para el análisis siguiente se supuso que la clasificación del experto fue perfecta. Con el
437 objetivo de poder cuantificar la performance de los diferentes modelos, se utilizó como
438 índice de bondad el valor de la proporción de puntos bien clasificados sobre el total de
439 cada elevación (*CSI*, por sus siglas en inglés *critical success index*).

440 El procedimiento aplicado fue el siguiente: se llama N a la cantidad de pixeles totales de
441 la muestra de entrenamiento, entonces:

- 442 1. Se seleccionaron de manera aleatoria y con reposición N puntos de la muestra de
443 entrenamiento.
- 444 2. Tomando como muestra para entrenar la resultante del ítem anterior, se ajustaron
445 los cuatro modelos (BN, LDA, QDA, LG).
- 446 3. Se clasificaron las 12 imágenes de los casos test.
- 447 4. Se aplicó el postprocesamiento.
- 448 5. Se calculó los valores de CSI para cada caso test y para cada modelo.
- 449 6. Se volvió al punto 1.

450 Se computó la secuencia anterior 100 veces (en adelante, replicaciones).

451 **4) RESULTADOS**

452 Se implementaron los cuatro métodos descritos en la sección anterior: LDA, QDA, BN
453 y LG. Tal como fue descrito previamente, luego de la clasificación inicial sobre la
454 muestra de testeo según cada una de las metodologías, se realizó un postprocesamiento
455 de forma de incluir la estructura espacial de los datos, siguiendo la regla de la mayoría.
456 Tal como se indicó en la Sección 2 todos los clasificadores fueron testeados en los
457 cuatro eventos de la Tabla 1.

458 En las Figuras 3, 4, 5, y 6 se muestran los resultados de los cuatro modelos para el caso
459 del 22 de Noviembre de 2009. Los resultados para el caso del 9 de Febrero de 2010 se
460 encuentran representados en las Figuras 7, 8, 9 y 10. Si bien es posible observar que en
461 ambos casos se dan situaciones en donde las dos clases (M y NM) se encuentran
462 presentes, es claro que en el caso de 2009 los ecos están más mezclados, en
463 comparación con la situación de 2010. Se puede observar que los resultados obtenidos
464 con los diferentes métodos de clasificación coinciden en un alto porcentaje. De todas
465 formas, es posible notar que las diferencias se encuentran en las zonas de frontera entre
466 ambas clases, obteniendo discrepancias mayores en la situación de 2009. Esto último es
467 de esperar ya que los puntos de frontera poseen cantidades similares de vecinos de las
468 dos clases lo que afecta por un lado al postproceso, que considera los valores de estas
469 cantidades, y por otro lado al comportamiento mismo de la variable SZDR en las zonas
470 de frontera. En este caso el efecto se debe a que SZDR se construye a partir de la
471 variabilidad espacial de Z_{DR} . Por consiguiente, sería posible que tanto el postproceso
472 como el valor de SZDR en las zonas de frontera introdujeran errores al momento de la
473 clasificación. De todas formas, la distribución espacial de los datos es sumamente
474 importante al momento de clasificar puntos que no se encuentren en las zonas de
475 frontera, ya que otorga una distinción importante en el comportamiento de las variables
476 en las dos clases (ZDR que suele ser más “ruidosa” en presencia de ecos no
477 meteorológicos) y a la vez permite capturar la continuidad de los ecos (postproceso).
478 Resulta entonces evidente que los puntos de frontera son los que proponen el desafío
479 mayor al momento de clasificar los datos en ecos meteorológicos o no meteorológicos.



Meteoro logica

480 Con respecto a la variable Velocidad Doppler, se observó que la misma no presenta
481 demasiada relevancia a la hora de la clasificación. La velocidad permite en términos
482 generales dar una idea de la naturaleza de los objetivos escaneados, pero puede llevar a
483 conclusiones erróneas si se la observa por sí sola. Es decir, ecos producidos por el
484 fenómeno de la propagación anómala poseen velocidades cercanas a cero, pero
485 objetivos como ecos biológicos pueden poseer velocidades similares a las de un eco de
486 precipitación. A su vez, puede suceder que objetivos meteorológicos posean velocidades
487 bajas. Quizás, si el objetivo de la clasificación fuese lograr una discriminación más fina
488 dentro de los ecos no meteorológicos (por ejemplo distinguir entre ecos biológicos, ecos
489 de terreno y propagación anómala) la variable velocidad podría tener mayor relevancia
490 para el análisis.

491 En la Tabla II se presentan los resultados obtenidos para la media, mediana, máximo,
492 mínimo, primer cuartil, tercer cuartil y desvío del índice CSI correspondiente a cada uno
493 de los cuatro métodos utilizados a lo largo de las 100 repeticiones del remuestreo. Es
494 posible notar como los cuatro métodos contienen valores similares de mediana, pero no
495 así de la media, lo que indica de alguna manera que hay diferencias en la dispersión de
496 los datos.

497 En las Figuras 11 y 12 se muestran los histogramas y los boxplots correspondientes a
498 los resultados de los valores de los CSI en cada uno de los cuatro métodos, basados en
499 las 100 repeticiones. Como primera observación es posible notar que si bien los valores
500 de las medianas en los cuatro modelos son muy semejantes entre sí, el modelo con
501 menor varianza es el modelo lineal. Además el modelo de Bayes Naive, si bien tiene
502 mayor nivel de varianza, es el que logra valores más altos de CSI. Teniendo en cuenta la
503 relación sesgo-varianza presente en estos modelos sería lógico esperar que el modelo
504 más simple (LDA), es decir el modelo con menos parámetros que ajustar sea el que
505 posea menor varianza y mayor sesgo. Es interesante notar cómo en este caso sucede lo
506 primero pero no lo segundo, o por lo menos no es evidente. El modelo Bayes Naive no
507 hace suposiciones sobre la forma de las funciones de densidad, otorgándole al método
508 mayor flexibilidad. Si bien esto último en principio es algo positivo ya que permite



Meteoro logica

509 obtener un modelo menos restrictivo, puede suceder que si el número de observaciones
510 no es suficiente, se obtenga una sobre-estimación que lleve a errores en las
511 predicciones. Es posible entonces que la estimación dada por este modelo se encuentre
512 bajo este escenario.

513 En general, elegir el nivel adecuado de flexibilidad para un modelo, no es una tarea
514 sencilla, depende en parte del problema y de los datos con los que se cuente. Por
515 consiguiente, a juzgar por los resultados obtenidos y en el caso de ser necesario elegir
516 únicamente alguno de los cuatro modelos, en principio parecería que el lineal es el que
517 mejor ajusta.

518 5) CONCLUSIONES

519 En este trabajo se presentaron cuatro técnicas de clasificación supervisada las cuales
520 buscan dar una solución al problema de discriminación entre ecos meteorológicos y no
521 meteorológicos en imágenes de radares meteorológicos muy importante al momento de
522 realizar un control de calidad de la información para su uso en múltiples aplicaciones
523 hidrometeorológicas. A partir de datos provenientes del radar Doppler banda C de doble
524 polarización instalado en la Estación Experimental Agropecuaria INTA Anguil, se
525 testearon cuatro modelos basados en las cuatro técnicas de clasificación presentadas con
526 distintos niveles de flexibilidad. Mediante técnicas de remuestreo se estudiaron los
527 desempeños de los mismos y se analizaron los resultados.

528 Las clasificaciones finales de los casos testeados en los cuatro métodos fueron
529 coincidentes y correctas en un alto porcentaje (siendo las zonas de frontera las menos
530 precisas). De todas formas, se vio que aplicar una metodología más flexible (BN) no
531 incorporaba precisión en las predicciones, obteniendo mejores clasificaciones con los
532 métodos más rígidos. Una posible razón puede ser que al no contar con un gran número
533 de observaciones los ajustes más flexibles sobreestimen los datos, resultando entonces
534 el modelo con mejor desempeño el Modelo Lineal.



Meteoro logica

535 No obstante, se podría suponer que si se contase con un set más amplio de datos para
536 entrenar los modelos, la clasificación resultante para el caso del modelo BN podría
537 resultar mejor. Obtener una muestra de entrenamiento mayor resultaría costoso, ya que
538 la misma fue realizada de forma manual por el experto meteorólogo, lo que implica una
539 mayor demanda de tiempo para poder incluir un número superior de casos. De todas
540 formas es importante aclarar que por la manera en que está planteado este modelo, para
541 cada punto a clasificar es necesario computar el estimador núcleo de la densidad con
542 cada uno de los puntos de la muestra de entrenamiento (Ecuación 6), lo que implica que
543 a medida que aumenta la muestra de entrenamiento, aumenta el tiempo de cómputo. Por
544 esta razón, habría que analizar la factibilidad de éste método si se aumentase
545 considerablemente la muestra de entrenamiento.

546 Si bien los resultados encontrados son muy alentadores, se cree necesario seguir
547 avanzando en esta línea de investigación a fin de incorporar una mayor cantidad de
548 casos y tener una mayor significancia de los resultados. Como primera medida, se cree
549 que incorporar nuevas variables derivadas del radar al análisis ayudaría a obtener
550 mejores resultados. Para ello será necesario trabajar en la calidad y calibración de las
551 mismas sobre los radares que se encuentran en funcionamiento. En segunda instancia tal
552 como se consideró la variable SZDR para incorporar nociones de la estructura espacial
553 en sentido horizontal, se estima que incluir también variables que den nociones de la
554 estructura espacial en el sentido vertical sería positivo. Por otra parte, como en este
555 trabajo se utilizaron técnicas clásicas de clasificación se cree necesario explorar y
556 evaluar algunas técnicas más avanzadas, como redes neuronales o técnicas de
557 clusterización. Por último se espera poder extender estas técnicas a otros radares de
558 doble polarización existentes en el país.

559 **AGRADECIMIENTOS:** La realización del presente trabajo fue financiada por los
560 proyectos PIDDEF N°5 2014-2017, PICT 2013-1299 y UBACyT 20020130100618BA.

561 **REFERENCIAS**

562 **Battan L.J. 1973:** Radar observation of the atmosphere. Q.J.R. Meteorol. Soc., 99: 793.

563

564 **Berenguer, M., Corral C., Sanchez-Diezma R. y Sempere-Torres D., 2006:** A fuzzy
565 logic technique for identifying non precipitating echoes in radar scans. J. of Atmos. and
566 Oceanic Tech., 23, 1157-1180.

567

568 **Bo Young Y., GyuWon L. y Hong-Mok P., 2015:** Identification and Removal of Non-
569 meteorological Echoes in Dual-polarization Radar Data Based on a Fuzzy Logic
570 Algorithm. Advances in atmospheric sciences, 32, 1217–1230

571

572 **Cho Y. H., Lee. G, Kim K. E. y Zawadski I., 2006:** Identification and removal of
573 ground echoes and anomalous propagation using the characteristics of radar echoes. J.
574 of Atmos. and Oceanic Tech., 23, 1206-1222.

575

576 **Delicado P., 2008:** Curso de Modelos no Paramétricos. Departamento de estadística e
577 investigación operativa, Universidad de Cataluña. 192 págs.

578

579 **Greku M. y Krajewski W. F., 2000:** An Efficient Methodology for Detection of
580 Anomalous Propagation Echoes in Radar Reflectivity Data Using Neural Networks. J.
581 of Atmos. and Oceanic Tech., 17, 121-129.

582

583 **Gourley J., Chatelet J. P. y Tabary P., 2006:** A Fuzzy Logic Algorithm for the
584 Separation of Precipitating from Nonprecipitating Echoes Using Polarimetric Radar
585 Observations, J. of Atmos. and Oceanic Tech, 24, 1439-1451.

586

587 **Hastie L., Friedman J. y Tibshirani R., 2009:** The Elements of statical learning: Data
588 Mining, Inference, and Prediction. Springer Text in Statistics. 747 págs.

589

590 **Hubbert J. C., Dixon M. y Ellis S. M., 2009:** Weather Radar Ground Clutter. Part II:
591 Real-Time Identification and Filtering. J. of Atmos. and Oceanic Tech., 26, 1181-1197.

592

593 **Kosko B., 1994:** Neural Networks and fuzzy systems: a dynamical systems approach to

- 594 machine intelligence. Prentice-Hall International Editions. 449 págs.
595
- 596 **Lakshmananetal V., Zhang J., y Howard K., 2010:** A Technique to Censor
597 Biological Echoes in Radar Reflectivity Data. J. of applied meteorology and
598 climatology, 49, 453-462.
599
- 600 **Moszkowicz S., Cian G. J. y Krajewski F., 1993:** Statistical Detection of Anomalous
601 Propagation in Radar Reflectivity Patterns. J. of Atmos. and Oceanic Tech, 11, 1026-
602 1034.
603
- 604 **Peña D., 2002:** Análisis de datos Multivariantes. S.A. McGraw-hill / Interamericana de
605 España. 515 págs.
606
- 607 **Rico-Ramirez M. A. y Cluckie I. D., 2008:** Classification of ground clutter and
608 anomalous propagation using dual-polarization weather radar. IEEE Transactions on
609 Geosciences and Remote Sensing, 46, 7, 1892-1904.
610
- 611 **Ryzhkov A. y Zrníc D. S., 1998:** Discrimination between Rain and Snow with a
612 Polarimetric Radar. J. of applied meteorology. 37, 1228-1240.
613
- 614 **Schuur T., Heinselman P. y Ryzhkov A., 2003:** Observations and classification of
615 echoes with the polarimetric WSR-88D radar. National Severe Storms Laboratory
616 (NOAA) and Cooperative Institute for Mesoscale Meteorological Studies (University of
617 Oklahoma).
618
- 619 **Silverman B. W., 1986:** Density estimation for statistics and data analysis. Chapman
620 and Hall/CRC. 176 págs.
621
- 622 **Siggia A. D. y Passarelli R. E., 2004:** Gaussian model adaptive processing (GMAP) for
623 improved ground clutter cancellation and moment calculation. Third European

624 Conference on Radar Meteorology (ERAD) 67–73.

625

626 **Steiner M. y Smith J. A., 2001:** Use of Three-Dimensional Reflectivity Structure for
627 Automated Detection and Removal of Nonprecipitating Echoes in Radar Data. *J. of*
628 *Atmos. and Oceanic Tech.*, 19, 673-685.

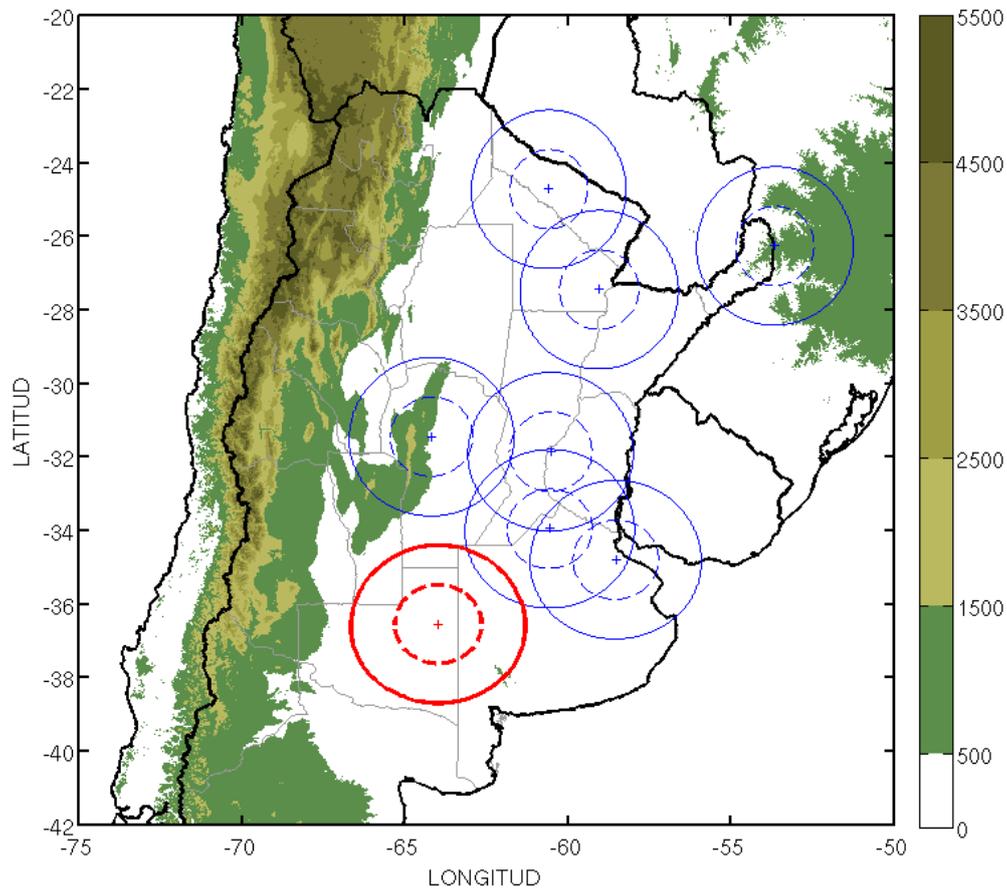
629

630 **Zawadzki, I., 1984:** Factors affecting the precision of radar measurement of rain, in 22
631 Conference on radar meteorology, edited by AMS, pp. 251-256, Zurich, Switzerland.

632

633 **Figuras y Tablas**

634



635

636

637 **Figura 1:** Ubicación del radar meteorológico Doppler banda C de doble polarización en
638 la Estación Experimental Agropecuaria INTA Anguil. En círculos rojos llenos se indica
639 el área de cobertura de 240 km, mientras que en punteado el de 120 km. En círculos
640 azules se muestra la posición de los radares de banda C existentes en Argentina. En
641 sombreado se muestra la topografía en metros.

642

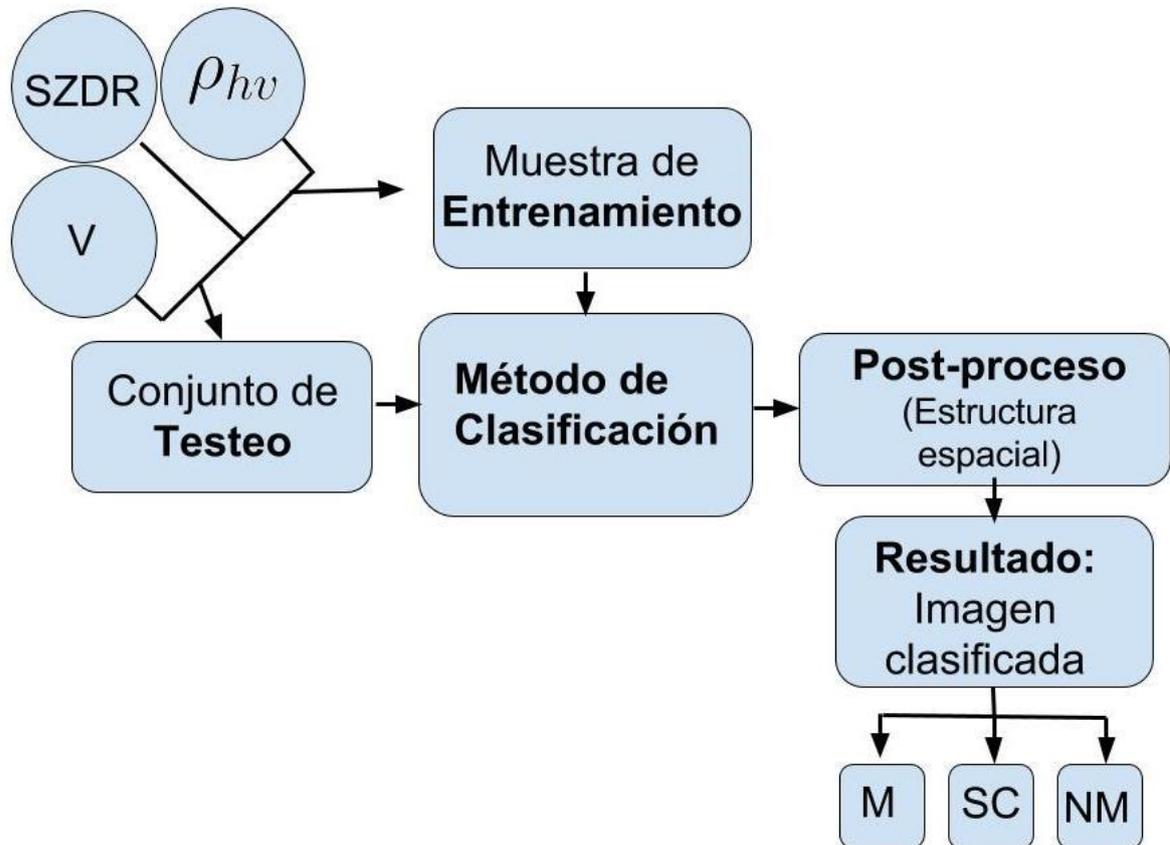
643

644

645

646

647



648

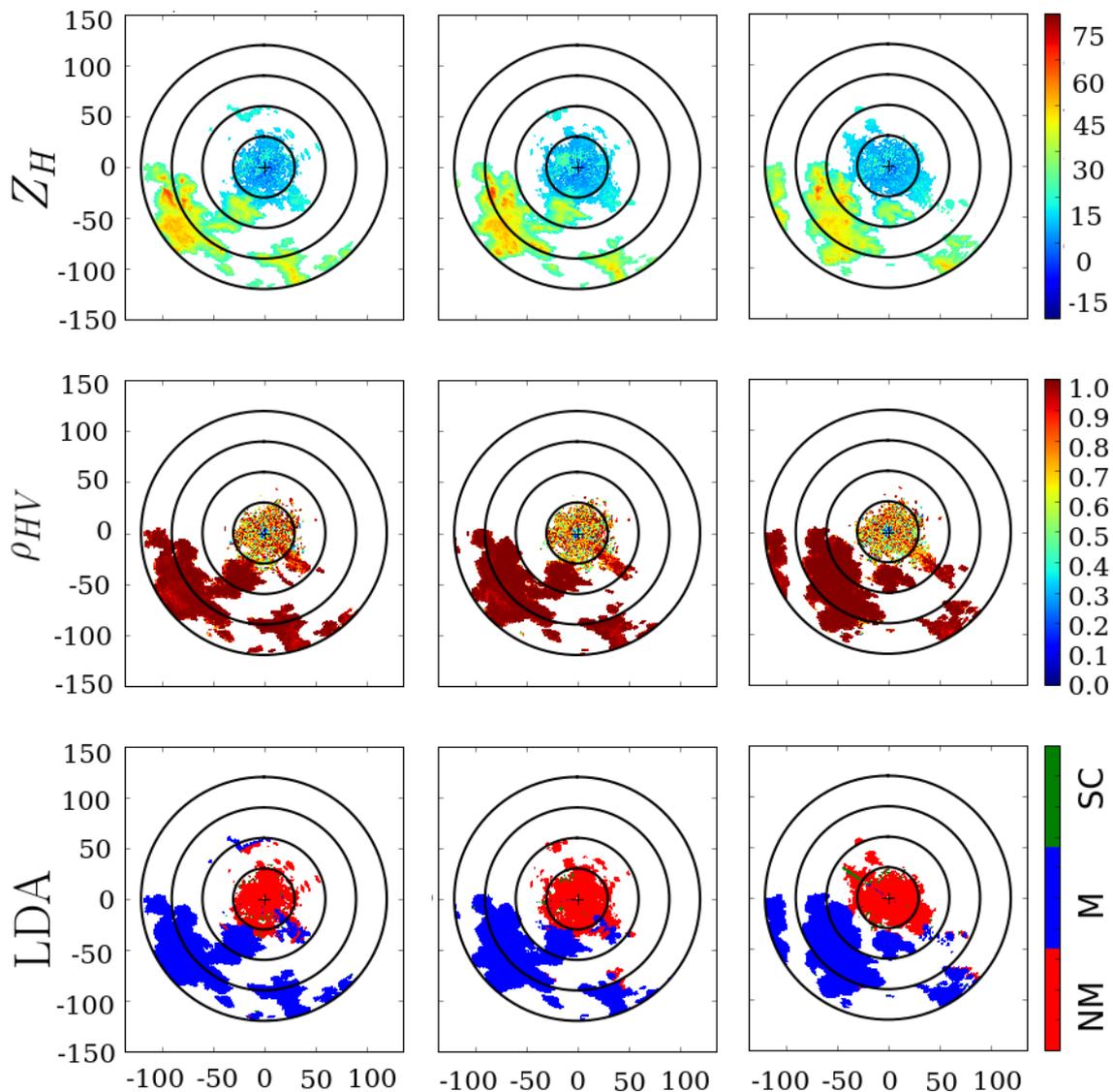
649

650 **Figura 2:** Diagrama de flujo sobre la metodología a seguir para el problema de
 651 clasificación en imágenes de radar propuesto en el presente trabajo. La letra M
 652 corresponde a la categoría Eco Meteorológico, la NM a Eco No Meteorológico y SC a
 653 datos sin clasificar.

654

655

656



657

658 **Figura 3:** Resultados obtenidos para el 22 de noviembre de 2009 en el clasificador659 LDA. Por fila y sobre la primera elevación (0.5°): reflectividad horizontal Z_H (dBZ),660 ρ_{HV} , resultado del clasificador: en azul los puntos clasificados como meteorológicos

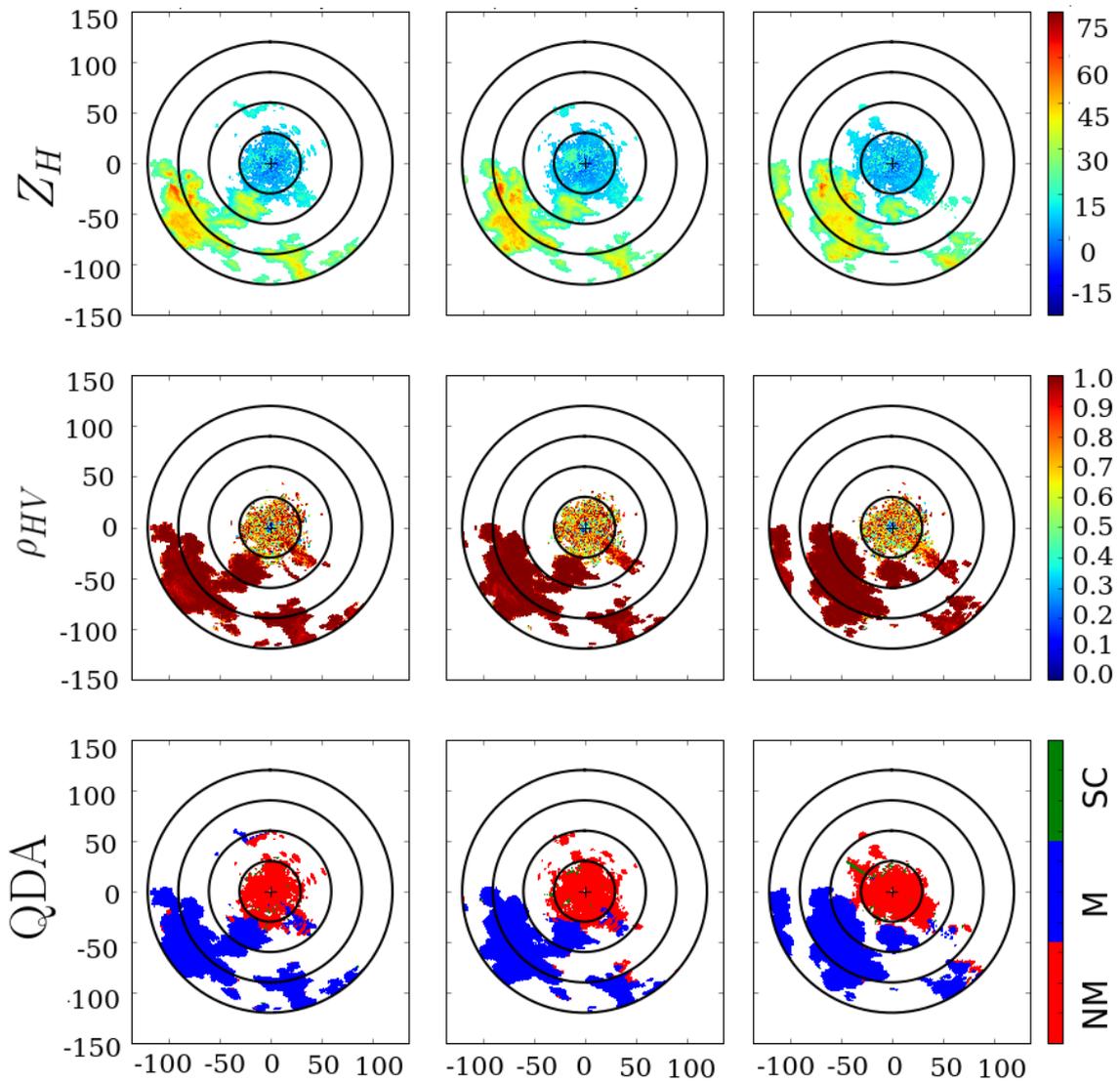
661 (M), en rojo los clasificados como no meteorológicos (NM), y en verde los

662 considerados sin clase (SC). Los anillos concéntricos poseen una separación de 30 km

663 entre sí. Por columnas: tiempos sucesivos correspondientes a 08:53 UTC, 09:03 UTC y

664 09:23 UTC.

665



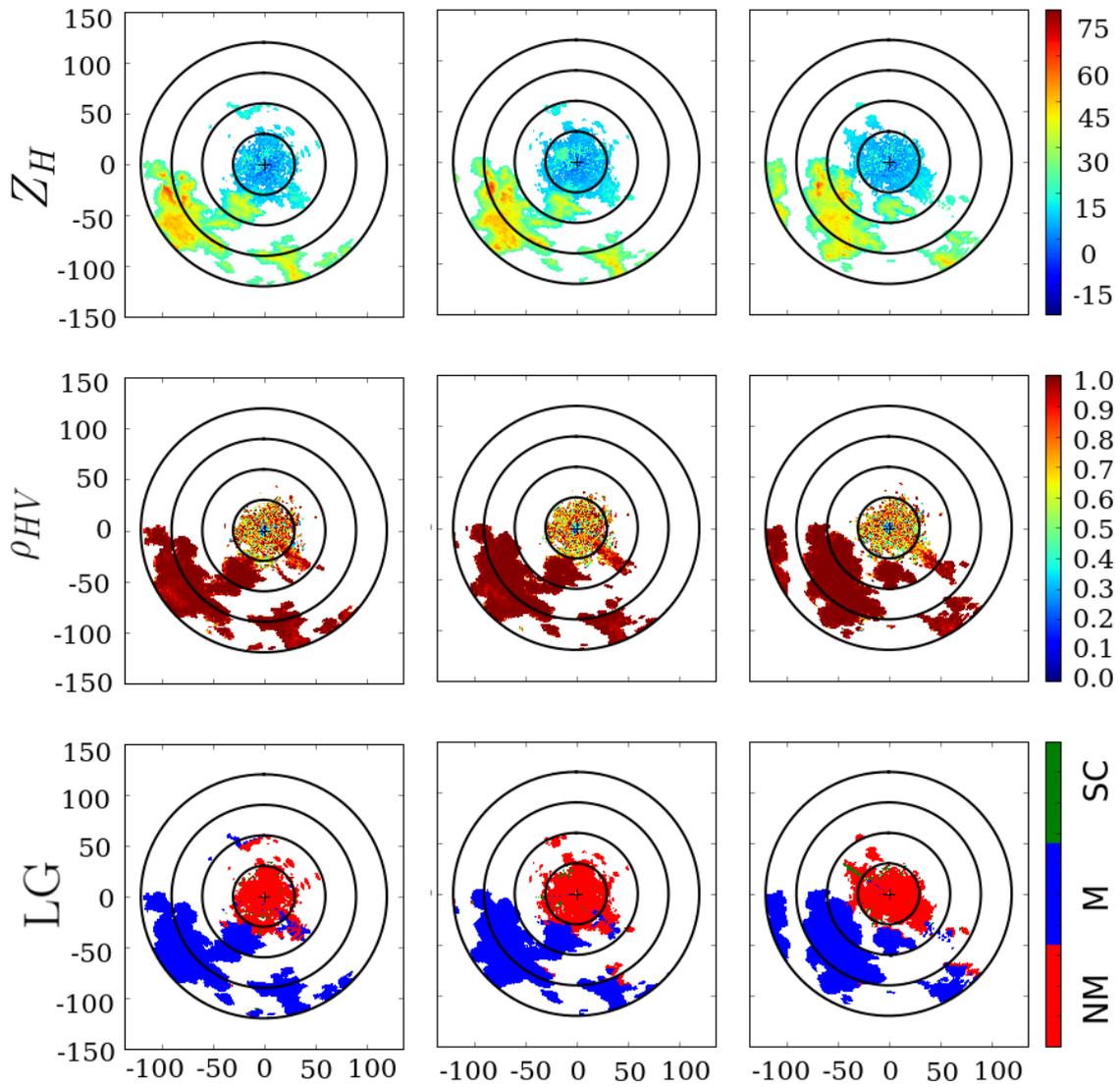
666

667

668

669

Figura 4: Ídem Figura 3 pero para clasificador QDA.



670

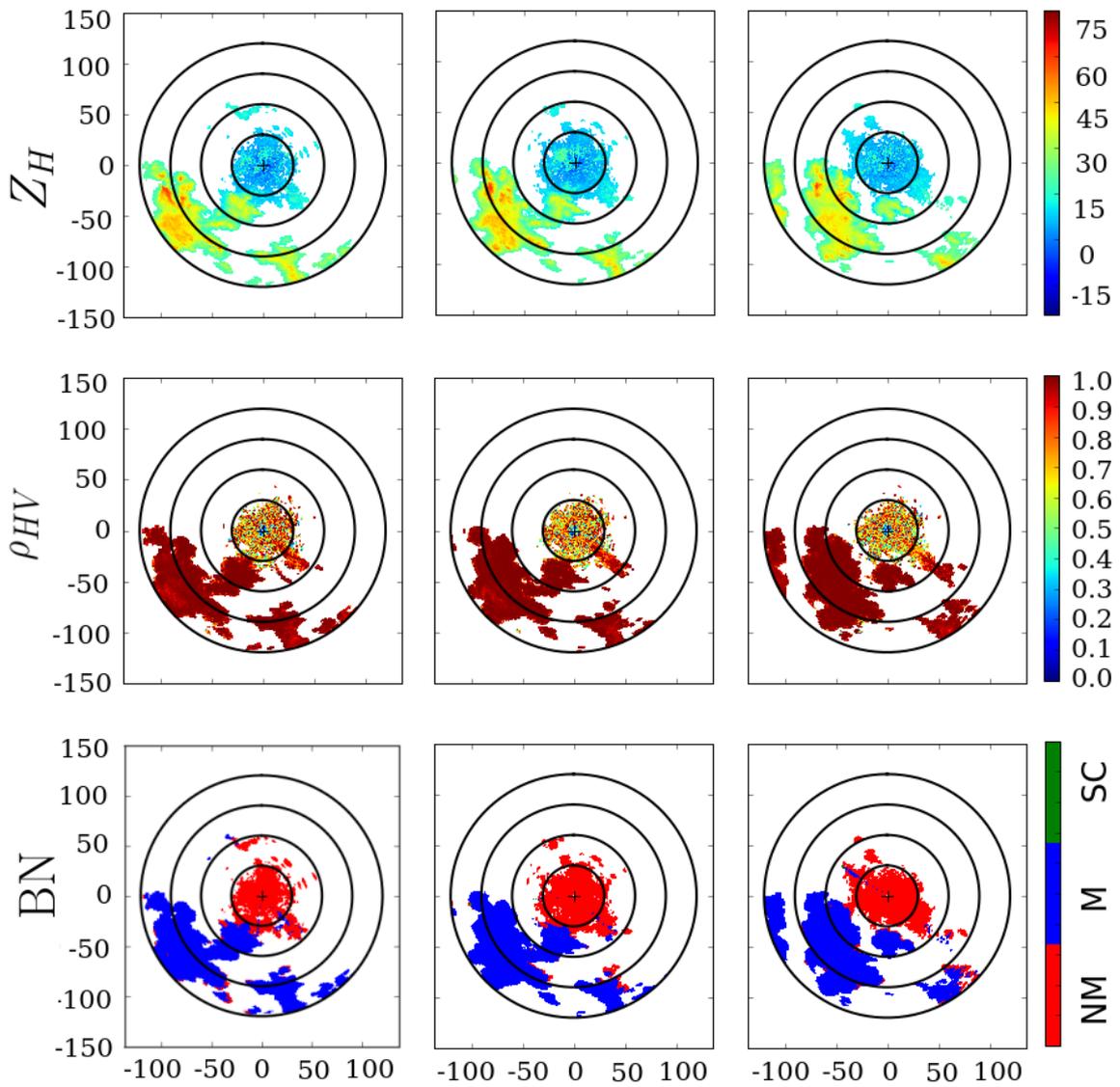
671

672

673

674

Figura 5: Ídem Figura 3 pero para clasificador LG



675

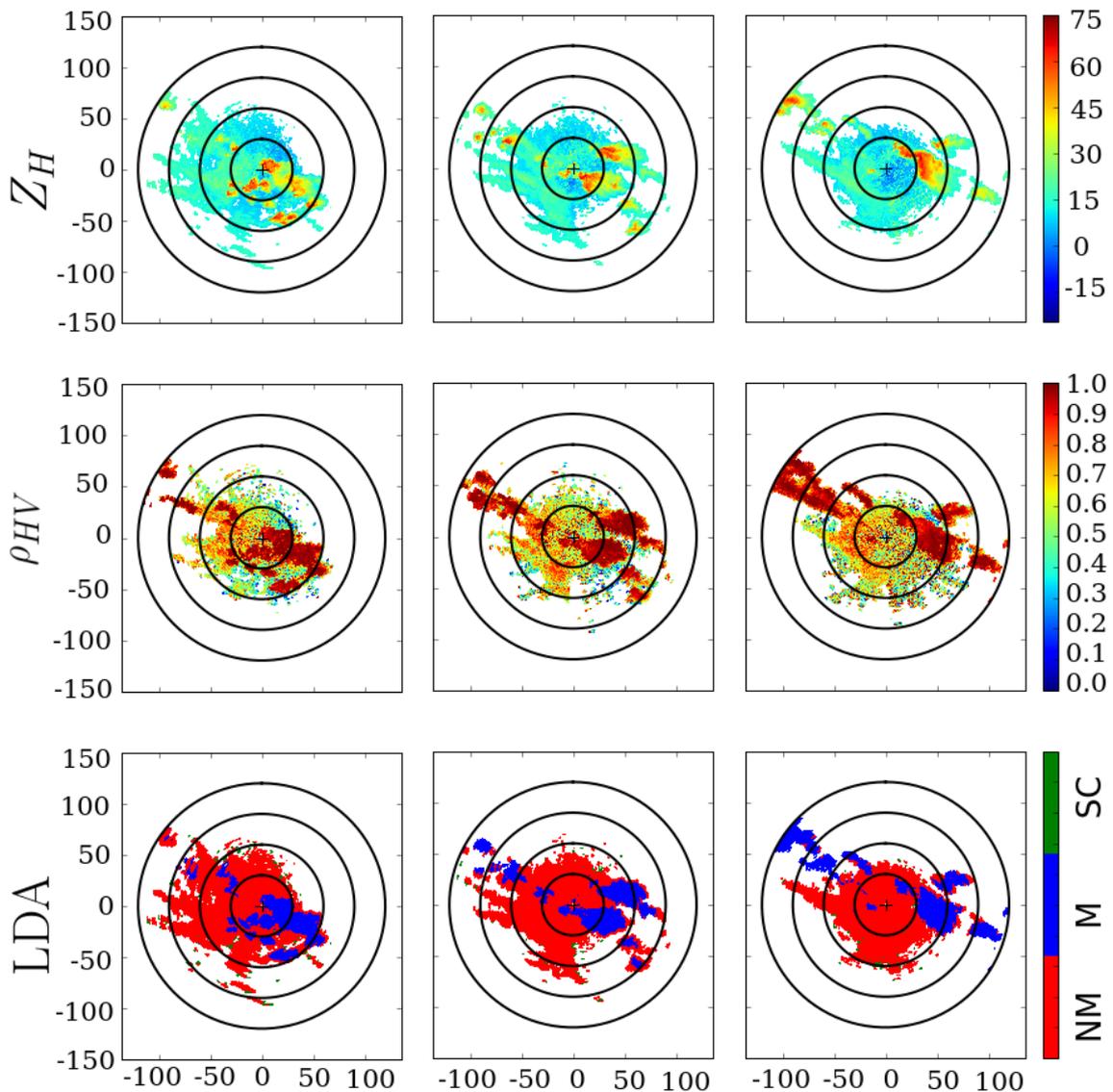
676

Figura 6: Ídem Figura 3 pero para clasificador BN.

677

678

679



680

681 **Figura 7:** Resultados obtenidos para para el 9 de febrero de 2010 en el clasificador682 LDA. Por fila y sobre la primera elevación (0.5°): reflectividad horizontal Z_H (dBZ),683 ρ_{HV} , resultado del clasificador: en azul los puntos clasificados como meteorológicos

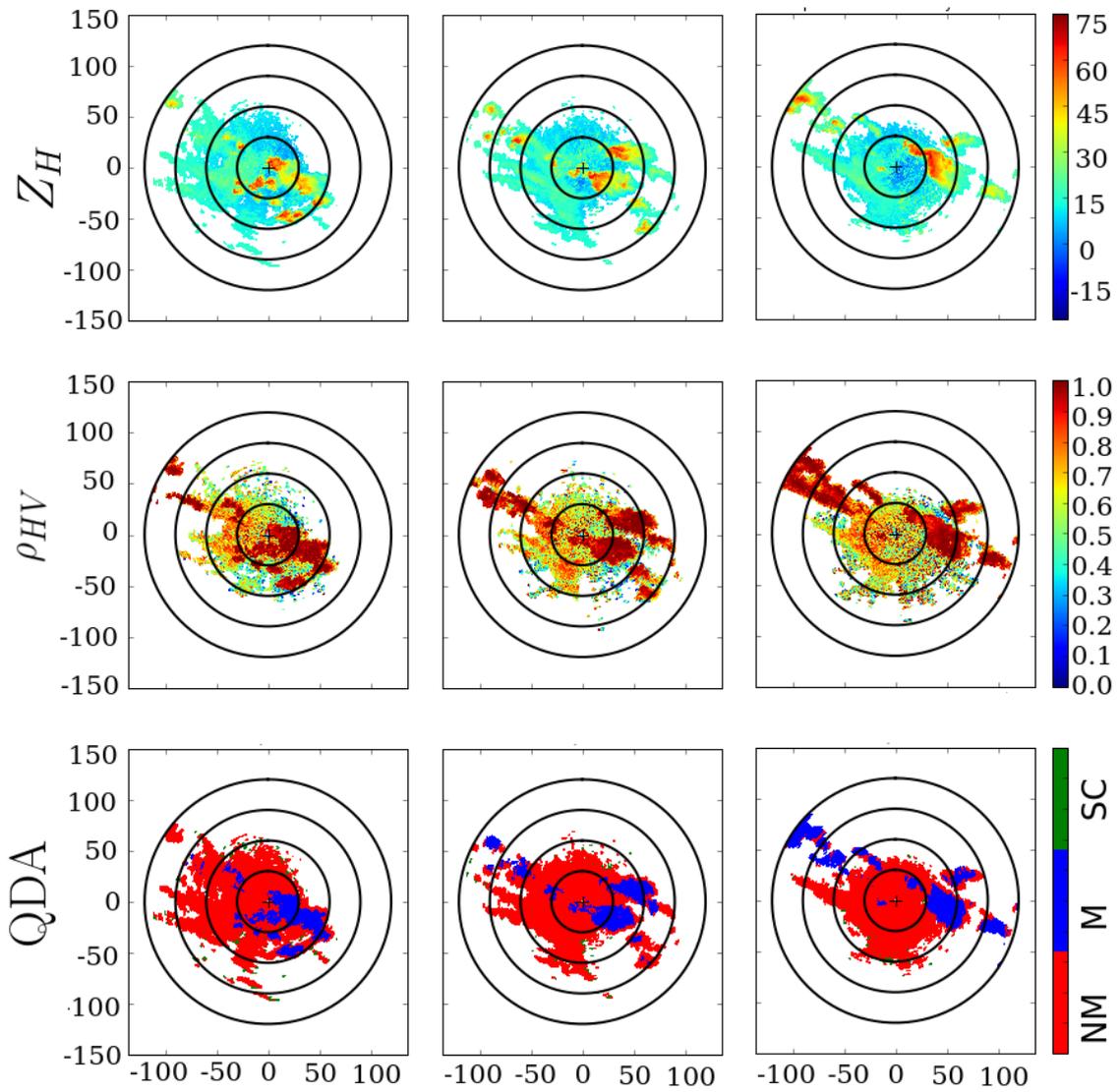
684 (M), en rojo los clasificados como no meteorológicos (NM), y en verde los

685 considerados sin clase (SC). Los anillos concéntricos poseen una separación de 30 km

686 entre sí. Por columnas: Por columnas: tiempos sucesivos correspondientes a 21:03 UTC,

687 21:43 UTC y 22:33 UTC.

688



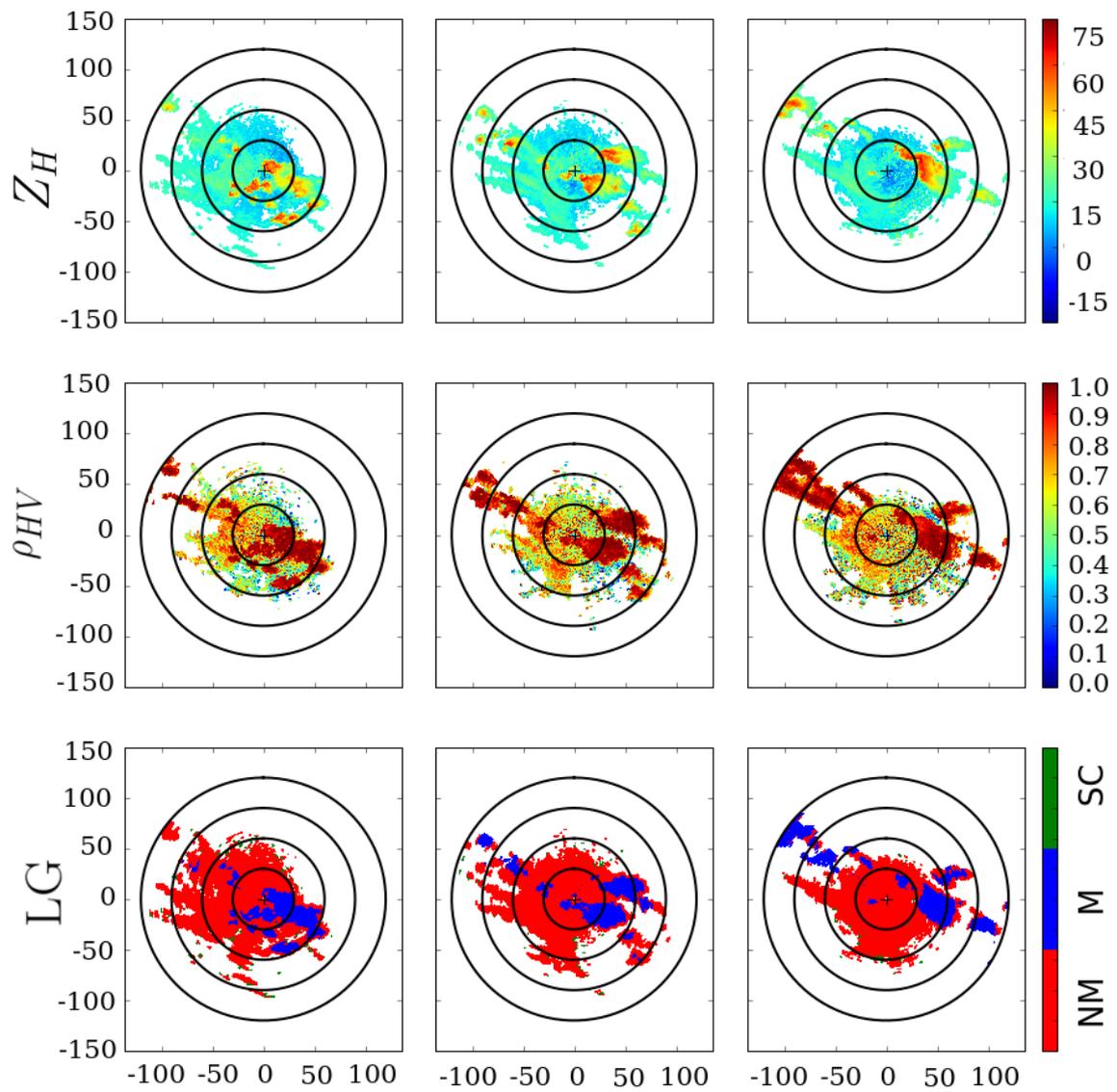
689

690

691

Figura 8: Ídem Figura 6 pero para el clasificado QDA

Meteorología

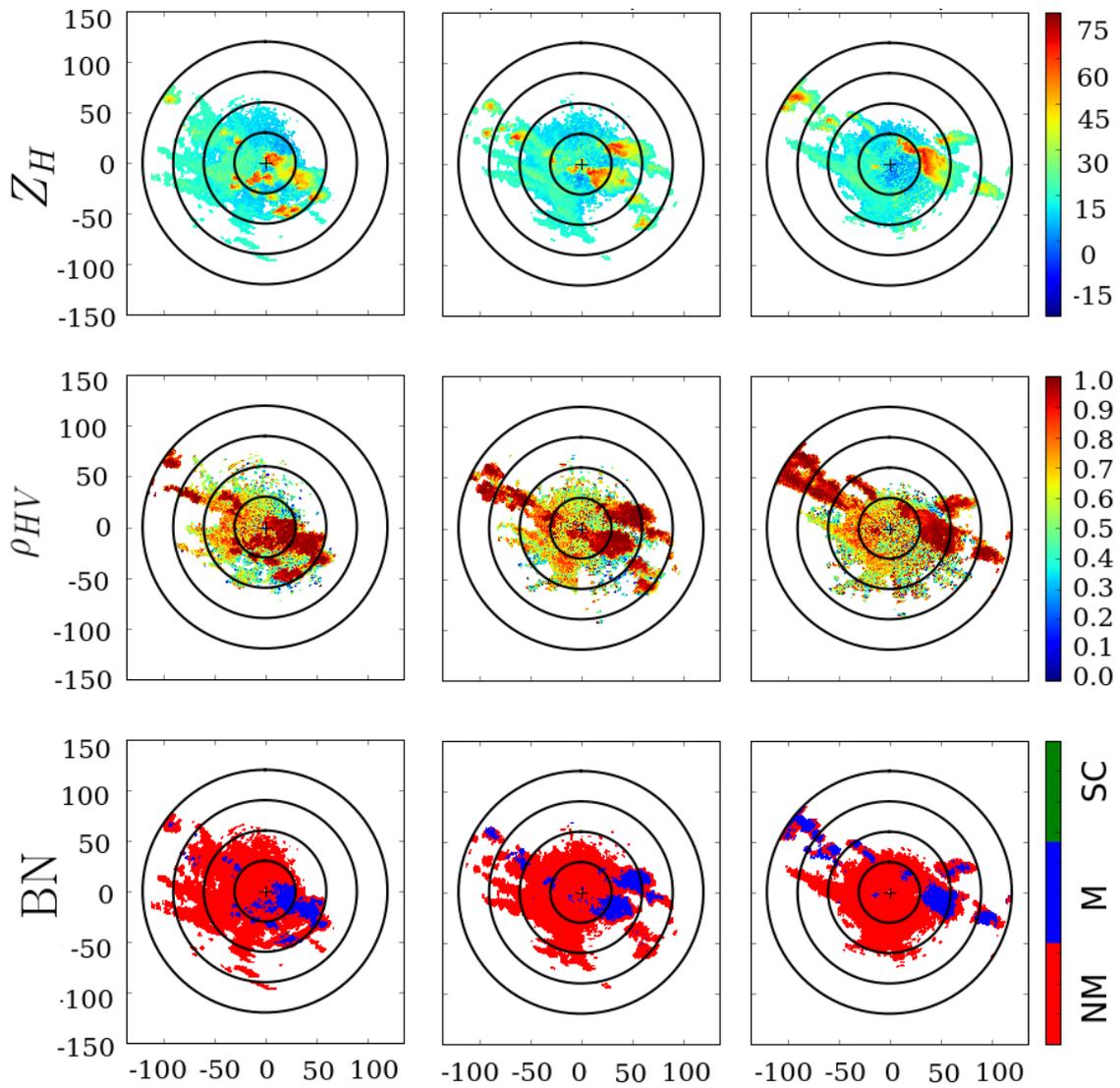


692

693

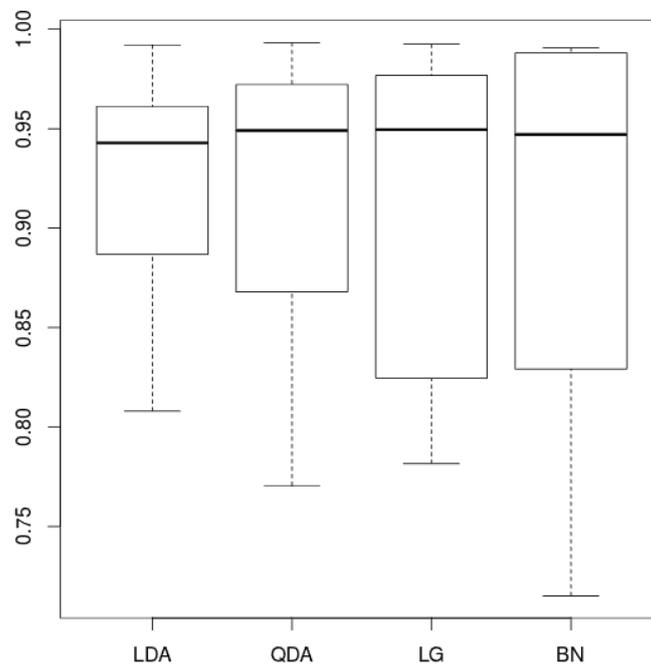
694

Figura 9: Ídem Figura 6 pero para el clasificado LG



695
696
697
698

Figura 10: Ídem Figura 6 pero para el clasificado BN

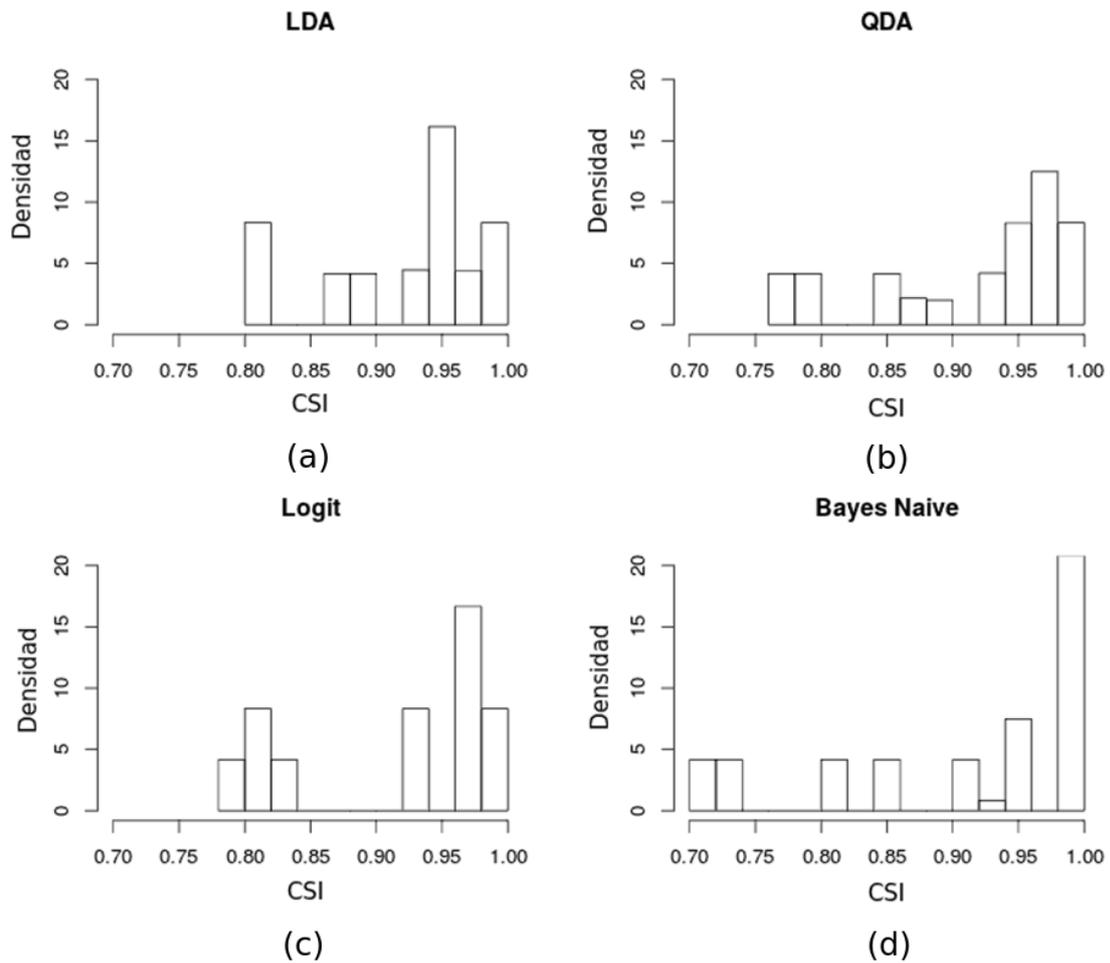


699

700

701 **Figura 11:** Boxplot de los valores de CSI obtenidos a partir de las 100 replicaciones en
702 los cuatro modelos. De izquierda a derecha: LDA, QDA, LG y BN. La línea negra
703 oscura marca el valor de la mediana en cada uno de los modelos. Los límites de las cajas
704 están dados por el primer cuartil (25%, límite inferior) y tercer cuartil (75%, límite
705 superior). Por último los bigotes de los extremos dan los valores del máximo (bigote
706 superior) y del mínimo (bigote inferior).

707



708

709

710 **Figura 12:** Histogramas de los valores de CSI obtenidos a partir de las 100

711 replicaciones en los cuatro modelos: (a) LDA, (b) QDA, (c) LG y (d) BN.

712

Fecha	Ventana Horaria	Características
20 de junio 2009	Desde 04:00 hasta 06:00 UTC	- Presencia de ecos meteorológicos extendidos espacialmente. - Sobre el sitio del radar se observa eco no meteorológico.
22 de noviembre 2009	Desde 08:30 hasta 09:30 UTC	- Presencia de eco meteorológico moderado. - Alrededor del radar: eco no meteorológico
9 de febrero 2010	Desde 21:00 hasta 23:00 UTC	- Mezcla de eco meteorológico y no meteorológico. - Eco meteorológico intenso.
27 de febrero de 2010	Desde 00:00 hasta 02:00 UTC	- Presencia de ecos no meteorológicos alrededor del sitio del radar aislados de ecos meteorológicos.

713

714

Tabla I: Casos para la validación

715

716

Modelo	Primer Cuartil	Mediana	Tercer Cuartil	Media	Desvío	Máximo	Mínimo
LDA	0,8912	0,9428	0,9606	0,9227	0,0577	0,9920	0,8080
QDA	0,8733	0,9490	0,9701	0,9177	0,0738	0,9932	0,7705
LG	0,8303	0,9494	0,9755	0,9134	0,0772	0,9924	0,7815
BN	0,8360	0,9470	0,9881	0,9040	0,0977	0,9907	0,7151

717

718 **Tabla II:** Resultados obtenidos a partir de las replicaciones. Los valores de la tabla son

719 referidos al índice de éxito crítico (CSI) que se muestra en la Figura 11.

720

721