

Radar Meteorológico de la EEA Anguil

Contribuciones para su uso en el sector agropecuario

EDITORAS

María Laura Belmonte

Yanina Bellini Saibene

EEA Anguil INTA



EEA Anguil “Ing. Agr. Guillermo Covas”
Instituto Nacional de Tecnología Agropecuaria

Diseño Gráfico

Dis. Gráf. Francisco Etchart

Impresión

Gustavo J. Moyano

Impreso en los talleres gráficos de la EEA INTA Anguil

“Ing. Agr. Guillermo Covas”

Tirada de 500 ejemplares

Septiembre de 2017



EDICIONES INTA

Centro Regional La Pampa-San Luis

EEA INTA Anguil “Ing. Agr. Guillermo Covas”

RN N°5 Km 580, CP 6326, Anguil, La Pampa, Argentina

Prólogo

La Estación Experimental Agropecuaria (EEA) INTA Anguil “Ing. Agr. Guillermo Covas” se ha destacado en recopilación de información sobre fenómenos meteorológicos, fruto del aporte ininterrumpido de las estadísticas climáticas. En este sentido, la institución, como proveedora de información meteorológica, generó una óptima interrelación con los diferentes usuarios de la actividad agropecuaria, de los recursos hídricos y del medio ambiente.

En los últimos tiempos, el INTA realizó una importante inversión para desarrollar infraestructura moderna de captura y procesamiento de datos climáticos. Una de estas iniciativas fue la instalación de una red de radares meteorológicos, de los cuales uno está ubicado en la EEA Anguil. Esta tecnología permite disponer de sistemas de monitoreo, en tiempo real, sobre el impacto del clima y sus variaciones en las diferentes regiones productivas de la región pampeana.

El monitoreo diario de algunas variables del clima llevó al estudio de las precipitaciones y de eventos como el granizo y su daño en los cultivos. Los trabajos presentados aportan información de la gestión y modelado de los datos del radar, con un enfoque de su uso aplicado al sector agropecuario.

Los documentos exhibidos en esta publicación de divulgación científica fueron posibles gracias al trabajo conjunto entre el grupo de Información Agropecuaria y AgroTICs de la EEA Anguil, el Instituto de Clima y Agua del INTA y la Facultad de Ingeniería de la Universidad de La Pampa (UNLPam). Algunos de ellos fueron presentados en diferentes congresos y lograron reconocimientos.

PhD. Jesús Pérez Fernández

Director EEA Anguil “Ing. Agr. Guillermo Covas”

Contenidos

1. Evaluación de la estimación de precipitación a partir de radar meteorológico	7
2. Ajuste meteorológico de precipitación estimada con radar	21
3. Imágenes de radar meteorológico como recurso en la planificación de actividades agropecuarias	27
4. Breve introducción a las tormentas de granizo, el daño en cultivos, los radares meteorológicos y las técnicas de Minería de Datos	37
5. Identificación de granizo y daño en cultivos con la utilización de variables polarimétricas de los radares de Paraná y Anguil	53
6. Estimación de ocurrencia de granizo en superficie mediante datos de radar meteorológico utilizando técnicas de Data Mining	61
7. Clasificación de daño en cultivos por granizo utilizando Gene Expression Programming y datos de radar meteorológico	81
8. HAR (HAIL ARCHIVE): sistema de información para registrar eventos de granizo, sus efectos en superficie y los datos de radares meteorológicos en la región pampeana central	105
9. RadINTA: un conjunto de herramientas de software libre para facilitar la explotación de datos de los radares meteorológicos del INTA	113
10. Hacia una arquitectura de procesamiento de datos del RADAR Meteorológico de INTA Anguil	125

Evaluación de la estimación de precipitación a partir de radar meteorológico

1.

Belmonte, María Laura¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
belmonte.maria@inta.gob.ar

RESUMEN

Los registros de precipitación mediante datos puntuales de pluviómetros, en redes de escasa densidad e irregular distribución, no reflejan adecuadamente la variabilidad espacial característica de la lluvia. Los radares meteorológicos ofrecen la capacidad de aportar información espacial detallada con la posibilidad de integrarla temporalmente. El objetivo de este trabajo fue evaluar el desempeño del radar meteorológico ubicado en la Estación Experimental de INTA Anguil, provincia de La Pampa, como estimador de la precipitación. Se generó el producto de precipitación acumulada por radar en imágenes con resolución espacial de 1 km², utilizando una sola relación entre reflectividad y tasa de precipitación. Los compuestos integrados diario, decádico y mensual se compararon con datos provenientes de una red pluviométrica de superficie existente en la región durante tres períodos: diciembre de 2011 a mayo de 2012, agosto a octubre de 2012, y enero a marzo de 2013. Como resultado se halló una correlación moderada a alta entre las estimaciones de radar y los registros de referencia. El ajuste lineal entre ambos conjuntos de datos fue significativo para los tres períodos (0,74; 0,59 y 0,89 $p < 0,05$) y se verificó una tendencia a la sobreestimación de la cantidad de lluvia de manera sistemática. La precisión de las estimaciones desmejoró con el incremento de la distancia al radar, haciéndose muy evidente a partir de los 90-100 km (RMSE 15,6 mm) en coincidencia con una amplia referencia bibliográfica. A distancias superiores a 50 km la pérdida de precisión se relacionó con lluvias más cuantiosas. Los eventos de precipitación de mayor intensidad fueron estimados con más dispersión y magnitud del error respecto de lluvias menos intensas. Las estimaciones de otoño y primavera presentaron menor

error relativo (1,3 a 2,3%) respecto de los meses de verano (3,6 a 7,6%). Se concluyó que las estimaciones de radar logran caracterizar la distribución espacial de las precipitaciones con un error aceptable y poco variable en el tiempo.

Palabras clave: lluvia, reflectividad radar, errores de estimación

ABSTRACT

Precipitation data of rain gauges from networks of low density and irregular distribution do not adequately reflect the characteristic spatial variability of rainfall. Meteorological radars provide detailed spatial information with the possibility of integrating it temporally. The objective of this work was to evaluate the performance of the meteorological radar located at the Experimental Station of INTA Anguil, province of La Pampa, as a precipitation estimator. Radar accumulated precipitation was generated in images with spatial resolution of 1 km², using a single relation between reflectivity and precipitation rate. Compounds integrated daily, decadic and monthly were compared with data from a surface rainfall network in the region for three periods: December 2011 to May 2012, August to October 2012, and January to March 2013. As a result A moderate to high correlation was found between radar estimates and baseline records. The linear adjustment between the two datasets was significant for the three periods (0.74, 0.59 and 0.89 $p < 0.05$) and showed a systematic tendency to overestimate the amount of rainfall. Coinciding with a wide bibliographical reference, the accuracy of the estimations weaken with the increase of the distance to the radar, becoming very evident from 90-100

km (RMSE 15,6 mm). The loss of precision was related to larger rains at distances larger than 50 km. The most intense precipitation events were estimated with more dispersion and magnitude of the error than less intense rains. The autumn and spring estimates presented a smaller relative error (1.3 to 2.3%) compared to the summer months (3.6 to 7.6%). It was concluded that the radar estimates characterize the spatial distribution of the precipitations with an acceptable error and almost constant in time.

Keywords: rain, radar reflectivity, estimation errors

INTRODUCCIÓN

La estimación de la lluvia mediante radar meteorológico (RM) es realizada indirectamente a partir de la variable de reflectividad (Z). La tasa o intensidad de precipitación (R) de un evento meteorológico, definida como la cantidad de precipitación ocurrida por unidad de tiempo, puede ser estimada a través de Z. Ambas variables están relacionadas debido a que son dependientes de la distribución de tamaño de gotas, la cual está sujeta a constantes variaciones de un episodio de lluvia a otro, incluso dentro de un mismo episodio esta distribución de gotas presenta variabilidad espacio-temporal (Zawadzki, 1984; Joss y Zawadzki, 1997). A partir de datos experimentales de Z y R se ha comprobado la existencia de una relación empírica entre ambas variables. De esta manera, para estimar R a partir de Z se recurre a este tipo de relaciones matemáticas, denominadas Z-R (Marshall y Palmer, 1948; Rinehart, 2007) y se expresan en una función matemática de tipo: $Z = aR^b$. Los coeficientes a y b de la función son constantes empíricas. R es un producto del contenido másico y de la velocidad de caída en un volumen de medición de radar. No existe una relación única entre la reflectividad del radar y la intensidad de las precipitaciones (PP), ya que los coeficientes a y b varían con el origen y tipo de lluvia, la distribución del tamaño de las gotas, localización geográfica, condiciones meteorológicas, entre otros (Rinehart, 2007). Estos coeficientes pueden determinarse experimentalmente para determinadas situaciones conociendo la verdadera distribución del tamaño de las gotas en un volumen dado.

Los métodos más comunes empleados para la obtención de estos parámetros en la calibración de un RM están basados en la medición directa del diámetro de las gotas de lluvia a partir del uso de disdómetros¹. Históricamente se determinaron muchas relaciones operativas para radares bajo diferentes situaciones geográficas, distintas latitudes y condiciones de tormenta.

La estimación de precipitación mediante radar está sometida a varias fuentes de error. Wilson y Brandes (1979) categorizaron estas fuentes de error como: errores en la estimación de la reflectividad radárica, errores debidos a la incertidumbre en la elección de las relaciones Z-R utilizadas, y el error debido a los diferentes tipos de muestreos de radares con respecto a otros medidores en superficie. El error concerniente a la estimación de la reflectividad radárica está relacionado por un lado a la calibración electrónica del hardware del RM. Esta calibración consiste en el control de todos los parámetros de operación del radar que intervienen el proceso de medición, tales como: la potencia transmitida, la ganancia de la antena, la longitud de onda de la radiación emitida, la longitud del pulso. Por otro lado, las fuentes de error no asociadas al sistema de hardware involucran: el bloqueo del haz, debido a que puede quedar parcial o totalmente bloqueado por la topografía u obstáculos entre el radar y el objetivo, provocando una subestimación de la reflectividad (Harrold y Kitchingman, 1975). La propagación anómala del haz, cuando ciertas condiciones de la atmósfera, tales como inversiones térmicas y presencia de humedad, distorsionan el trayecto del haz del radar haciéndolo refractar hacia el suelo, provocando ecos erróneos (Battan, 1973; Brandes y Sirmans, 1976). La atenuación o pérdida de potencia de la energía emitida a medida que el haz se propaga en la atmósfera, debida a la precipitación y a los gases atmosféricos (Wexler y Atlas, 1963; Blake, 1970), y por último, error debido al incremento de la altura de muestreo con la distancia, y a la variación vertical de la reflectividad. Debido a los efectos de la abertura del haz, de su inclinación y de la curvatura de la tierra, las mediciones de RM se aplican a PP en altitud promediadas sobre una profundidad considerable que dependen

¹ Instrumento óptico láser que es capaz de medir el tamaño de las gotas de lluvia o del hidrometeoro del que se trate, y su velocidad de caída.

de la distribución vertical de la precipitación y pueden contribuir a errores en las estimaciones a nivel de la superficie del suelo (Joss y Waldvogel, 1990). En relación a la incertidumbre en la elección de una relación matemática entre Z y R, que sea adecuada a cada situación, puede ocurrir que al utilizar de forma general una sola relación en la estimación de un evento de precipitación ésta no represente la situación real en ese momento y en ese lugar. Los errores que pueden cometerse han alcanzado hasta un factor de dos con respecto a las mediciones en tierra realizadas con pluviómetros como referencia, (Wilson y Brandes, 1979; Joss y Waldvogel, 1990). Una serie de relaciones empíricas entre Z y R fue presentada por Battan (1973) con más de 70 alternativas obtenidas en diferentes partes del mundo para radares de banda C y para diferentes condiciones meteorológicas. Una forma de mitigar este error, consistiría en clasificar cada volumen de datos del radar en zonas con diferentes tipos de PP, básicamente en estratiformes y convectivas y aplicar la relación específica para cada tipo de lluvia (Churchill y Houze Jr., 1984; Rosenfeld *et al.*, 1995; Sempere Torres *et al.*, 1999). Como antecedente en Argentina, Vidal, *et al.*, (2012), a partir de mediciones *in situ* de las distribuciones del tamaño de gotas mediante disdrómetros, obtuvieron valores de coeficientes a y b proponiendo distintas relaciones entre las variables Z y R bajo diferentes tipos de tormentas, con el objetivo de aplicar dichas relaciones al radar ubicado en Ezeiza (Bs. As.) y derivar un producto de precipitación más adecuado para la región.

En relación al error introducido por el tipo de muestreo de RM respecto a otros medidores en superficie, como los pluviómetros, la principal diferencia se debe al llamado error de representatividad (Kitchen y Blackall, 1992; Moreau *et al.* 2009), y se refiere a la diferencia en la observación que presentan ambos métodos (Zawadzki, 1975; Austin, 1987; Moreau *et al.*, 2009) debido a que un pluviómetro mide el agua efectivamente precipitada, mientras el RM estima el agua potencialmente precipitable a una cierta altura. La necesidad de una base de comparación para evaluar las estimaciones de lluvia por radar, hace que frecuentemente los pluviómetros sean aceptados como referencia (Brandes *et al.*, 1999). El objetivo de este trabajo fue evaluar el desempeño del RM de Anguil como estimador de la precipitación meteorológica. Para ello se estudió la relación entre estimaciones de precipitación acumu-

lada del radar y registros en pluviómetros co-localizados de una red terrestre como referencia.

MATERIALES Y MÉTODOS

Datos de radar meteorológico

Se trabajó con archivos de información proveniente del RM de INTA (Instituto Nacional de Tecnología Agropecuaria) ubicado en la Estación Experimental de Anguil, (36,539 S; 63,990 W).

La calibración electrónica del hardware se realizó durante la instalación y periódicamente se verifica la potencia de la señal emitida. La atenuación de la señal normal que presentan los radares de banda C es mitigada en gran parte mediante parámetros polarimétricos que aplica correcciones a la atenuación del retorno. Este radar estuvo configurado para completar una serie de giros a 360° que se repite para 12 ángulos de elevación, entre 0,5° de base y 15,1° de tope en un rango de 240 km (Hartmann, *et al.*, 2010). La frecuencia de escaneo se programó cada 10 minutos, y bajo funcionamiento normal totalizó 144 adquisiciones diarias. Como resultado de cada adquisición se generó una cantidad de datos numéricos que caracterizan un volumen tridimensional de la atmósfera situado dentro del haz de radiación, en el rango de los 240 km con centro en la antena. Estos datos corresponden a variables convencionales y polarimétricas que fueron almacenados en archivos, denominados volúmenes, en servidores *in situ*. En este trabajo se utilizaron los volúmenes que contenían la variable: Z reflectividad, a los cuales se accedió por vía remota.

Se descargaron archivos con resolución de 10 minutos, durante 3 períodos: primer día de diciembre 2011 al 30 de mayo 2012; primer día de agosto al 31 de octubre de 2012 y primer día de enero al 30 de marzo de 2013. Estos archivos presentaban un formato caracterizado por una sección XLM (Extensive Markup Language) y una sección de datos en formato binario con compresión, lo que requirió realizar una transformación a un formato ASCII (American Standard Code for Information Interchange) para lectura e interpretación (Hartmann, *et al.*, 2010).

A partir de los archivos de Z en mm⁶/m³ se estimó la R en mm/h. Para ello se utilizó una relación matemática entre estas variables y se mantuvo fija, correspondió a la propuesta por Marshall y Palmer

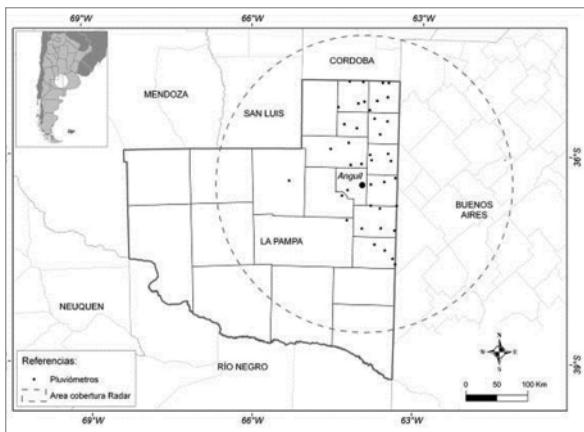


Figura 1. Distribución de los pluviómetros o estaciones utilizadas para la validación.

(1948), cuyos parámetros se configuraron a partir del software Rainbow® 5 según protocolo del fabricante (Gematronik, 2005). Posteriormente se generaron los productos hidrológicos SRI (*Surface Rain Intensity*), algoritmo que genera una imagen con valores asociados de intensidad de precipitación en mm/h estimada para cada celda o unidad mínima de la misma. A partir de la integración de los SRI se obtuvo el producto PAC (*Precipitation Accumulation*). Este fue el dato final analizado en este trabajo que representa un acumulado de lluvia en mm en imágenes de 8 bits que constan de 240 x 240 pixeles con una resolución espacial de 1 km² y 1°, por píxel o celda y resolución temporal de 10 minutos.

Datos de estaciones pluviométricas

Se utilizaron en este trabajo datos de PP diaria proveniente de la red pluviométrica de la provincia de La Pampa. Esta red está compuesta por 95 puntos de observación y posee pluviómetros (Pv) oficializados por el SMN. Para la realización de este trabajo fueron seleccionados 40 Pv (estaciones). La selección se realizó descartando aquellas estaciones que; estaban fuera del área de cobertura del radar, presentaban registros incompletos o, no cumplían estrictamente con las normas de instalación y mantenimiento de los pluviómetros. Para cada estación seleccionada fueron registradas sus coordenadas geográficas mediante un dispositivo de sistema de posicionamiento global (GPS) (Fig. 1).

Los datos de precipitación correspondieron a registros diarios medidos en mm observados a las 9.00 a.m., correspondientes a un día pluviométrico (9

a.m a 9 a.m) en Pv homologado de acuerdo a normas internacionales de la Organización Meteorológica Mundial (OMM) y según protocolo de observación del Servicio Meteorológico Nacional (SMN). Fueron excluidos aquellos registros que presentaron inconsistencias de acuerdo a las características de cada suceso de PP ocurrido en la zona de ubicación de cada estación. Para esto se llevó un registro de cada evento de lluvia o tormenta ocurrido durante los períodos de selección de información.

Metodología

Se generaron compuestos de PP acumulada en tres integraciones temporales: diarios (24-horarios), decádicos (diez días) y mensuales. Para los compuestos diarios se integraron 144 imágenes obtenidas desde las 9 a.m a las 9 a.m del día siguiente, (coincidente con el registro pluviométrico diario medido en las estaciones). Los datos constituidos con menos de 144 imágenes fueron descartados. Los compuestos decádicos se obtuvieron integrando los compuestos diarios en diez días. Los compuestos mensuales se obtuvieron integrando los compuestos diarios del día 1º hasta el día 28, 30 ó 31 según el mes considerado. En los casos donde existió faltante de imágenes radar se consideraron días incompletos y fueron reemplazados con la lectura de los Pv correspondiente a ese día. Los compuestos de radar fueron divididos en 3 grupos según distancia al centro: hasta 50 km; entre 51 km y 100 km, y más de 100 km.

A partir de registros pluviométricos de estaciones y análisis visual de imágenes de z se seleccionaron algunas fechas con ocurrencia de eventos de PP de diferente intensidad, durante el primer periodo de datos. Sobre las fechas elegidas se generaron los productos SRI y se diferenciaron en dos grupos: lluvias débiles a aquellos eventos donde los SRI fueron menores a 10 mm/h y otro grupo de lluvias denominadas intensas que presentaron SRI mayor a 10 mm/h, tomando como referencia esos umbrales propuestos por Nzeukou *et al.* (2004). Posteriormente se generaron los PAC diarios para cada fecha. Se seleccionaron los datos de PP acumulada correspondientes a los Pv involucrados en el área contemplada por cada evento.

Para coincidir ambas medidas (pixel radar y Pv) en la misma ubicación geográfica, se ubicó el píxel correspondiente de la imagen radar centrado en

cada Pv de la muestra. Los compuestos de radar se validaron contra datos de estaciones pluviométricas, y se analizó la relación entre estimaciones de radar y datos de referencia:

- Para tres períodos sucesivos con datos integrados diarios.
- En tres rangos de distancia al radar con datos integrados diarios del periodo 1.
- En dos intensidades de precipitación con datos integrados diarios del periodo 1.
- En distintos meses del año, con datos integrados mensuales de los períodos 1 y 2.
- Con tres integraciones temporales (diarios, decádicos y mensuales) con datos del periodo 1.

Se realizó un análisis exploratorio de datos, con el objetivo de detectar la existencia de datos atípicos y la distribución de las variables. Se construyeron gráficos de dispersión, mediante éstos se evaluó la linealidad de la relación entre ambas medidas. Se realizaron ajustes lineales mediante el método de mínimos cuadrados, considerando los datos de estaciones pluviométricas como variable regresora y su respectiva estimación por RM como variable dependiente. Se calculó el coeficiente de determinación (R^2). Para analizar el grado de exactitud de los ajustes se trazaron líneas de regresión 1:1 entre ambas variables. Los estadísticos calculados para analizar el grado de exactitud de las estimaciones fueron los parámetros de ajuste lineal (pendiente y ordenada al origen). En todas las comparaciones efectuadas se calculó el coeficiente de Pearson (r) para medir el grado de correlación lineal entre medidas estimadas y de referencia. Para evaluar la significancia de los coeficientes de correlación, y parámetros de ajuste se empleó la prueba estadística t-Student. Se calcularon los índices:

RMSE, RMSE relativo, y BIAS o error sistemático medio, el error cuadrático medio y la desviación estándar o STDE.

RESULTADOS Y DISCUSIÓN

Relación entre estimaciones de radar y datos de referencia

Los estadísticos de posición calculados para datos integrados diarios de radar superaron a los

correspondientes a los Pv en todos los períodos analizados. El rango de datos de radar fue superior al de los Pv. La distribución de los datos tanto de RM como de Pv presentó asimetría hacia la derecha, se corroboró una distribución distinta a la normal mediante prueba de Shapiro-Wilk.

La correlación entre estimaciones radar y medidas de referencia resultaron estadísticamente significativas ($p < 0,05$) en los tres períodos ($r = 0,87, 0,77$ y $0,89$ respectivamente) los ajustes lineales explicaron entre un 59% y 89% de la varianza común entre ambas medidas. Estos resultados fueron coherentes con lo hallado en la bibliografía referida a RM de doble polarización con datos 24-horarios. Holleman, (2006) evaluando productos de software Rainbow con radar de banda C en Holanda, obtuvo correlaciones similares con datos sin corregir y corregidos por atenuación, ($r = 0,79$ y $0,87$) respectivamente. Moreau, et al., (2009) en Francia obtuvo correlación de $0,8$, y $0,85$ a $0,92$ aplicando un algoritmo de corrección por atenuación. Brandes et al., (1998) obtuvo correlaciones entre $0,78$ y $0,95$. También con radar en banda C, Hannesen y Gysi (2002) hallaron correlaciones entre $0,48$ y $0,64$ con distintos algoritmos de acumulación. Sin embargo en la misma región geográfica donde se obtuvieron los datos para este trabajo, Della Chiesa, (2011), con radar de simple polarización (SP) ubicado en Pergamino, Argentina, no obtuvo buen ajuste con datos de estaciones pluviométricas ($R^2 = 0,18$). Con el mismo radar SP y el radar ubicado en Paraná de DP, Hartmann, et al. (2010) hallaron alta correlación ($r = 0,92$ para ambos) pero con datos integrados mensualmente. La mejora en los valores de correlación obtenidos con radares de DP podría ser parcialmente explicada debido a que éstos utilizan mayor cantidad de variables relacionadas con z (variables polarimétricas) para estimar la lluvia. Estas medidas proporcionan información adicional relacionada con las características de la distribución del tamaño de las gotas proporcionando mayor precisión en las estimaciones de la PP (Bringi, et al., 1982; Zrnic y Ryzhkov, 1999). En comparación con información pluviométrica realizada con teledetección satelital Magliano et al., (2012) con productos de radar montado sobre satélite (TRMM, NASA-JAXA) y sobre la misma región estudiada no encontraron buenas correlaciones con datos de estaciones pluviométricas a escala diaria ($r = 0,1$).

A partir de los diagramas de dispersión de la figura 2 se observó la linealidad de los valores de PP

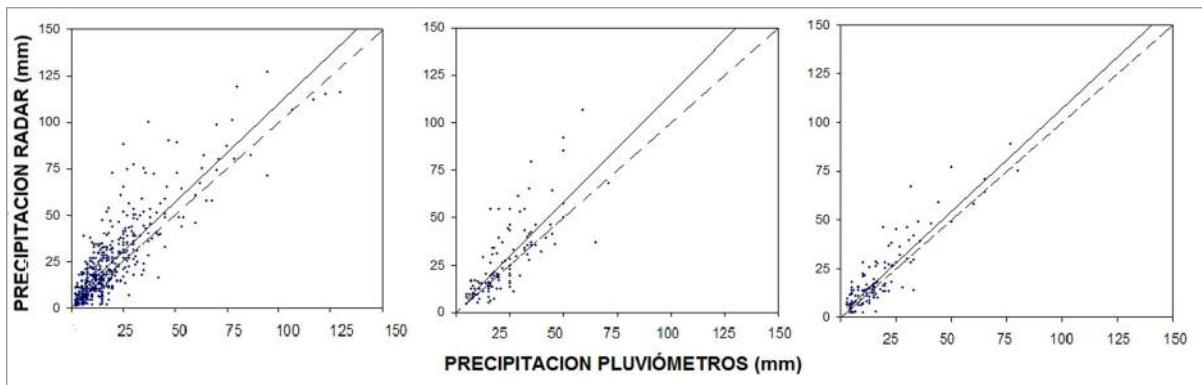


Figura 2. Gráficos de dispersión y ajuste lineal (línea llena) entre la precipitación diaria estimada por radar y la observada en pluviómetros de referencia. La línea punteada representa la relación 1:1

estimados respecto a la referencia. Ambos conjuntos de datos presentaron ajuste lineal significativo ($p < 0,05$) en los tres períodos estudiados, lo que implica que las rectas ajustadas presentaron pendientes estadísticamente distintas de 0. Las pendientes de las rectas obtenidas para cada período no mostraron diferencias estadísticamente significativas al 5%.

Los valores de error sistemático medio hallados (Tabla 1) son coherentes con la literatura, Harrison, *et al.*, (2000) con RM de banda C, obtuvo valores de 0,83 a 1,3 con datos no corregidos y de 0,28 luego de aplicar correcciones por atenuación y Perfil Vertical de Reflectividad (PVR). Sin embargo con datos radar sin calibraciones previas el valor de BIAS muy frecuentemente supera el factor de 2 o más (Harrold *et al.*, 1974; Wilson y Brandes, 1979; Zawadzki, 1984; Joss y Waldvogel, 1990; Meischner *et al.*, 1997).

Mediante los parámetros de ajuste y error sistemático (signo positivo para los tres períodos) se ha podido identificar una tendencia en la relación entre los pares de datos (Hannesen y Gysi, 2002) que resulta en una sobreestimación global del radar de la cantidad de lluvia respecto a la referencia, entre un 17 y 26% y la existencia de una componente sistemática del error. Esta sobreestimación se confirmó estadísticamente mediante prueba t de diferencia de medias para los tres períodos ($p < 0,05$). Con los radares ubicados en la región pampeana argentina, Hartmann, *et al.*, (2010) encontró subestimación de los valores de lluvia con el radar de SP de Pergamino y sobreestimación con el radar de DP de Paraná, en un solo período de datos y con acumulaciones mensuales.

La tendencia a la sobreestimación se verifica para los tres períodos sucesivos analizados en este trabajo, lo que permitió inferir que no existieron variaciones marcadas del error sistemático en el tiempo. Dentro de las posibles causas que podrían explicar la tendencia a la sobreestimación en todo el dominio de datos, se encuentra la posible inexactitud de la constante de calibración del radar (Smith y Krajewski, 1991; Kitchen y Blackall 1992; Smith *et al.*, 1996; Anagnostou, *et al.*, 1998). Más precisamente la incertidumbre de la correcta intensidad de la potencia de emisión, causando sobreestimación de z con una consecuente sobreestimación de la lluvia. Otras causas posibles de sobreestimación de z es la presencia de lo que se denomina banda brillante (Vignal *et al.*, 1999; Bordoy, *et al.*, 2009); la presencia de precipitación sólida o granizo (ambos relacionados con el tipo de precipitación), y la propagación anómala (Rinehart, 2007). Sin embargo, estas condiciones se producen de manera aleatoria y no constantes en el tiempo y en el espacio como para determinar un sesgo sistemático.

A partir de los gráficos de la figura 2 se observaron variaciones en la dispersión del error (R-P) a lo largo de las rectas, con un marcado aumento a partir de valores de precipitación acumulada diaria de 20 mm aproximadamente.

La dispersión del error, en todo el dominio de

Tabla 1. Parámetros de error sistemático medio y ajuste lineal para los tres períodos de estudio.

Período	BIAS	β_0	β_1	R^2
1	0,26	3,74	1,06	0,74
2	0,19	1,12	1,13	0,59
3	0,17	1,05	1,9	0,89

datos y principalmente a partir de determinado valor de lluvia acumulada permitió inferir que existieron efectos aleatorios causantes de error en las estimaciones, además de la existencia de una componente sistemática. Los efectos aleatorios están relacionados con múltiples factores de error en las estimaciones de radar, bien estudiados en la bibliografía sobre el tema. Dentro de ellos podemos mencionar a los relacionados a la variabilidad de los eventos que ocurrió en cada periodo, a las diferentes distancias y alturas de detección de la lluvia, a la intensidad variable de la precipitación entre eventos y en un mismo evento, al tipo de PP (estratiforme, convectivas o ambos al mismo tiempo) y a la atenuación de la señal de emisión, relacionada con las características del mismo radar y con la intensidad y tipo de PP ocurridas durante el estudio (Sauvageot, 1994; Krajewski y Smith, 2002; Lee *et al.*, 2007; Rinehart, 2007).

Distancia al radar

Los ajustes lineales por mínimos cuadrados resultaron significativos para los tres grupos de distancia analizados ($p < 0,05$). Las estaciones ubicadas a menos de 50 km de distancia presentaron un mejor grado de precisión. No se encontraron diferencias estadísticamente significativas cuando se compararon las pendientes de las rectas de ajuste de los tres grupos ($p > 0,05$).

Se evidenció un notorio aumento de la dispersión del error en las estaciones ubicadas a más de 50 km. a partir de un umbral aproximado a los 20 mm diarios. Esto se confirmó a través de los índices de dispersión (STDE) que figuran en la tabla 2. Estos resultados permitieron inferir que en las estaciones más alejadas, las lluvias más intensas se estimaron con mayor variabilidad que en las estaciones más cercanas al radar.

Múltiples investigaciones coincidieron en una fuerte implicancia del factor distancia sobre las estimativas de lluvia de RM. Wilson, (1976); Fabry *et al.*

Tabla 2. Índices de error y dispersión calculados por estación y promediados según tres distancias. Datos del periodo I.

km	Hasta 50	Entre 51 y 100	Más de 100
RMSE (mm)	9,5	10,5	15,6
RMSE rel (%)	4,9	5,3	5,8
STDE rel (%)	4,8	5,2	5,4

(1992); Kitchen y Jackson (1993); Koistinen *et al.* (2003), concluyeron en que las diferencias en milímetros estimadas por radar y registradas en Pv en la misma ubicación, aumentan a medida que las estimaciones se efectúan más lejos al sensor. Bech, *et al.* (2007) en el norte europeo, y Bordoy, *et al.* (2010) en España observaron cómo en general el error sistemático respecto a medidas pluviométricas aumentó a medida que se incrementó la distancia al radar. Kitchen y Jackson (1993) en el Reino Unido, y Llort (2010) en España, encontraron que la correlación lineal en comparaciones radar-pluviómetro decrecía conforme aumentaba la distancia al radar. El rango de valores de RMSE para todas las distancias evaluadas (12 a 172 km) fue de 6,2 a 22,2 mm. El índice calculado para las estaciones ubicadas a más de 100 km presentó diferencia estadísticamente significativa respecto a los índices obtenidos con los otros dos grupos a distancias menores (Tabla 2). Los RMSE obtenidos fueron coherentes con valores citados en la literatura con radares de DP y acumulaciones 24-horarios, por ejemplo Jayakrishnan *et al* (2004) en EE.UU obtuvo RMSE entre 4 y 24 mm, Bordoy *et al* (2010) obtuvo errores cuadráticos entre 2,14 a 10 mm y 8 a 28,7 mm durante episodios invernales y estivales respectivamente en comparaciones de diferentes configuraciones de productos estimativos de lluvia con RM en España.

Se observó una tendencia al aumento de las diferencias, en términos de RMSE, a medida que las comparaciones se alejan del radar y un incremento de la dispersión muy marcado a partir de 90 km aproximadamente (Fig. 3).

Resultados similares obtuvo Pastor García (2013) los RMSE aumentaban con la distancia y los mayo-

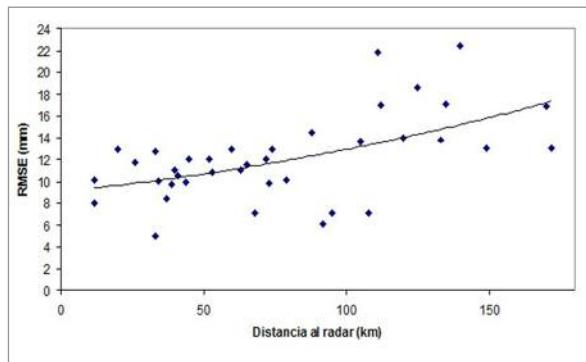


Figura 3. Representación de los RMSE calculados para las estaciones de la muestra en función de la distancia al radar, con datos del periodo I. La línea representa tendencia.

res índices los halló entre 80 y 130 km, con un máximo de situaciones en 110 km de distancia al radar. Kitchen y Jackson, (1993) concluyeron que a distancias mayores a 100 km las medidas de referencia superaban en un valor de 2 a las estimaciones. Collier (1986) observó que a partir de 80 km el radar subestimaba la PP registrada por los Pv en forma creciente con la distancia, más precisamente en un factor de 1:4 a 120 km y 1:8 a 160 km. Similares resultados obtuvo Llort (2010), quien corroboró que a partir de 20 km de distancia comenzaba un aumento sostenido de la dispersión del error. Afirman que existe una distancia umbral a partir de la cual un RM pierde significativamente su precisión como estimador de lluvia es un tema controvertido en la investigación existente, debido a que se ha demostrado que existen múltiples factores influyentes para cada situación particular que hacen variar ampliamente la distancia umbral (Collier, 1986; Joss y Waldvogel, 1990; Fabry *et al.*, 1992; Kitchen y Jackson, 1993), sin embargo son numerosos los trabajos que coinciden en que a distancias mayores a 100 km no deben esperarse estimaciones cuantitativas precisas con radar. Para el menor ángulo de elevación ($0,5^{\circ}$) el haz de radiación alcanza a 1,5 km de altura respecto del nivel del mar, a 100 km de distancia al radar, y aproximadamente a 6 km de altura a 250 km de distancia (Rinehart, 2007), por eso, al tomar directamente las medidas del radar como representativas del valor de z en el suelo se cometen errores cuya magnitud depende de la importancia de la variación vertical de la reflectividad.

Es amplia la bibliografía sobre el tema que argumenta el efecto distancia como distorsivo de las estimaciones de radar. Las estimaciones de lluvia a nivel del suelo se infieren de los datos de z medidos por el radar en la primera elevación del barido volúmetrico, lo que se conoce como PVR. A medida que aumenta la distancia al sensor se produce un incremento de la altura de la estimación y por consiguiente en el PVR ocurren fuertes variaciones debido a fenómenos como evaporación o incremento de la PP, corrientes verticales o cambios de fase de las partículas de agua. El error causado por el incremento de la altura de muestreo depende de la variación vertical de la reflectividad entre la altura muestreada y la superficie. Cabe esperar que cuanto mayor sea la altura de muestreo, esto es, cuanto más alejado del radar se encuentre el volumen de muestreo, mayor sea la posibilidad de una diferencia de z importante

entre los dos niveles (Fatorelli, 1999; Franco 2008).

El RM del estudio utilizó una configuración teórica de PVR y se mantuvo constante en la distancia y en el tiempo. Es probable que esta configuración obtenga diferentes resultados al variar la distancia de estimación o no sea la más adecuada para cada distancia, ya que no existen estudios preliminares efectuados para el radar de Anguil con distintas configuraciones de PVR para evaluar su desempeño. Numerosos métodos se han propuesto en la investigación sobre el tema para la corrección del error asociado a la variación vertical de la reflectividad con la distancia (Zawadzki *et al.*, 1986; Koistinen, 1991; Joss y Lee, 1995; Franco, 2008), y su desarrollo depende de las características del propio radar y la climatología de la zona donde está instalado. La corrección del PVR ha sido incorporada a la operación de varias redes de RM como la de Suiza (Joss y Lee, 1995) o España (Aguado, 1999).

Intensidad de la precipitación

A partir de los gráficos de dispersión y ajuste lineal de la Fig. 2 fue posible inferir que la dispersión del error estuvo relacionada con los montos de lluvia, haciéndose más evidente un aumento a partir de los 20 mm diarios. Tal como afirman cuantiosos trabajos relacionados con estimación de lluvia por radar, la intensidad de precipitación junto con las características de cada evento en particular son variables que afectan los resultados de las estimaciones de RM (Harrold, *et al.*, 1974; Wilson y Brandes, 1979; Klazura y Kelly, 1995). De manera general la intensidad de la precipitación está relacionada con los montos acumulados en superficie, las lluvias más intensas acumulan mayores montos en un tiempo considerado respecto de lluvias con menor intensidad en el mismo periodo. Considerando los umbrales de intensidad previamente definidos en la metodología, se analizaron de manera individual distintos eventos de lluvia con una duración definida y que afectaron un área determinada, involucrando varios Pv de referencia, para el análisis de cada una por separado. La relación P/R promedio tomó valores por debajo de 1 en todos los eventos analizados, confirmando la sobreestimación de la lluvia por radar independientemente de la intensidad y características de cada evento. La dispersión relativa de la relación P/R resultó mayor para eventos de lluvias intensas (Tabla 3).

Tabla 3. Análisis de tormentas individuales según intensidad: a) lluvias débiles (< 10 mm/h); b) lluvias intensas (> 10 mm/h). P/R promedio= Promedio de los ratios entre las medidas de PV y la estimación correspondiente en el pixel de la imagen de RM en cada evento. CV= coeficiente de variación.

a) Tormenta	Nº de pluviómetros	P/R promedio	Dispersion relativa de P/R	
			CV (%)	CV promedio (%)
1	13	0,88	34,9	39,9
2	20	0,96	48,4	
3	13	0,8	36,3	
4	13	0,92	55,6	
5	16	0,97	41,3	
6	14	0,84	23,3	

b) Tormenta	Nº de pluviómetros	P/R promedio	Dispersion relativa de P/R	
			CV (%)	CV promedio (%)
1	13	0,81	40,2	51,4
2	12	0,98	61	
3	15	0,91	52	
4	12	0,65	47,5	
5	13	0,79	61,5	
6	11	0,82	46	

De acuerdo con Lee *et al.*, (2006) a mayor intensidad de lluvia más interferencias se producen en la señal del radar y mayor error de estimación se produce. Varios autores coincidieron en que el ratio P/R típicamente varía de tormenta a tormenta cuando la precipitación acumulada se estima a partir de la reflectividad radar (Harrold *et al.*, 1974; Wilson y Brandes, 1979; Klazura y Kelly, 1995) y frecuentemente alcanza a un factor de 2 (Sauvageot, 1994).

La variación del error de estimación que se produce evento a evento de lluvia está relacionada a la variabilidad temporal y espacial en la distribución de tamaño de las gotas. Para estimar la tasa de PP a partir de z se aplica en la configuración del producto de lluvia por radar, una relación funcional Z-R de la forma: $Z = aR^b$. Los coeficientes a y b dependen de la distribución de diámetros de las gotas y por lo tanto de las características del evento climatológico. Según Brandes *et al.*, (1998) y Rosengaus (1999) la variación del error de estimación que se produce evento a evento, está relacionada al uso de una misma relación Z-R en tormentas que difieren en sus características, estableciendo un límite definitivo a su precisión. Wilson y Brandes (1979) afirmaron que la utilidad de una sola relación de conversión de Z a R en diferentes tipos de tormentas está indicada por la variabilidad en la relación P/R. De acuerdo con esto y los resultados obtenidos en este trabajo, la relación Z-R utilizada en la configuración del producto evaluado, en principio estimaría de manera más homogénea las lluvias débiles que las lluvias de mayor intensidad, considerando que los coeficientes utilizados en la configuración del producto radar

son aplicables a eventos de tipo estratiforme en latitudes medias (Marshall y Palmer, 1948). Sin embargo los eventos muy intensos presentarían errores considerables de estimación. En la Región Pampeana los eventos de mayor intensidad de origen convectivo son producidos en estación primavero-estival (Hoffman, 1988; Roberto *et al.*, 1994; Sarochar *et al.*, 2005) siendo los que aportan milímetros de agua considerables para su utilización en actividades agropecuarias. Como único antecedente en la Región Pampeana Argentina, y con el RM ubicado en Ezeiza (Bs.As.), Vidal *et al.* (2012) a través de regresiones lineales con las variables Z y R, y mediante la utilización de disdrómetros, determinaron coeficientes 306,8 para a y 1,57 para b, considerando las condiciones climatológicas dominantes de la región.

Estación del año

Con la integración mensual del dato se evaluó el comportamiento de las estimaciones en diferentes meses del año. Durante el periodo 1 el RMSEr presentó variaciones importantes entre los meses considerados en este análisis, tomando valores más altos en enero, febrero y marzo, siendo los meses donde mayor cantidad de milímetros aportaron las lluvias. Dicho de otro modo en los meses donde ocurrieron mayores PP, la diferencia entre estimaciones radar y registros de referencia fue mayor. Esto fue un indicio de que existió una relación entre la precisión del radar y el nivel de precipitación registrado. El mes de febrero presentó una marcada diferencia en la dispersión del error respecto a los otros meses, este comportamiento fue atribuido a la ocurrencia de varios eventos de mayor intensidad de PP durante este mes y de tormentas con granizo y vientos intensos, que causaron mayores interferencias en la señal radar. Por otro lado la mayor diferencia entre las estimaciones y medidas de referencia también podrían explicarse en el hecho de que los pluviómetros miden con mayor error la columna de agua en presencia de viento y granizo, incrementando el error de representatividad, debido a que no captan en un solo punto de medición la variabilidad espacio-temporal de la PP (Kitchen y Blackall, 1992). Los meses de otoño y primavera presentaron los mínimos errores relativos y coeficientes de dispersión (Tabla 4). Della Chiesa (2011) en la región pampeana, estudiando la variación intra-anual de la preci-

Tabla 4. Coeficientes de correlación, error relativo y dispersión del error según meses para datos mensuales en el periodo 1 y 2. (MPP pluv: monto promedio de PP registrado por Pv)

	Periodo 1						Periodo 2	
	DIC	ENE	FEB	MAR	ABR	MAY	SEP	OCT
r	0,81	0,83	0,85	0,87	0,76	0,87	0,77	0,72
RMSEr (%)	3,6	5,1	7,6	4,9	4,1	2,3	1,3	2,1
CV error (%)	91,3	77,3	119,7	92,3	92,2	89,7	46,3	58
MPP pluv(mm)	32	115	157	96	53	33	20	28

sión del RM de Pergamino obtuvo los mejores resultados durante los meses de septiembre y octubre y los peores resultados en febrero. En España Bordoy *et al.* (2010) halló los mayores errores cuadráticos entre estimativas de RM y medidas pluviométricas terrestres durante episodios de verano, que en varios casos superaban en un factor de 3 a los índices calculados en episodios invernales. Asimismo, el fenómeno de atenuación o pérdida de potencia de la señal a medida que se propaga en la atmósfera es intensificado durante las PP intensas y también en presencia de granizo (Battan, 1973; Zawadzki, 1984; Rinehart, 2007). Este efecto podría explicar la mayor dispersión del error ocurrido durante y entre los episodios ocurridos en la estación estival, ocasionando subestimaciones y sobreestimaciones de la reflectividad de manera alternada. La atenuación de la señal debida a la precipitación en radares de banda C, (como el estudiado en este trabajo) es muy significativa pudiendo alcanzar valores del 80% para intensidades de lluvia de 5 mm/h (Zawadzki, 1984).

Integración temporal del dato

Con la integración mensual del dato se obtuvo el mejor grado de precisión y los menores índices de error (Tabla 5). Correlaciones similares habían hallado Hartmann *et al.* (2010) con los radares ubicados en la misma región geográfica para acumulaciones mensuales. Della Chiesa (2011) con el radar ubicado en Pergamino halló diferencias al integrar temporalmente los datos, no encontrando buenos ajustes con acumulaciones diarias, pero con un notable aumento de la precisión al integrarlas semanal, quincenal y mensualmente.

La lluvia es caracterizada por una fuerte irregularidad en el tiempo, por eso a medida de que se integren las observaciones tenderá a mejorar la precisión de las estimaciones. La integración en 10 días no presentó diferencias significativas respecto al acumulado diario, aunque obtuvo un menor índice

Tabla 5. Coeficiente de determinación de ajuste lineal y error cuadrático medio relativo para distintas integraciones temporales durante el periodo 1

INTEGRACIÓN	R ²	RMSE rel (%)	BIAS
DIARIO	0,74	5,3	0,26
DECÁDICO	0,76	6,4	0,26
MENSUAL	0,9	1,7	0,24

de error relativo. Si bien la integración temporal mejora el dato en este trabajo se encontró una aceptable precisión a escala diaria. Los acumulados diarios pueden ser indicador razonablemente bueno de la precisión del algoritmo de estimación de PP con RM (Gjertsen *et al.*, 2003).

CONCLUSIONES

Se evaluó el desempeño del RM de Anguil como estimador de la precipitación en su área geográfica, bajo condiciones meteorológicas zonales. Se estudió la relación entre datos radar y referencia pluviométrica de una red terrestre, bajo su cobertura espacial, en tres períodos sucesivos. Se halló una correlación moderada a alta entre ambos conjuntos de datos. Los buenos ajustes encontrados entre los valores de PP estimados respecto de los valores de referencia revelaron en principio la capacidad del RM para representar la distribución espacial de la lluvia en su área de cobertura, aportando información confiable principalmente en aquellas regiones donde la red de estaciones meteorológicas es poco densa o inexistente. Se verificó una tendencia a la sobreestimación de la cantidad de lluvia de manera sistemática, esto se atribuyó a deficiencias en la calibración electrónica del hardware radar que sobreestima la reflectividad de los blancos interceptados y en consecuencia los valores de lluvia, como un error sistemático y poco variable en el tiempo. A partir del análisis de los resultados se evidenció la variabilidad espacio-temporal de los errores de estimación lo que puso de manifiesto la existencia de diversos factores que interfieren en el proceso de medida de la reflectividad radar. El estudio se dirigió a evaluar el efecto de la distancia en las estimaciones de lluvia y se concluyó que la precisión del radar desmejora con el incremento de la distancia al mismo, haciéndose muy evidente a partir de los 90-100 km, en coincidencia con una amplia referencia bibliográfica. La mejor performance del radar se

obtuvo en un radio menor a 50 km. A distancias mayores a ésta, la pérdida de precisión del radar se relacionó con lluvias más cuantiosas. Los eventos de precipitación de mayor intensidad fueron estimados con una mayor dispersión y magnitud de error respecto de las lluvias más débiles. Según los resultados de este trabajo se infiere que, en parte este error podría atribuirse a la utilización de una sola relación de conversión de reflectividad a intensidad de PP considerando que la distribución de los tamaños de gotas y las características de los eventos de lluvia zonales varían constantemente en el espacio y en el tiempo. Del análisis obtenido en los distintos meses del año se concluyó que durante la estación de verano la calidad de la estimación del radar empeora respecto de la estaciones otoño y primavera. Esto se explicó por las características climatológicas de las tormentas ocurridas durante el periodo estival evaluado, con lluvias de moderada a alta intensidad, y con varios casos asociados a la presencia de granizo y vientos intensos. Dichas tormentas producirían mayores interferencias en la señal radárica, atenuación de la potencia emitida y mayor ruido en la estimación de reflectividad.

Al aumentar la integración temporal del dato se mejoró la precisión de la estimación a partir de radar. Este hecho podría explicarse por la fuerte irregularidad que caracteriza a la lluvia tanto en el espacio como en el tiempo; por lo tanto a medida de que se integren observaciones, los valores acumulados tenderán a ser más homogéneos, aumentando la confiabilidad de las estimaciones. Sin embargo la precisión del dato en resolución diaria es considerada aceptable a partir de los buenos índices de correlación y ajustes hallados.

A partir de los resultados se concluyó que será posible caracterizar la distribución de las precipitaciones con margen de error conocido, en resolución espacial de 1 km² y temporal diaria o mayor. La información de lluvia aportada por este radar permitirá observar de manera confiable cómo se distribuyen los campos de precipitación, mejorando los resultados de observaciones puntuales por pluviómetros ante un patrón de alta variabilidad o irregularidad en la ocurrencia de lluvias. Bajo algunas condiciones como durante la época estival y lluvias provenientes de fuertes tormentas asociadas a granizo, como a distancias mayores a los 100 km, las estimaciones serán menos confiables.

REFERENCIAS BIBLIOGRÁFICAS

- Belmonte, M.L. 2014. Estimación de precipitación a partir de radar meteorológico: evaluación y desarrollo de productos para la toma de decisiones en planificación agropecuaria. Tesis presentada para optar al grado de Magíster en Ciencias Agrarias. Universidad Nacional del Sur.
- Aguado, F. 1999. Corrección por Perfil Vertical de Reflectividad de los datos de precipitación radar. IV Simposio de Predicción. Serie Monografías. Ministerio de Medioambiente de España
- Anagonostou, E.N., Krajewski, W.F. Seo, D.J., Jhonson, E.R. 1998. Mean field rainfall bias Studies for WSR-88D.J Hydrol. Eng., 3, (pp. 149-159).
- Austin, P.M. 1987. Relation between measured radar reflectivity and surface rainfall. Monthly Weather Review, 115, (pp. 1053-1069).
- Battan, L.J. 1973. Radar Observation of the atmosphere. The University of Chicago Press, Chicago, III, 324 pp.
- Bech, J., Gjertsen, U., Haase, G. 2007. Modelling weather radar beam propagation and topographical blockage at northern high latitudes. Quarterly Journal of the Royal Meteorological Society, 133(626), (pp. 1191-1204).
- Blake, L. V., 1970. Prediction of radar range. Radar Handbook, edited by M. I. Skolnik, (pp. 2-51- 2-55). McGraw-Hill, New York.
- Bordoy, R., Bech, J., Rigo, T., Pineda, N. 2010. Analysis of a method for radar rainfall estimation considering the freezing level height. Journal of Mediterranean Meteorology & Climatology, 7, (pp 25-39).
- Brandes, E.A. Sirmans, D. 1976: Convective rainfall estimation by radar: Experimental results and proposed operational análisis technique. Preprints, Conference on Hydroclimatology (Ft. Worth), AMS, Boston, (pp. 54-59).
- Brandes, E. A., Vivekanandan, J., Wilson, J. W. 1999. Comparison of radar reflectivity estimates of rainfall from collocated radars. J. Atmos. Oceanic Technol. 16, (pp 1264–1272).
- Bringi, VN. Goddard, J, Cherry, SM. 1982. Comparison of dual polarization radar measurements of rain with ground based disdrometer measurements. J Appl Meteor. 21, (pp 252–64).
- Churchill, D. D., Houze Jr., R. A. 1984: Development and structure of winter monzón cloud clusters on 10 December 1978. Journal of the Atmospheric Sciences, 41, (pp 933-960).
- Collier, C. 1986. Accuracy of rainfall estimates by radar, Part I: Calibration by telemetering raingauges. Journal of Hydrology, 83(3), (pp. 207-223).
- Della Chiesa, T. 2011. Estimación del balance de agua del suelo en la región pampeana utilizando datos provenientes

- de sensores remotos. Trabajo correspondiente al ciclo de intensificación para acceder al título de ingeniero agrónomo. Universidad de Buenos Aires. Facultad de Agronomía
- Fabry, F., Austin, G. L., Tees, D., 1992. The accuracy of rainfall estimates by radar as a function of range. *Quart. J. Roy. Meteor. Soc.* 118, (pp 435–453).
 - Franco, M. 2008. Estimación cuantitativa de la lluvia mediante radar meteorológico. Corrección del error asociado a la variación vertical de la
 - Gematronik, 2005. Rainbow® 5 Products & Algorithms. Gematronik GmbH., Raiffeneisenstr. 10, 41470 Neuss, Germany
 - Gjertsen, U., Salek, M., Michelson, D. B. 2003. Gauge-adjustment of radar-based precipitation estimates—a review. COST-717 working document no. WDD, 2(200310), 1.
 - Hanesen, R., Gysi, H. 2002. An enhanced precipitation accumulation algorithm for radar data. In Proc. ERAD (pp. 266-271)
 - Harrison, D. L., Driscoll, S. J., Kitchen, M. 2000. Improving precipitation estimates from weather radar using quality control and correction techniques. *Meteorological Applications*, 7(2), (pp.135-144).
 - Harrold T.W., Kitchingman, P.G. 1975. Measurement of surface rainfall using radar when the beam intersects the meeting layer. Preprints, 16th radar Meteorology conference (Houston), AMS, Boston, (pp. 473-478)
 - Harrold, T. W., E. J. English. Nicholass, C.A. 1974: The accuracy of radar-derived rainfall measurements in hilly terrain. *Quart. J. Roy. Meteor. Soc.*, 100, (pp. 331–350).
 - Hartmann, T., Tamburrino, M. S., Bareilles, F. 2010. Análisis preliminar de datos obtenidos por la red de radares del INTA para el estudio de precipitaciones en la Región Pampeana. 39º Congreso Argentino de Agroinformática. 2das Jornadas Argentinas de Informática. JAIIo, sep. 2010
 - Holleman, I. (2006). Bias adjustment of radar-based 3-hour precipitation accumulations. Koninklijk Nederlands Meteorologisch Instituut.
 - Jayakrishnan, R., Srinivasan, R., Arnold, J. G. 2004. Comparison of rain gauge and WSR-88D Stage III precipitation data over the Texas-Gulf basin. *J. Hydrol.* 292, (pp135-152)
 - Joe, P.J. Smith, P. L., 2001. Summary of the radar calibration workshop. Preprints, 30th International Conference on Radar Meteorology, Munich, Germany, American Meteorological Society, (pp. 174-176).
 - Joss, J., Lee, R., 1995. The application of radar gauge comparisons to operational precipitation profile corrections. *J. Appl. Meteor.* 34, (pp. 2612–2630).
 - Joss, J., Waldvogel, A. 1990. Precipitation measurements and hydrology. In: D. Atlas (ed.), *Radar in Meteorology*, Am. Meteorol. Soc., Boston, MA, 577–606.
 - Joss, J., Zawadzki, I. 1997. Raindrop size distribution again? Preprints, 28th Conf. on Radar Meteorology, Austin, TX, Amer. Meteor. Soc., 326–327.
 - Kitchen, M., Blackall, R.B. 1992. Representativeness errors in comparisons between radar and gauge measurements of rainfall. *J. Hydrolog.* 134, (pp 13-33).
 - Kitchen, M., Jackson, P. M., 1993. Weather Radar Performance at Long Range - Simulated and Observed. *J. Appl. Meteor.* 32, (pp. 975–985)
 - Klazura, G. E.; Kelly, D.S. 1995: A comparison of high resolution rainfall accumulation estimates from the WSR-88D precipitation algorithm with rain gauge data. Preprints, 27th Conf. on Radar Meteorology, Vail, CO, Amer. Meteor. Soc., 31–34.
 - Krajewski, W. F., Smith, J. A. 2002. Radar hydrology: rainfall estimation. *Advances in water resources*, 25(8), (pp. 1387-1394).
 - Lee, G. 2006. Sources of errors in rainfall measurements by polarimetric radar: variability of drop size distributions, observational noise, and variation of relationships between R and polarimetric parameters. *J. Atmos. Oceanic Technol.*, 23, (pp.1005-1028).
 - Magliano, PN; Mercau, JL; Jobbágy, EG. 2012 ¿Es el radar TRMM (NASA, JAXA) un buen estimador de las precipitaciones de la región central Argentina? XI Congreso Argentino de Meteorología. Argentina.
 - Marshall, J.; Palmer, W. 1948. The distribution of raindrops with size. *J. Meteorol.* 4, (pp. 186–192).
 - Meischner, P.; Collier, C.; Illingworth, A.; Joss, J.; Randeu, W. 1997. Advanced weather radar systems in Europe: The COST 75 action. *Bulletin of the American Meteorological Society*, 78(7), (pp. 1411-1430).
 - Moreau, E.; Testud, J.; Le Bouar, E. 2009. Rainfall spatial variability observed by X-band weather radar ant its implication for the accuracy of rainfall estimates. *Advances in Water Resources*, 32, (pp. 1011-1019).
 - Nzeukou, A.; Sauvageot, H.; Ochou, A. D.; Kebe, C. 2004. Raindrop size distribution and radar parameters at Cape Verde. *Journal of Applied Meteorology*, 43(1), (pp. 90-105).
 - Pastor Garcia, V. 2013. Comparació de la pluja estimada per radar i la mesurada per pluviògrafs. Tesis de master. Universidad de Barcelona. Barcelona, España.
 - Llort, X. 2010. A study of the structure of radar rainfall and its errors. 2010. Tesi Doctoral Programa d'Enginyeria Civil Universitat Politècnica de Catalunya. center of Applied Research in Hydrometeorology (CRAHI). Catalunya, España.
 - Rinehart, R. 2007. Radar for meteorologists. Dept. of Atmospheric Sciences, Center for Aerospace Sciences. University of North Dakota. 428 pp.

- Roberto, Z.E., Casagrande, G., Viglizzo, E.F. 1994. Lluvias en la pampa central. Tendencias y variaciones del siglo. Publicación N°2. INTA, Centro Regional La Pampa-San Luis, Argentina.
 - Rosenfeld, D.; Amitai, E.; Wolff, D.B. 1995. Classification of rain regimes by the threedimensional properties of reflectivity fields. *Journal of Applied Meteorology*, 34, (pp. 198-211).
 - Rosengaus, M. 1995. Revisión Temática. Fundamentos de radares meteorológicos: Aspectos clásicos. *Ingeniería Hidráulica en Mexico*. 10, (pp. 55-74).
 - Sarochar, R.H.; Ciappesoni, H.H.; Ruiz, N.E. 2005. Precipitaciones convectivas y estratiformes en la Pampa Húmeda: una aproximación a su separación y aspectos climatológicos de ambas. *Meteorologica*, 30(1-2), (pp. 77-88).
 - Sauvageot, H. 1991. Radar meteorology. Artech House, Inc, 1991. 315 pp.
 - Sauvageot, H. 1994. Rainfall measurement by radar: A review. *Atmospheric Research*, 35(1), (pp. 27-54).
 - Sempere-Torres, D.; Sánchez-Diezma, R.; Zawadzki, I.; Creutin, J. D. 1999. DSD identification following a pre-classification of rainfall type from radar analysis. Preprints, 29th International Conference on Radar Meteorology, Montreal, Quebec, Canada, American Meteorological Society, (pp. 632-635).
 - Smith, J. A.; Krajewski, W.F. 1991. Estimation of the mean field bias of radar rainfall estimates. *J Appl. Meteorol.*, 30, (pp. 397-412).
 - Smith, J.A.; Seo, D.J.; Baeck, M.L.; Hudlow, M.D. 1996. An intercomparison study of NEXRAD precipitation estimates. *Water resour.* 32, (pp. 2035-2045).
 - Vidal, L.; Salio, P.; Pappalardo, L. 2012. Uso combinado de datos TRMM/PR y disdrómetro para corrección de reflectividad de radares meteorológicos en Argentina", 1^º Encuentro de investigadores en Formación en Recursos Hídricos (IFRH 2012), 15-16 Junio 2012, INA, Ezeiza.
 - Vignal, B., Andrieu, H., Creutin, J. D. 1999. Identification of vertical profiles of reflectivity from volume scan radar data. *Journal of applied meteorology*, 38(8), (pp. 1214-1228).
 - Wexler, R., Atlas, D. 1963. Radar reflectivity and attenuation of rain. *J. Appl. Meteorol.*, 2, (pp. 276-280).
 - Wilson, J. 1976. Radar-rain gage precipitation measurements: A summary. Preprints, Conference on Hydrometeorology (Ft. Worth), AMS, Boston, (pp. 72-75)
 - Wilson, J., Brandes, E. 1979: Radar measurement of rainfall. A summary. *Bull. Amer. Meteor. Soc.*, 60, (pp. 1048-1058)
 - Zawadzki, I., 1984. Factors affecting the precision of radar measurement of rain. Preprints, 22th International Conference on Radar Meteorology, Zurich, Switzerland, American Meteorological Society, (pp. 251-256).
 - Zrnic, DS. Ryzhkov, AV. 1999. Polarimetry for weather surveillance radar. 1999. *Bull Am Meteor Soc* 1999;80, (pp. 389-406).
-

Ajuste meteorológico de precipitación estimada con radar

2.

Belmonte, María Laura¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
belmonte.maría@inta.gob.ar

RESUMEN

Ajuste meteorológico se describe como un procedimiento aplicado para adaptar las medidas de lluvia estimadas por radar meteorológico (RM) a partir de reflectividad hacia una cantidad de lluvia medida, por ejemplo la precipitación registrada en superficie a través de pluviómetros. Los métodos de ajuste de información pluviométrica estimada por radar con datos de referencia terrestre presentan la ventaja de corregir, en un proceso simple, todas las fuentes de error que interfieren en el proceso de estimación. El objetivo de este trabajo fue aplicar una corrección a los valores estimados por el RM ubicado en la Estación Experimental de INTA Anguil, provincia de La Pampa. A partir de la información pluviométrica registrada durante diciembre de 2011 a mayo de 2012 por el RM y por una red de pluviómetros localizada bajo la cobertura del radar se calcularon y aplicaron factores de corrección mediante dos procedimientos de ajuste simple en dos períodos de registro diferentes. Las rectas de ajuste obtenidas luego de la aplicación de las correcciones mejoraron la exactitud del modelo en ambos períodos. Con el procedimiento de corrección I el RMSE disminuyó un 19,5% y 25% para los períodos 1 y 2, respectivamente. Aplicando la corrección II, este índice disminuyó un 17% para el período 1 y 21% para el período 2. En términos de eficiencia el modelo I superó al modelo de corrección II. Las correcciones también lograron reducir el error sistemático medio, encontrando el mejor resultado con la corrección II en el período 1.

Palabras clave: lluvia, radar meteorológico, corrección, pluviómetros

ABSTRACT

Meteorological adjustment is a procedure to adapt rainfall measurements estimated by meteorological radar (RM) from reflectivity to a measured amount of rainfall, eg precipitation recorded on surface through rain gauges. The methods of adjusting rainfall information estimated by radar with terrestrial reference data have the advantage of correcting, in a simple process, all the sources of error that interfere in the estimation process. The objective of this work was to apply a correction to the values estimated by the RM located in the INTA Anguil Experimental Station, province of La Pampa. From the rainfall data recorded during December 2011 to May 2012 by the RM and a network of rain gauges located under radar coverage, correction factors were calculated and applied by two simple adjustment procedures in two different recording periods. The adjustment straight obtained after the corrections improved the accuracy of the model in both periods. With correction procedure I the RMSE decreased by 19.5% and 25% for periods 1 and 2, respectively. Applying correction II, this index decreased by 17% for period 1 and 21% for period 2. In terms of efficiency the model I exceeded the model of correction II. The corrections also depleted the average systematic error, finding the best result with correction II in period 1.

Keywords: rain, weather radar, correction, rain gauges

INTRODUCCIÓN

Existen varios enfoques metodológicos para la corrección de la lluvia derivada de la estimación de

radares meteorológicos (RM). En este sentido en la comunidad de investigadores actualmente existe una clara separación entre dos términos que por muchos años se utilizaron indistintamente: “calibración” y “ajuste” (Gjertsen, et al., 2003), así la calibración se entiende como el conjunto de procedimientos aplicados al hardware de un RM para garantizar un rendimiento estable del mismo dentro de las especificaciones dadas y el ajuste meteorológico se describe como un procedimiento aplicado para adaptar las medidas de lluvia estimadas por RM hacia una cantidad deseada, por ejemplo la precipitación (PP) registrada en superficie a través de medidores como pluviómetros (Koistinen y Puhakka, 1986) o como cualquier procedimiento mediante el cual las características de los datos de RM se modifican para que se correspondan lo mejor posible con la cantidad de lluvia medida por pluviómetros de referencia (Michelson, 2003). Algunos métodos de ajuste se focalizan en mantener una relación fija entre reflectividad radar (z) y tasa de precipitación estimada (R) y utilizan las observaciones de pluviómetros o colectores de lluvia ubicados en superficie para ajustar las estimaciones derivadas de radar (Wilson y Brandes, 1979). Gjertsen, (2003) clasifica a las técnicas de ajuste de información pluviométrica de radar mediante el uso de pluviómetros en: simples y sofisticadas. Los métodos simples ajustan los datos de radar a partir de comparaciones puntuales con datos de pluviómetros, pudiendo ser el ajuste uniforme, para aquellos casos donde se determina un factor de ajuste medio que se aplica a todos los datos del radar (Harrold, et al., 1974; Barnston y Thomas, 1983); o variable, donde se admite que los factores de ajuste varían en el espacio y/o en el tiempo. Las técnicas sofisticadas utilizan interpolación óptima o geostadística, siendo algunos ejemplos las técnicas de kriging y cokriging (Krajewski, 2002; Seo et al., 1998), estos métodos aprovechan la información más valiosa de cada fuente: de los pluviómetros el valor cuantitativo de lluvia y del radar la estructura espacial de la precipitación.

Los métodos de ajuste de información pluviométrica estimada por radar con datos de referencia terrestre presentan la ventaja de tratar, en un proceso simple, con todas las fuentes de error que interfieren en el proceso de estimación. Por ejemplo algunas técnicas ajustan a una línea recta los pares de datos radar-pluviómetros co-localizados. (Wilson y Brandes, 1979; Collier, 1986), mediante compara-

ciones radar-pluviómetros se obtienen factores de corrección aplicables a las medidas radar con el objeto de minimizar su sesgo respecto a los datos de los pluviómetros (Brandes, 1975; Wilson y Brandes, 1979; Gjertsen et al., 2003).

Este trabajo tuvo como objetivo aplicar una corrección meteorológica a los valores estimados por RM de Anguil. Para ello se evaluaron dos procedimientos de ajuste simples.

MATERIALES Y MÉTODOS

Datos de RM

Se trabajó con archivos de información proveniente del RM de INTA (Instituto Nacional de Tecnología Agropecuaria) ubicado en la Estación Experimental de Anguil, (36,539 S; 63,990 W). La calibración electrónica del hardware se realizó durante la instalación y periódicamente se verifica la potencia de la señal emitida.

A partir de los archivos de Z en mm^6/m^3 se estimó la R en mm/h . Para ello se utilizó una relación matemática entre estas variables y se mantuvo fija, correspondió a la propuesta por Marshall y Palmer (1948), cuyos parámetros se configuraron a partir del software Rainbow® 5 según protocolo del fabricante (Gematronik, 2005). Posteriormente se generaron los productos hidrológicos SRI (*Surface Rain Intensity*), algoritmo que genera una imagen con valores asociados de intensidad de precipitación en mm/h estimada para cada celda o unidad mínima de la misma. A partir de la integración de los SRI se obtuvo el producto PAC (*Precipitation Accumulation*). Este fue el dato final analizado en este trabajo que representa un acumulado de lluvia en mm en imágenes de 8 bits que constan de 240 x 240 pixeles con una resolución espacial de 1 km² y 1°, por píxel o celda y resolución temporal de 10 minutos.

Datos de referencia

Como datos de referencia se utilizaron en este trabajo información de PP proveniente de la red pluviométrica de la provincia de La Pampa. Para la realización de este trabajo fueron seleccionados 40 pluviómetros. Los datos de PP correspondieron a registros diarios medidos en mm observados a las 9.00

a.m, correspondientes a un día pluviométrico (9 a.m a 9 a.m) en pluviómetro homologado de acuerdo a normas internacionales de la Organización Meteorológica Mundial (OMM) y según protocolo de observación del Servicio Meteorológico Nacional (SMN).

Metodología

Los archivos de RM 10-minutales se integraron en compuestos de PP acumulada diaria (24-horarios).

A partir de la información pluviométrica registrada durante el periodo diciembre de 2011 a mayo de 2012 por el RM y los pluviómetros se calcularon factores de corrección mediante dos procedimientos de ajuste simples (Gjertsen, 2003) que se detallan a continuación:

Procedimiento I: determinación de un factor de ajuste Mean Field Bias (F) (Wilson y Brandes., 1979), según la ecuación 1. Este factor fue aplicado uniformemente a todas las estimaciones de cada píxel de la imagen del radar para su corrección (ec. 2):

$$F = \frac{\sum_{i=1}^N P}{\sum_{i=1}^N R} \quad (1)$$

donde ΣP es la sumatoria de datos de lluvia acumulada diaria registrada por los pluviómetros y ΣR es la sumatoria de los compuestos diarios de radar en los pixeles co-localizados, registrados durante todo el periodo considerado.

$$R_{corregido} = F \times R \quad (2)$$

Procedimiento II: determinación de un factor de ajuste (F) según la ecuación 3. Este factor presenta una transformación logarítmica para que la relación matemática entre las medidas de pluviómetros y las estimaciones correspondientes de radar, tienda a estar normalmente distribuida (Cain y Smith, 1976; Koistinen y Michelson, 2002). Es aplicado uniformemente a todas las estimaciones de cada píxel de la imagen del radar para su corrección (ec. 4).

$$F = \text{Exp}[\log(P/R)] \quad (3)$$

$$R_{corregido} = F \times R \quad (4)$$

Con los factores obtenidos mediante los procedimientos de corrección I y II (CI y CII, respectivamente) se afectaron los productos PAC de lluvia acumulada diaria en dos periodos posteriores de registro de datos:

- **Periodo 1:** registros obtenidos durante agosto, septiembre y octubre de 2012.
- **Periodo 2:** registros obtenidos durante enero, febrero y marzo de 2013.

Posteriormente los valores corregidos de radar se volvieron a comparar con los acumulados de los pluviómetros correspondientes a cada píxel de la imagen.

La performance de los procedimientos de ajuste aplicados y su comparación con las estimaciones previo al ajuste se evaluó a partir de indicadores RMSE y el índice BIAS.

RESULTADOS Y DISCUSIÓN

A partir de los gráficos de dispersión de las figuras 1 y 2, se observó cómo las rectas de ajuste obtenidas luego de la aplicación de las correcciones se aproximaron más hacia la recta 1:1 en ambos períodos, en comparación con los datos previo al ajuste, mejorando la exactitud del modelo.

Como resultado de la aplicación de las dos correcciones se evidenció una mejora de todos los índices calculados en ambos períodos. Con la corrección I el RMSE disminuyó un 19,5% y 25% para los períodos 1 y 2, respectivamente. Aplicando la corrección II, este índice disminuyó un 17% para el periodo 1 y 21% para el periodo 2 (Tabla 1). Woodley *et al.* (1974); Brandes, (1975) y Huff y Towery, (1978), aplicaron factores de ajuste obtenidos mediante la comparación de estimaciones radar y medidas pluviométricas. Estos trabajos lograron reducir el error promedio entre un 30 y 75% con estimaciones de RM de simple polarización. Wilson y Brandes, (1979) lograron reducir la diferencia entre ambos medidores a valores de entre 10 a 30% mediante la comparación con información proveniente de redes de pluviómetros, sus resultados mejoraron entre 10 y 20%, al aumentar la densidad de pluviómetros utilizados.

En términos de eficiencia el modelo I superó al modelo de corrección II.

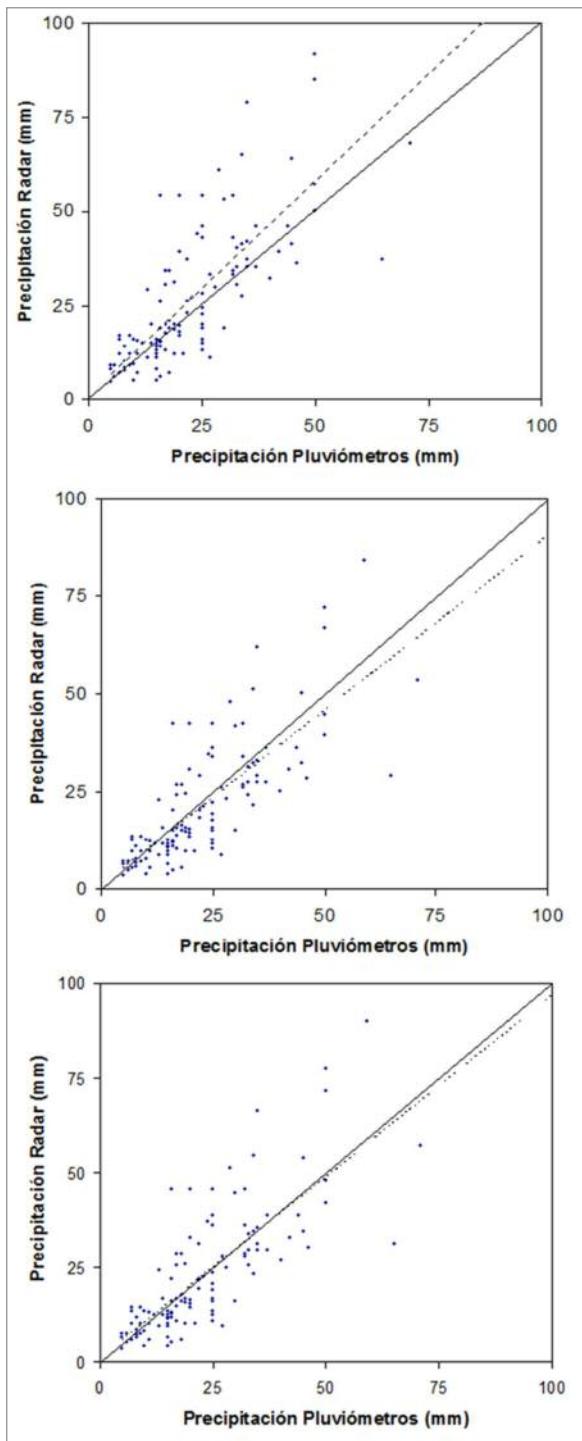


Figura 1. Gráficos de dispersión y ajuste lineal (línea partida) entre la precipitación diaria estimada por radar y la observada en pluviómetros de referencia para el periodo 1. Datos radar sin ajustar (arr). Datos radar aplicando la corrección I (abajo izq.), datos radar aplicando la corrección II (abajo der.). La línea llena representa la relación 1:1

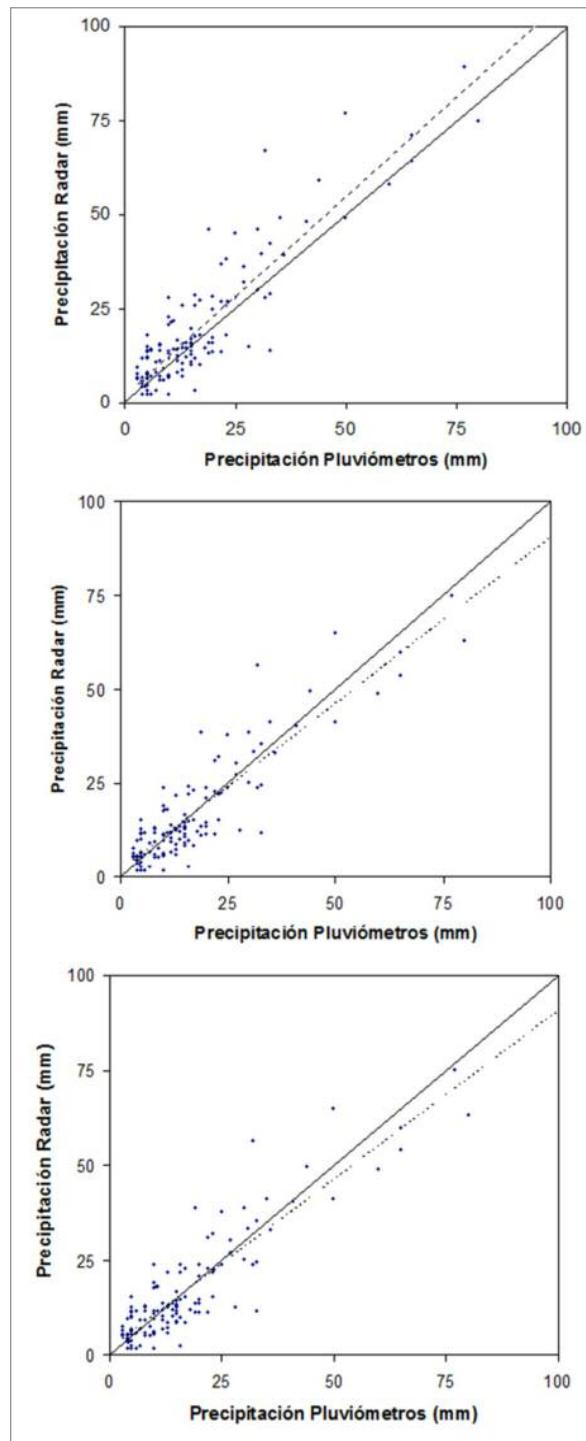


Figura 2. Gráficos de dispersión y ajuste lineal (línea partida) entre la precipitación diaria estimada por radar y la observada en pluviómetros de referencia para el periodo 2. Datos radar sin ajustar (arr). Datos radar aplicando la corrección I (abajo izq.), datos radar aplicando la corrección II (abajo der.). La línea llena representa la relación 1:1

Tabla 1. Índices comparativos de los resultados obtenidos al aplicar dos correcciones a dos períodos diferentes.

PERÍODO		RMSE (mm)
1	Previo a corrección	13,2
	Corrección I	9,9
	Corrección II	10,4
2	Previo a corrección	8,2
	Corrección I	6,6
	Corrección II	6,8

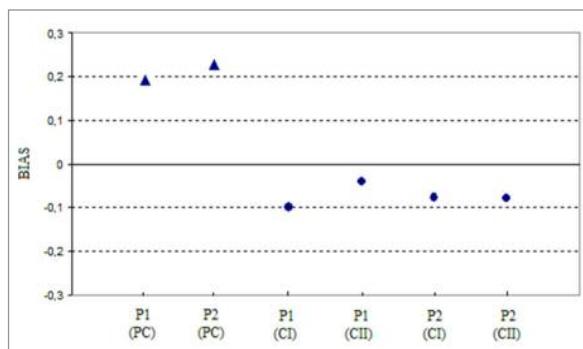


Figura 3. Variación del BIAS previo a la aplicación de procedimientos de corrección (PC) y luego de las correcciones I (CI) y II (CII). P1 y P2 son periodo de registro 1 y 2 respectivamente

Las correcciones también lograron reducir el error sistemático medio, encontrando el mejor resultado con la corrección II en el período 01. (Fig. 3)

CONCLUSIONES

Los procedimientos de ajuste simple evaluados resultaron en la reducción del sesgo de las estimaciones del radar meteorológico ubicado en Anguil en relación a datos de referencia. Estas correcciones producen una mejora en la exactitud de las estimaciones.

REFERENCIAS BIBLIOGRÁFICAS

- Belmonte, M.L. 2014. Estimación de precipitación a partir de radar meteorológico: evaluación y desarrollo de productos para la toma de decisiones en planificación agropecuaria. Tesis presentada para optar al grado de Magister en Ciencias Agrarias. Universidad Nacional del Sur.
- Barnston, A. G., Thomas, J. L. 1983. Rainfall measurement accuracy in FACE: A comparison of gage and radar rainfalls. Journal of climate and applied meteorology, 22(12), (pp. 2038-2052).
- Brandes, E. A. 1976. Convective rainfall estimation by radar: Experimental results and proposed operational analysis technique. Preprints, Conference on Hydrometeorology (Ft. Worth), AMS, Boston, (pp. 54-59)
- Cain, D. E., Smith, P. L. 1976. Operational adjustment of radar-estimated rainfall with rain-gauge data-statistical evaluation. In bulletin of the american meteorological society. 57, no. 7,(pp. 876-876). 45 beacon st, boston, ma 02108-3693: amer meteorological soc.
- Collier, C. 1986. Accuracy of rainfall estimates by radar, Part I: Calibration by telemetering raingauges. Journal of Hydrology, 83(3), (pp. 207-223).
- Gematronik, 2005. Rainbow® 5 Products & Algorithms. Gematronik GmbH., Raiffeneisenstr. 10, 41470 Neuss, Germany
- Gjertsen, U., Salek, M., Michelson, D. B. 2003. Gauge-adjustment of radar-based precipitation estimates—a review. COST-717 working document No. WDD, 2(200310), 1.
- Harrold, T. W., E. J. English., Nicholass, C.A. 1974: The accuracy of radar-derived rainfall measurements in hilly terrain. Quart. J. Roy. Meteor. Soc., 100, (pp. 331–350).
- Huff, F. A., Towery, N. G. 1978. Utilization of radar in operation of urban hydrologic systems. Preprints. In 18th Conference on Radar Meteorology, Amer. Meteor. Soc, Boston (pp. 437-441).
- Koistinen, J., Michelson, D. B. 2002. BALTEX weather radar-based precipitation products and their accuracies. Boreal environment research, 7(3), (pp. 253-263).
- Koistinen, J., Puhakka, T., 1986. Can we calibrate radar with raingauges? Geophysica (Helsinki) 22, (pp.119–129).
- Krajewski, W. F., Smith, J. A. 2002. Radar hydrology: rainfall estimation. Advances in water resources, 25(8), (pp. 1387-1394).
- Marshall, J.; Palmer, W. 1948. The distribution of raindrops with size. J. Meteorol. 4, (pp. 186–192).
- Michelson, D.B. 2003. Quality control of weather radar data for quantitative application, PhD thesis, Telford Institute of Environmental Systems, University of Salford, Salford, UK, 281 pp.
- Seo, D. J.; Fulton, R. A.; Breidenbach, J. P. 1998. Rainfall estimation in the WSR-88D era for operational hydrologic forecasting in the National Weather Service. Preprints. In Special Symposium on Hydrology. Am. Meteor. Soc., Boston, MA.
- Wilson, J., Brandes, E. 1979: Radar measurement of rainfall. A summary. Bull. Amer. Meteor. Soc., 60, (pp. 1048–1058)
- Woodley, W. L., Olsen, A., Herndon, A., Wiggert, V. 1974. Optimizing the measurement of convective rainfall in Florida.

Imágenes de radar meteorológico como recurso en la planificación de actividades agropecuarias

3.

Belmonte, María Laura¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
belmonte.maria@inta.gob.ar

RESUMEN

El objetivo de este trabajo fue desarrollar productos basados en la estimación de las precipitaciones a partir de radar meteorológico (RM) e información complementaria. Se utilizaron imágenes de precipitación acumulada de RM ubicado en la Estación Experimental de INTA Anguil. Como información complementaria se utilizaron los promedios históricos mensuales de precipitación (PP) de localidades de La Pampa y provincias limítrofes; así como valores de contenido de humedad en suelos a capacidad de campo de la región NE de La Pampa y mapas de evapotranspiración real derivada de imágenes satelitales. El periodo de trabajo fue el año 2012 completo; y el primer trimestre de 2013. Se generaron anomalías mensuales de PP, acumulados de PP durante barbecho de trigo, y estimación de reservas de agua en el suelo durante el ciclo productivo de cultivos estivales e invernales de la región mediante un balance hídrico seriado. Las estimaciones de almacenaje de agua edáfica fueron comparadas con mediciones a campo. Los productos desarrollados se presentaron como mapas en formato ráster en diferente resolución temporal y resolución espacial de 1 km². El análisis de correlación entre los valores estimados de almacenaje y los datos de campo durante el ciclo de cultivos de invierno arrojó un índice de correlación de 0,84 en tanto que para cultivos de verano fue de 0,69. El error cuadrático medio calculado para septiembre y octubre fue de 20 mm y para enero, febrero y marzo fue de 37 mm. Los productos derivados de imágenes de PP de RM se presentan como una metodología apta para el estudio espacial y temporal de esta variable, a su vez posibilita la generación de indicadores útiles para valorar el desarrollo y los potenciales rindes de los cultivos regionales. Esta información es presen-

tada en formatos gráficos digitales utilizables e integrables a sistemas de información geográfica (SIG).

Palabras clave: *precipitaciones; sensores remotos; cultivos regionales; SIG*

ABSTRACT

The aim of this work was to develop products based on the estimation of precipitation from meteorological radar (RM) and complementary information. Images of accumulated precipitation of RM located in the Experimental Station of INTA Anguil were used. Historical average monthly precipitation (PP) from locations of La Pampa province and bordering provinces were used as complementary information. Also were included values of moisture content in soils at field capacity of the NE region of La Pampa and maps of real evapotranspiration derived from satellite images. The work period was the 2012 and the first quarter of 2013. Monthly anomalies of PP, accumulated PP during wheat fallow, and estimation of soil water storage during the productive cycle of summer and winter crops of the region were generated by a serial water balance. Soil water storage were compared to field measurements. The developed products were presented as raster maps with different temporal resolution and with an spatial resolution of 1 km². The correlation analysis between estimated storage values and field data during the winter crop cycle yielded a correlation index of 0.84 while for summer crops it was 0.69. The calculated mean squared error for September and October was 20 mm and for January, February and March was 37 mm. The products derived from MR PP images are presented as a suitable methodology for the spatial and temporal study of this variable,

which in turn makes it possible to generate useful indicators to assess the development and potential yields of the regional crops. This information is presented in usable digital graphic formats and integrable with geographic information systems (GIS).

Keywords: precipitation; remote sensing; Regional crops, GIS

INTRODUCCIÓN

Las precipitaciones influyen en los ciclos productivos agrícolas, desde la siembra a la cosecha de un cultivo, como así también durante labores culturales, almacenamiento de granos y hasta el transporte de la producción. A su vez determinan la expansión o retracción de la superficie sembrada y la adopción de técnicas agrícolas. Por esta razón la disponibilidad de los registros pluviométricos en base diaria, decádica y mensual resultan de utilidad en la planificación agropecuaria y en sus actividades operativas relacionadas. Sin embargo, las series de datos diarios de precipitaciones muchas veces son incompletas o no son de rápido acceso. El radar meteorológico perteneciente al INTA localizado en Anguil, La Pampa constituye una posibilidad de estimar lluvias en una resolución temporal y espacial nunca antes obtenida para la región. Asimismo genera información en lugares donde en la actualidad no existen otros instrumentos de medición, o donde la densidad de éstos es insuficiente, mejorando la representación de la distribución espacial de las precipitaciones en la región. El procesamiento de datos generados por el sensor junto con la integración de otras fuentes de información implica una mejora en disponibilidad de herramientas para la toma de decisiones en la actividad agropecuaria. Existen antecedentes en Argentina que emplean datos de lluvia proveniente de sensores remotos con la finalidad proporcionar información objetiva a escala regional (Fernandez Long *et al.*, 2011; Della Chiesa, 2011).

El objetivo de este trabajo fue desarrollar una metodología basada en la estimación de las precipitaciones a partir del radar e información complementaria, que establece indicadores a escala regional. Tiene por finalidad la propuesta de productos que resulten de utilidad en la toma de decisiones en la planificación de actividades agropecuarias.

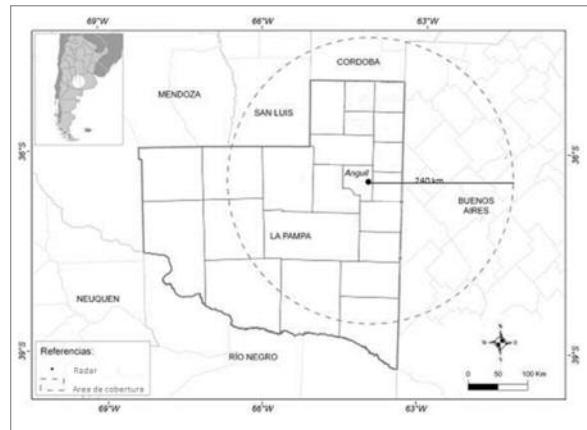


Figura 1. Ubicación geográfica del RM y su área de cobertura

Área de estudio

El RM se encuentra localizado en cercanías de la localidad de Anguil, provincia de La Pampa, dentro de la región pampeana argentina. Su radio de acción es de 240 km, abarca gran parte de esta provincia, parte oeste de Buenos Aires y una pequeña proporción del sur de Córdoba y San Luis, alcanzando a cubrir un área aproximada a 18 millones de hectáreas (Fig. 1).

Los principales cultivos de cosecha de la región son trigo y cebada en época invernal, y girasol, soja, maíz y sorgo en la época estival (Lorda, *et al.*, 2008). Prevalecen los sistemas de producción mixtos, basados en rotaciones de cultivos anuales y pasturas perennes en base a leguminosas, siendo la más común alfalfa (Viglizzo *et al.*, 2001). Casi la totalidad de la producción se realiza en condiciones de secano, dependiendo del agua proveniente de las precipitaciones y de la capacidad de almacenaje del suelo.

El régimen térmico no resulta limitante para las actividades agropecuarias sin embargo como la producción extensiva en esta zona se realiza bajo condiciones de secano, la lluvia resulta el elemento climático de mayor influencia en la región, tanto por su cantidad como por su distribución anual. Según Viglizzo *et al.* (1991) La Pampa es una región de transición donde coexisten los ambientes subhúmedo, semiárido y árido. La zona bajo estudio se encuentra entre las isohietas de 700 mm y 850 mm de precipitación anual. Estudios realizados por Galmarini (1961), Hoffman (1988) y Roberto *et al.* (1994) coinciden en que la cantidad de lluvia varía fuertemente

en el sentido este-oeste en la provincia de La Pampa y que existe una marcada concentración en la distribución estacional de lluvias en primavera verano con relación a otoño invierno.

MATERIALES Y MÉTODOS

Información de radar

Se trabajó con archivos de reflectividad (Z) proveniente del radar meteorológico de INTA (Instituto Nacional de Tecnología Agropecuaria) ubicado en la Estación Experimental de Anguil, (36,539 S; 63,990 W). Se trata de un radar doppler de doble polarización que opera en banda C a una frecuencia de 5,64 Ghz y longitud de onda de 5,4 cm, de industria alemana, marca Gematronik modelo Meteor 600C.

Este radar estuvo configurado para completar una serie de giros a 3600 que se repite para 12 ángulos de elevación, entre 0,50 de base y 15,10 de tope, en un rango de 240 km (Hartmann, et al., 2010), representando una exploración o escaneo del volumen completo. La frecuencia de escaneo se programó cada 10 minutos, y bajo funcionamiento normal totalizó 144 adquisiciones diarias. Como resultado de cada adquisición se generó una cantidad de datos numéricos que caracterizan un volumen de la atmósfera situado dentro del haz de radiación. Estos datos corresponden a variables que fueron almacenados en archivos denominados volúmenes, en servidores *in situ*. En este trabajo se utilizaron los volúmenes que contenían la variable Z , a los cuales se accedió por vía remota.

Se descargaron archivos con resolución temporal de 10 minutos, durante el año 2012 y primer trimestre de 2013. Los datos en bruto proporcionados por el radar presentaban un formato de archivos definido por el fabricante, caracterizado por un encabezamiento en XML (Extensible Markup Language), y un área de datos en formato binario con compresión. Esta situación impone una restricción al uso de herramientas estándares para manipular XML. Fue necesario realizar una transformación a archivos de tipo ASCII (American Standard Code for Information Interchange) y posteriormente se transformaron a imágenes con formato TIFF (Tagged Image File Format). Para este procedimiento se utilizaron herramientas de software para poder

leer, interpretar y escribir la información en estos formatos adecuados, generadas por Hartmann et al., (2010).

A partir de los archivos de Z en mm^6/m^3 se estimó la intensidad de precipitación (R) en mm/h . Para ello se utilizó una relación matemática entre estas variables propuesta por Marshall y Palmer (1948), cuyos parámetros se configuraron a partir del software Rainbow® 5 según protocolo desarrollado por fabricante del radar (Gematronik, 2005). Dicha relación corresponde a:

$$Z = 200 R^{1,6}$$

Posteriormente se generó el producto hidrológico SRI (*Surface Rain Intensity*). Este algoritmo genera una imagen con valores asociados de intensidad de precipitación en mm/h estimada para cada celda o unidad mínima de la misma. A partir de la integración de los SRI se obtuvo otro producto: PAC (*Precipitation Accumulation*) que representa un acumulado de lluvia en milímetros, en imágenes de 8 bits que constan de 240 x 240 píxeles con una resolución espacial de 1 km^2 y 1°, por pixel o celda y resolución temporal de 10 minutos.

Información complementaria

Se utilizaron promedios históricos mensuales generados con no menos 30 años de las distintas localidades de la provincia de La Pampa y algunas localidades de las provincias limítrofes (Córdoba, Mendoza, San Luis y Buenos Aires).

Se utilizaron imágenes de Evapotranspiración real (ETR) proveniente del sensor AVHRR de los satélites de la serie NOAA con una resolución espacial de 100 ha (1 x 1 km). Los valores de ETR son derivados según el modelo de Di Bella et al. (2000), a partir de un índice de vegetación normalizado (NDVI) y la temperatura de la superficie (T_s), como:

$$\text{ETR (mm/mes)} = -88,3439 + 1,77636 T_s + 286,406 \text{ NDVI}$$

Se utilizó información de valores de contenido de humedad a capacidad de campo, la cual se obtuvo a partir de una base de datos histórica de la región NE de la provincia de La Pampa suministrada por el laboratorio de suelos de la EEA INTA Anguil.

Métodos

Se integraron los archivos 10-minutales en compuestos de 24 hs (diarios). A partir de estos compuestos, se realizaron integraciones en otras resoluciones temporales. Se generó una primera serie de compuestos de PP acumulada: decádicos y mensuales. A partir de éstos y la información complementaria se desarrolló una segunda serie de productos. Estos productos fueron: anomalías mensuales de PP, PP acumulada durante el periodo de barbecho de trigo y estimación de reservas de agua en el suelo mediante un balance hídrico seriado.

La reproyección y georreferenciación de imágenes se realizó con el software GVSIG 1.1.2 (<http://www.gvsig.org>) y para el procesamiento, cálculos, composición y análisis de imágenes y productos se utilizó el complemento calculadora de mapas raster del mencionado software. Para poder efectuar cálculos numéricos entre imágenes radar y satelitales se reproyectaron las imágenes de ETR a coordenadas geográficas Lat/Long, datum WGS-84.

Para calcular la PP decádicos y mensuales se integraron los compuestos diarios por década (10 días), y en un total de 30 o 31 días, según el mes.

Anomalías mensuales de precipitación

Con los datos pluviométricos de las localidades de la provincia de La Pampa y localidades de provincias limítrofes se generó una imagen ráster, mediante algoritmo de interpolación inverso a la distancia (IDW). La imagen resultante fue de precipitación promedio mensual en resolución de 1 x 1 km. A partir de la misma y las imágenes de compuestos de precipitación acumulada mensual de radar en la misma resolución espacial, se calcularon las anomalías de precipitación mensuales para los meses del año 2012, mediante el cálculo siguiente:

$$\text{AMP}_i(\%) = \frac{\overline{PP}_i - \overline{PP}_i}{\overline{PP}_i} \times 100$$

AMP_i es la anomalía mensual de precipitación del mes i , expresada en porcentaje.

\overline{PP}_i es la precipitación acumulada mensual del mes i

\overline{PP}_i es la precipitación mensual promedio del mes i

Precipitación acumulada a la siembra de trigo

En las regiones semiáridas y subhúmedas la acumulación de agua previa a la siembra del cultivo, asociado con una adecuada capacidad de retención por parte de los suelos, resulta clave para compensar los requerimientos de agua que normalmente no son cubiertos por las precipitaciones ocurridas durante el ciclo. Por lo tanto las precipitaciones ocurridas durante los períodos de barbecho muchas veces definen la decisión de sembrar y estrategias de fertilización y/o aplicación de herbicidas.

Se calculó la precipitación acumulada durante el periodo de barbecho de este cultivo en la región, considerado entre el inicio del mismo y el inicio de la siembra. Se consideraron dos fechas de siembra: 15 de junio para trigo de ciclo de crecimiento intermedio-largo y 1º de agosto para trigos de ciclo corto. Como el 74% del trigo para cosecha de grano cultivado en La Pampa tiene como antecesor a los cultivos de cosecha gruesa (Belmonte, et al., 2010) se consideró como inicio de barbecho el 15 de abril. Se integraron los compuestos diarios de radar durante el periodo de barbecho de este cultivo.

Estimación de reservas de agua del suelo durante el ciclo productivo de cultivos

Se realizó el seguimiento de la reserva de agua almacenada en suelo a través de un balance hídrico decádico. La información provista por sensores remotos capta la variabilidad espacial de dos de los componentes más importantes del balance hídrico edáfico (BH), la PP y la evapotranspiración (ET). Se integró la información de PP estimada por RM y ET estimada a partir de imágenes satelitales, a través del modelo de balance hídrico seriado (BHS) desarrollado por Thornthwaite y Mather (1955) y modificado por Pascale y Damario (1977). Este balance se establece entre la demanda atmosférica del agua a través del proceso de evapotranspiración y la oferta del agua dada por las precipitaciones y el agua almacenada en el suelo en períodos previos. El valor del almacenaje de un periodo t se obtuvo sumando al almacenaje del periodo $t-1$ el valor de la diferencia de PP y ET ($\text{PP}-\text{ET}$) si esta es positiva:

$$\text{Alm } t = \text{Alm } t-1 + (\text{PP}-\text{ET})$$

Si la diferencia PP-ET es negativa, el almacenaje del periodo t se calcula como:

$$\text{Alm } t = \text{Alm } t-1 \times e^{[(PP- ET)/CC]}$$

Si la diferencia PP-ET es positiva, pero si esta diferencia sumada al almacenaje calculado para el periodo sucesivo anterior supera la CC:

$$\text{Alm } t = CC$$

Alm es el agua almacenada en el suelo en milímetros (mm), PP y ET son la precipitación y evapotranspiración del periodo en mm y CC es la capacidad de campo del suelo en mm.

Los valores de CC fueron asignados a partir de una base de datos de la provincia de La Pampa suministrada por el laboratorio de suelos de la EEA Anguil de INTA. A partir de esta información se calculó un valor promedio para cada departamento provincial involucrado en el cálculo. Para el resto del área de las provincias vecinas se utilizaron estos mismos datos, por ejemplo a los partidos del oeste de Bs. As., se asignó los valores de los departamentos pampeanos más cercanos a cada uno. Posteriormente se generó una imagen ráster que contenía un valor de CC asignado a cada píxel de 1 km².

En un BHS a medida que va aumentando la unidad de tiempo o periodo de cálculo del almacenaje, también aumenta el error en la estimación del agua edáfica con una tendencia a la subestimación de la humedad real del suelo (Forte Lay y Villagra, 1983), esto último se debe fundamentalmente a la mecánica intrínseca del modelo de balance, que considera como humedad entrante en el suelo, a la cantidad de agua que proviene de la diferencia precipitación y evapotranspiración de la unidad de tiempo considerada, y al hecho de que las precipitaciones no se reparten en forma equitativa durante todos los días del periodo, sino que en general, hay unos pocos días con precipitación y el resto carece de ella. Es por esta razón que a partir de la estimación del dato diario de lluvia que provee el radar y la disponibilidad del dato de ETR decádico a escala regional aportado por imágenes satelitales, se realizó el cálculo de almacenaje de un BHS en un periodo decádico con el fin de mejorar el resultado a obtener. En total se obtuvo el cálculo de almacenaje para 24 décadas

consecutivas, desde principios de agosto 2012 a fines de marzo de 2013.

La metodología del BHS obtiene el almacenaje correspondiente de un periodo a partir del almacenaje del anterior. En este trabajo se consideró como periodo inicial la primera década de agosto 2012, debido a que se contaba con el valor de humedad medido a campo a partir de una base de datos georeferenciada del estado hídrico del suelo mensual que lleva a cabo la RIAN (Red de Información Agropecuaria Nacional del INTA) cuya metodología de obtención se describe en Belmonte, et al., (2011). Como estos datos no tenían resolución espacial de 1 km² (valor del píxel), se utilizó el promedio de los datos disponibles dividiendo la región de estudio en dos zonas: norte y sur (Fig. 2). Para la zona norte se partió de un almacenaje de 100 mm y para la zona sur de 60 mm. Con estos valores se generó una imagen ráster denominada Humedad Inicial, para ingresarla a los cálculos posteriores.

Las estimaciones de almacenaje se validaron contra mediciones a campo de contenido de agua útil en el suelo, obtenidas mediante el método gravimétrico. Fue realizado durante las últimas décadas de los meses de septiembre y octubre de 2012 en lotes de trigo y durante febrero y marzo de 2013

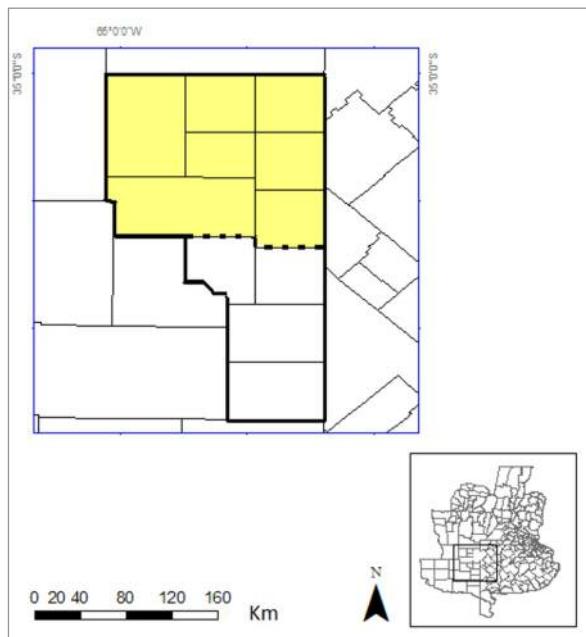


Figura 2. División de zonas utilizadas para el cálculo del promedio de humedad inicial. El área delimitada con línea oscura corresponde al área provincial de cálculo del BH. La parte coloreada corresponde a la zona norte, el área sin colorear a la zona sur

sobre lotes cultivados con soja, en la región NE de provincia de La Pampa. Para ello se utilizó información de la base de datos georreferenciada de la RIAN (descripta en párrafos precedentes). Los datos utilizados en este trabajo correspondieron a milímetros (mm) de agua útil almacenada en suelo hasta el metro de profundidad.

Los datos estimados mediante el BHS se correlacionaron con los datos medidos a campo en los pixeles centrados en posición de cada lote de la muestra. Se calculó el índice de Pearson, el nivel de significancia del análisis de correlación fue de 5%, y se calculó el índice RMSE.

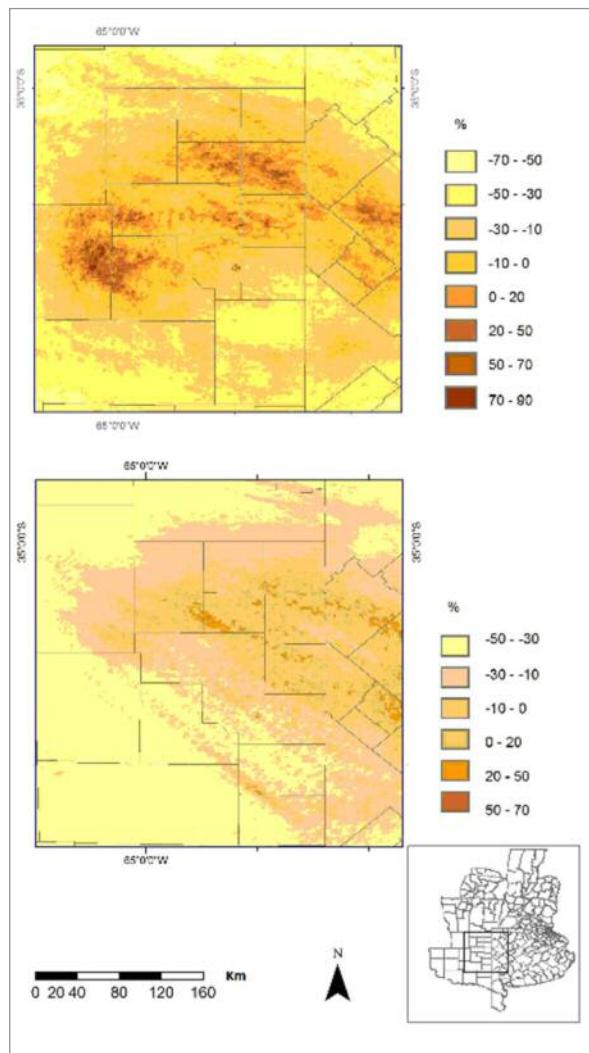


Figura 3. Mapas de anomalías de precipitación mensual para los meses de octubre (arriba) y septiembre (abajo) del año 2012

RESULTADOS

Anomalías mensuales de precipitación

Este producto se presenta como mapa de anomalías mensuales de precipitación, expresadas en porcentaje respecto del promedio de lluvias mensual, correspondiente al periodo 2012. Se exhiben los mapas correspondientes a los meses de septiembre y octubre de 2012 (Fig. 3)

Precipitación acumulada a la siembra de trigo

Este producto se presenta en mapas de precipitación formado por una composición de imágenes correspondientes al periodo desde el inicio de bar-

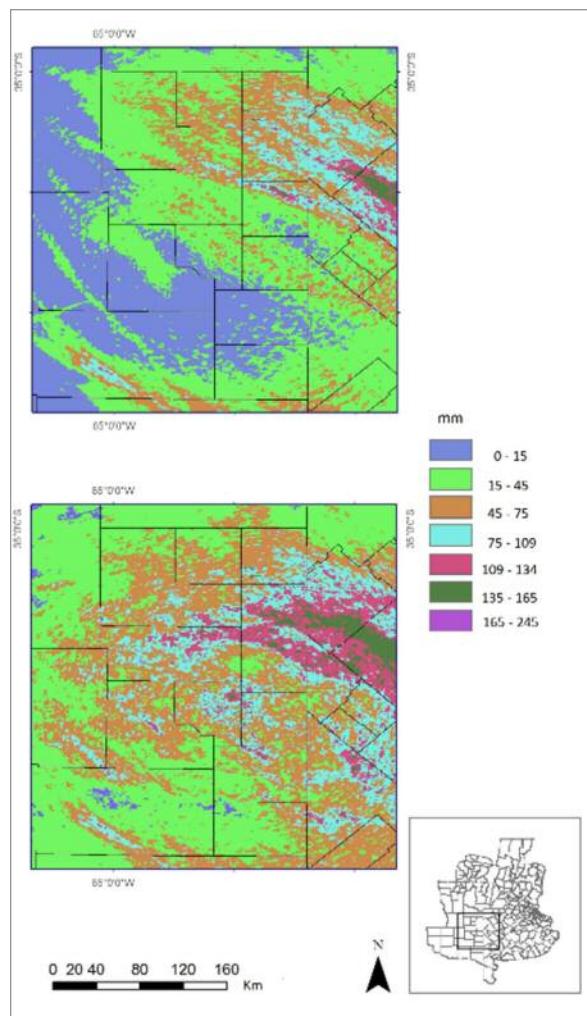


Figura 4. Mapas de precipitación acumulada a la primera fecha de siembra (arriba) y segunda fecha de siembra (abajo) del cultivo de trigo en la región

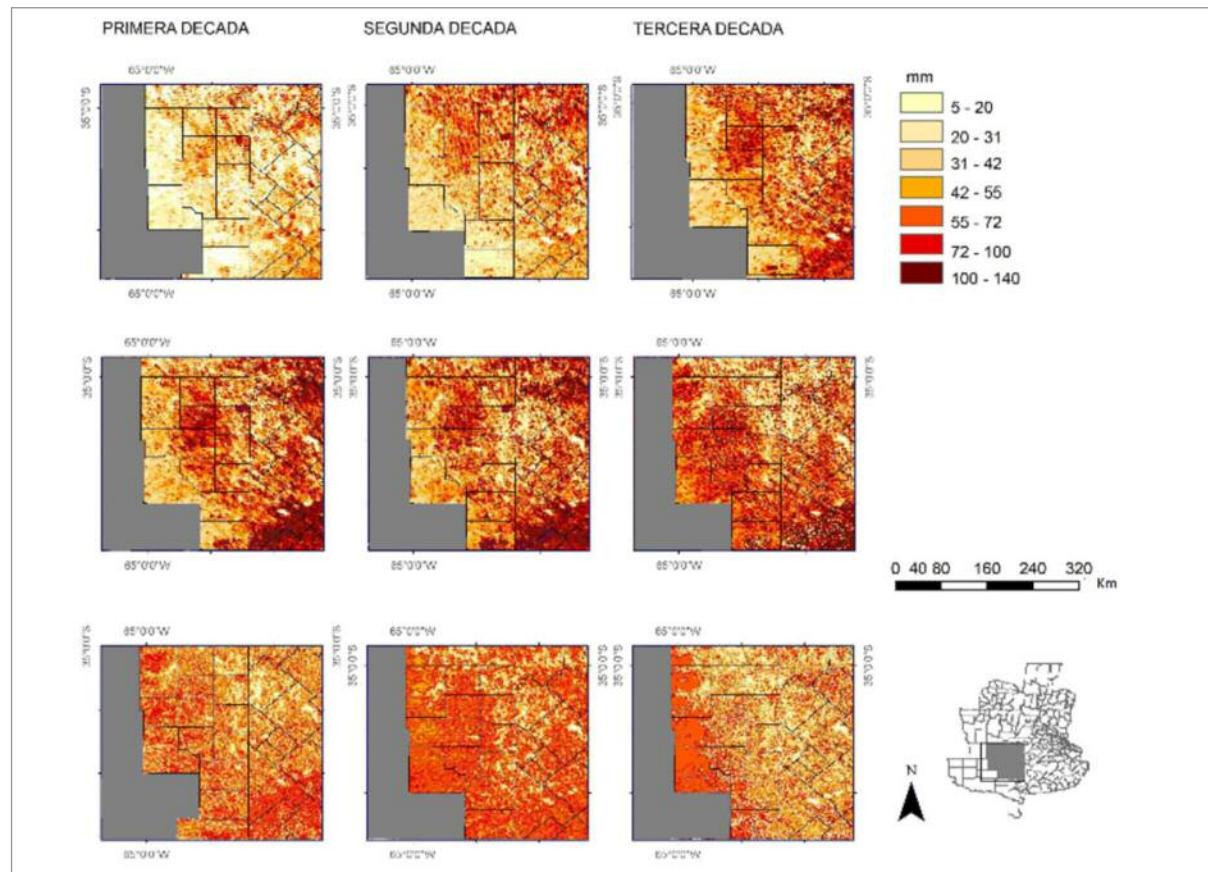


Figura 5. Mapas de almacenaje de agua en el suelo estimado durante las tres décadas de los meses de enero (arriba), febrero (centro) y marzo (abajo) del año 2013

becho a primera fecha de siembra y desde el inicio de barbecho hasta la segunda fecha de siembra de este cereal en la región (Fig. 4).

Estimación de reservas de agua del suelo durante el ciclo productivo de cultivos

Se generaron mapas de almacenaje de agua en el suelo por década desde agosto 2012 a marzo 2013. En este trabajo se presentan los mapas correspondientes a las décadas de enero, febrero y marzo de 2013 (Fig. 5)

Las imágenes logran representar un patrón de distribución espacial complejo y muy dinámico década a década.

Los valores de la muestra a campo utilizados en la validación mostraron buena amplitud de condiciones de humedad, entre 47 y 134 mm en lotes de trigo y entre 9 y 120 mm en lotes de soja. Estos valores fueron coherentes con los rangos de estimación del

modelo durante el periodo de toma del muestreo. El análisis de correlación entre los valores estimados de almacenaje de agua en suelo y los datos medidos a campo durante las décadas analizadas en el ciclo de cultivos de invierno arrojó un índice de correlación de 0,84 en tanto que para las décadas analizadas para cultivos de verano fue de 0,69 ($p < 0,05$). (Fig. 6).

El RMSE calculado para el periodo septiembre-octubre fue de 20 mm y para el periodo febrero-marzo fue de 37 mm. Los resultados marcan una tendencia a la sobreestimación de las reservas, la respuesta fue mejor para los meses de primavera respecto de los de verano.

Una probable fuente de error en la validación realizada podría deberse al efecto de la heterogeneidad a nivel de lote utilizado para el análisis de la humedad medida a campo, que puede surgir por diferencias en la topografía del terreno, situación frecuente en potreros productivos provinciales.

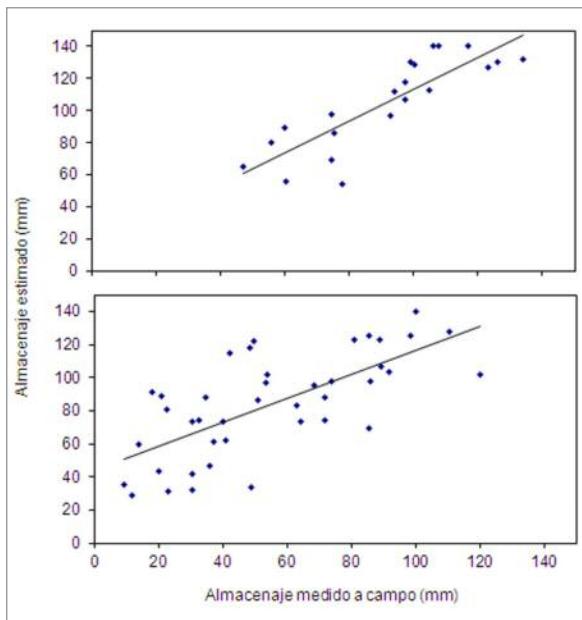


Figura 6. Gráficos de dispersión entre almacenaje estimado con el producto de BHS y el almacenaje medido a campo sobre lotes de trigo durante la última década de septiembre y octubre (arriba) y sobre lotes de soja durante la última década de febrero y marzo (abajo)

CONCLUSIONES

A partir de la precipitación estimada por RM e información complementaria, se desarrolló y presentó una serie de información cartográfica con indicadores a escala regional, en resolución espacial 1 km² y en distintas resoluciones temporales. Las imágenes de radar son generadas cada 10 minutos e integradas diariamente de manera automática, lo que permite contar con un dato actualizado cada 24 hs, de simple accesibilidad y con posibilidad de ser integrado en otras resoluciones temporales y generar nuevos productos. La información es factible de ser presentada en forma accesible en formatos gráficos digitales fácilmente utilizables.

El modelo de BHS implementado, considera las principales variables de entrada y salida de agua de los sistemas agrícolas de la región y se comprobó la sensibilidad del mismo para detectar las fluctuaciones temporales y espaciales de dicha variable. El mismo presenta como principal desventaja la realización de observaciones indirectas, es decir inferidas a partir de la reflectividad captada por sensores remotos. Operativamente en la actualidad se aplican métodos de extrapolación de datos puntuales

de estaciones meteorológicas. Esas técnicas implican grandes costos económicos y de tiempo, y se caracterizan por tener poca utilidad a escala regional debido a que no consideran la variabilidad espacial de las precipitaciones. En este contexto, las imágenes de radar se presentan como una metodología apta para el estudio espacial y temporal de esta variable climática. A su vez que posibilita la generación de indicadores útiles para valorar el desarrollo y los potenciales rendimientos de los cultivos regionales. De esta manera, constituye una herramienta de utilidad y operativa para la toma de decisiones en la planificación de actividades agropecuarias.

REFERENCIAS BIBLIOGRÁFICAS

- Belmonte, M.L. 2014. Estimación de precipitación a partir de radar meteorológico: evaluación y desarrollo de productos para la toma de decisiones en planificación agropecuaria. Tesis presentada para optar al grado de Magister en Ciencias Agrarias. Universidad Nacional del Sur.
- Belmonte, M.L., Fernández, M., Bellini Saibene, Y., Lorda, H., Fuentes, M.E.; Rossi, A., Garay, J., Rivarola, R. Ed. Alfredo Bono, Alberto Quiroga e Ileana Frasier. 2010. Caracterización tecnológica y productiva del cultivo de trigo y otros cereales de invierno para la región semiárida pampeana central. El cultivo de trigo en la región semiárida y subhúmeda pampeana. EEA Anguil INTA Ediciones.
- Belmonte, M.L., Fuentes, M., Bellini Saibene, Y., Fernández, J., Baez, A., Schaab, L., Ramos, L. 2011. Estado hídrico de los suelos de la región durante el ciclo de los cultivos. Anuario RIAN-RIAP 2009-2010. Boletín de divulgación técnica Nº 105. ISSN 0325-2167. (pp. 21-24).
- Della Chiesa, T. 2011. Estimación del balance de agua del suelo en la región pampeana utilizando datos provenientes de sensores remotos. Trabajo correspondiente al ciclo de intensificación para acceder al título de ingeniero agrónomo. Universidad de Buenos Aires. Facultad de Agronomía
- Di Bella, C.M., Rebella, C.M., Paruelo, J.M. 2000. Evapotranspiration estimates using NOAA AVHRR imagery in the Pampa Region of Argentina. International Journal of Remote Sensing. 21(4): (pp. 791-797).
- Fernández Long M.E., García Skabar, Y., Ferreira, L., Spescha, L. 2011. Balance De Agua En El Suelo Con Precipitación Estimada A Partir De Sensores Remotos. Presentado no XVII Congresso Brasileiro de Agrometeorología – 18 a 21 de Julho de 2011 – SESC Centro de Turismo de Guarapari, Guarapari - ES.
- Forte Lay, J.A., Villagra, M.M. 1983. El balance hidrológico

- de Thornthwaite y Mather (1955) aplicado bajo diferentes unidades de tiempo. Verificación con valores observados y comparación de resultados. Actas del XI Congreso Nacional del Agua. Córdoba (Argentina), mayo de 1983. Tomo I, Aguas Superficiales 1, (pp. 89-119).
- Galmarini, A.G. 1961. Caracterización climática de la provincia de La Pampa. Operación carnes (CAFADÉ). Publicación técnica 13.
 - Gematronik, 2005. Rainbow® 5 Products & Algorithms. Gematronik GmbH., Raiffeneisenstr. 10, 41470 Neuss, Germany
 - GVSIG. Software de SIG (Sistemas de Información Geográfica). <http://www.gvsig.org>
- Hartmann, T., Tamburino, M. S., Bareilles, F. 2010. Análisis preliminar de datos obtenidos por la red de radares del INTA para el estudio de precipitaciones en la Región Pampeana. 39º Congreso Argentino de Agroinformática. 2das Jornadas Argentinas de Informática. JAII0, sep. 2010
- Hoffman, J. 1988. Las variaciones climáticas ocurridas en la Argentina desde fines del siglo pasado hasta el presente. P 275-290. En: El deterioro del ambiente en la República Argentina. FECIC, Bs As, Arg.
 - INTA., Gobierno de La Pampa., UNLPam. 1980 Inventario integrado de recursos naturales de la provincia de La Pampa. ISAG Buenos Aires. 493 pp.
 - Lorda, H.; Roberto, Z.; Bellini, Y.; Sipowicz, A.; Belmonte, M. 2008. Descripción de zonas y subzonas agroecológicas RIAP. Área de Influencia de la EEA Anguil RIAN-RIAP. Boletín de divulgación técnica N° 96. ISSN 0325-2167
 - Marshall, J.; Palmer, W. 1948. The distribution of raindrops with size. J. Meteorol. 4, (pp. 186–192).
 - Pascale A., Damario E. 1977. El balance hidrológico seriendo y su utilización en estudios agroclimáticos. Rev. de la Fac. de Agronomía de La Plata. LIII (1-2), (pp 15-34).
 - Roberto, Z.E., Casagrande, G., Viglizzo, E.F. 1994. Lluvias en la pampa central. Tendencias y variaciones del siglo. Publicación N°2. INTA, Centro Regional La Pampa-San Luis, Argentina
 - Selex-Gematronik. 2008. Rainbow® 5 User guide. Products and Algorithms. Gematronik Weather Radar System. Germany. 298 pp
 - Viglizzo, E.F., Roberto, Z.E., Brockington, N.R. 1991. Agroecosystem performance in the semiarid pampas of Argentina and their interactions whit the environment. Agric. Ecosyst Environment 55, (pp. 7-16)
 - Viglizzo, E.F.; Lérpora, F.; Pordomingo, A.J.; Bernardos, J.N.; Roberto, Z.E; Del Valle.
-

Breve introducción a las tormentas de granizo los radares meteorológicos y las técnicas de Minería de Datos

4.

Bellini Saibene, Yanina¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
bellini.yanina@inta.gob.ar

RESUMEN

El análisis de tormentas severas que generan granizo es importante para el sector agropecuario, debido al alto impacto de los daños que ocasionan en los cultivos y en la infraestructura. Existen nuevas herramientas de recolección de datos remotos, como los radares meteorológicos de doble polarización que tiene el Instituto Nacional de Tecnología Agropecuaria (INTA) en su red, que permiten conocer la estructura de una tormenta, su posición geográfica y la composición de la misma. Este tipo de sensor remoto tiene grandes ventajas sobre redes terrestres de observación, ya que abarca una gran cantidad de superficie con una resolución de tiempo y espacio unificada. Debido a la naturaleza del instrumento, el radar genera un importante volumen de información que no puede ser procesado con herramientas tradicionales de análisis de datos. Por esta razón estudiar si las técnicas de minería de datos son adecuadas para el tratamiento de este tipo de información, aportará a incrementar el abanico de utilidades para aprovechar los datos registrados por los radares, aportando al conocimiento del granizo en la región y la generación de productos para el sector agropecuario. Este trabajo presenta conceptos básicos sobre minería de datos (DM), formación y comportamiento del granizo en Argentina y funcionamiento del radar meteorológico. Finalmente se detallan los antecedentes que usan en conjunto estos tres elementos.

Palabras clave: tormentas severas, riesgo agroclimático, teledetección, análisis de datos

ABSTRACT

The analysis of severe storms that generate hail are important for the agricultural sector, due to the high impact of the damages in crops and infrastructure. There are new tools for collecting remote data, such as the dual polarization meteorological radars from the National Institute of Agricultural Technology (INTA) network, which allow to know the structure of a storm, its geographical position and the composition of the storm. This type of remote sensor has great advantages over terrestrial observing networks, as it covers a large amount of surface with a unified resolution of time and space. Due to the nature of the instrument, radar generates an important amount of information that can not be processed with traditional data analysis tools. For this reason to study if the techniques of data mining are suitable for the treatment of this type of information, will contribute to increase the range of utilities to take advantage of the data recorded by the radars, contributing to the knowledge of hail in the region and the generation of products for the agricultural sector. This paper presents the basic concepts about data mining (DM), meteorological radar performance and hail training and behavior in Argentina. Finally we detail the works that use these three elements together.

Keywords: Severe storms, agro-climatic risk, remote sensing, data analysis

INTRODUCCIÓN

El Instituto Nacional de Tecnología Agropecuaria (INTA) cuenta con una red de tres radares meteorológicos (figura 1) cuyo objetivo se resume en mejorar

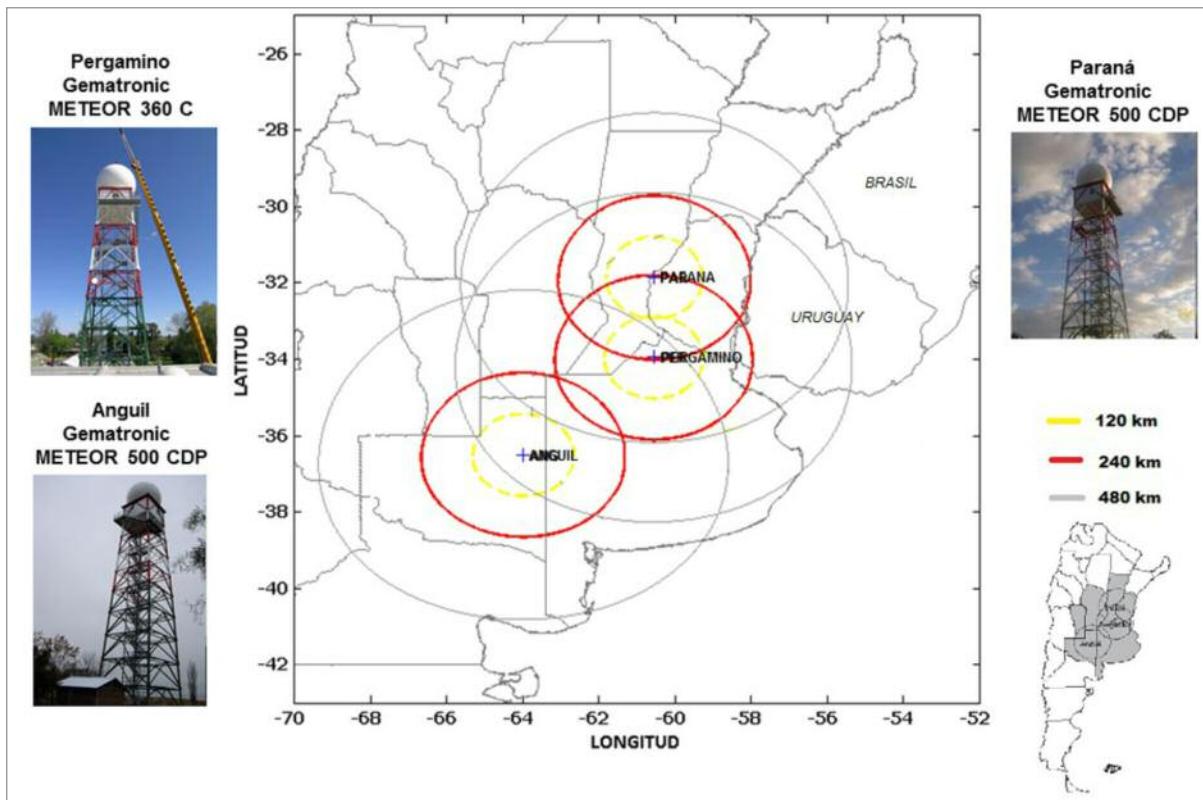


Figura 1. Mapa del área de influencia de 240 km de radio del radar de la EEA Anguil (Bellini Saibene, Y., 2015).

el conocimiento del ambiente climático en el que se desarrollan las plantas y los animales. Las fluctuaciones del clima constituyen uno de los factores que mayor incertidumbre generan dentro de un ecosistema agropecuario, contribuyendo a la complejidad en el proceso de la toma de decisiones y condicionando la eficiencia del manejo empresarial agropecuario (Casagrande, *et al.* 2006). El granizo es uno de los fenómenos meteorológicos capaz de infligir daños cuantiosos al deteriorar seriamente cosechas, edificios, infraestructura y medios de transporte (Ponce de Leon, 1985). En la región Pampeana Argentina¹, la máxima frecuencia de ocurrencia de granizo se encuentra en Córdoba y La Pampa (Mezher, *et al.* 2008, Mezher, *et al.* 2008b, Mezher, *et al.* 2008c, Mezher, *et al.* 2012). El estudio del granizo es de utilidad para: 1) los servicios meteorológicos nacionales y la aviación (alertas) (Hohl R., *et al.* 2002), 2) la industria de los seguros (estimación de daño, mitigación y transferencia, riesgo forense) (Hohl R., *et al.*

al. 2002; K. Casellas, *et al.*, 2012) y 3) para la comunidad agropecuaria (protección, mitigación y transferencia del riesgo, caracterización de riesgos agropecuarios, riesgo forense, lucha antigranizo) (Hohl R., *et al.* 2002; K. Casellas, *et al.*, 2012). Debido a la reducida extensión espacial y temporal de las tormentas de granizo, su detección y ocurrencia en superficie es una tarea difícil y costosa (Bustos C. y Videla H., 1982; Sánchez J. L., *et al.* 2013; Tabary P., *et al.* 2009; López L. y Sánchez J. L., 2009; L. López, *et al.* 2007), por lo que sensores remotos como los radares meteorológicos son una alternativa a las redes terrestres de mediciones, con la ventaja de abarcar una gran superficie y disponer de una única resolución en tiempo y espacio (Hohl R., *et al.* 2002).

Desde la década del 50 se han realizado numerosos estudios que exploran la relación de las variables medidas por los radares con el granizo y se han generados diferentes modelos con los objetivos de clasificar hidrometeoros, determinar su precipitación, tamaño o daño que ocasionaron; algunos de estos trabajos utilizan técnicas de minería de datos (DM) (ej.: Sánchez J.L., *et al.* 2013; Billet, J. *et al.*, 1997; Alexiuk,M. *et al.*, 1999; Li, P. C. *et al.*, 2000;

¹ La región Pampeana está compuesta por las provincias de Buenos Aires, este de La Pampa y centro sur de Santa Fe, Córdoba y Entre Ríos (Ligier 2009).

Holleman, I., 2001; Marzban C. y Witt A., 2001; Ramirez, L. *et al.* 2001; Gagne, D. J. *et al.* 2009). Estos modelos deben ser ajustados a la región de estudio y a los instrumentos de medición utilizados (Bustos C. y Videla H., 1982; López L. y Sánchez J. L., 2009; Marzano, F. S. *et al.*, 2007; Ceperuelo Mallafré, M., 2007). La mayor parte de la bibliografía presente para Argentina hace referencia a trabajos para radares de banda S, de simple polarización y localizados en la provincia de Mendoza (ej.: Bustos C. y Videla H., 1982; Sánchez J. L., *et al.* 2013; López L. y Sánchez J. L., 2009; Pérez, R.C., 2006; Simonelli, S.C. y Nicolini, M., 2000; Makitov, V., 2007). La técnica de DM más utilizada en estos trabajos es la regresión logística (ej.: Sánchez J. L., *et al.* 2013; López L. y Sánchez J. L., 2009). Los estudios que usan radares de banda C, analizan variables de simple polarización en Mendoza (ej.: Foote, G.B., 2005) y variables de doble polarización en las provincias de Buenos Aires, Entre Ríos y La Pampa (ej.: (Mezher, R.N., 2012b; Mezher, R.N. y Mercuri, P.A. 2010; Mezher, R.N., 2011). La técnica de DM más utilizada es la RL (Bustos C. y Videla H., 1982). En los trabajos realizados con datos de los radares de INTA analizan solamente la primera elevación de datos (ej: Mezher, *et al.* 2012; Bustos C. y Videla H., 1982).

Uno de los radares de INTA se localiza en la Estación Experimental Agropecuaria (EEA) Anguil, en La Pampa. Realiza mediciones cada diez minutos en un área de 240 km de radio generando un gran volumen de información diaria (Bellini Saibene, Y.; 2015). El radar cuenta con un software provisto por el fabricante que implementa el algoritmo Z-based Hail Warning (ZHAIL) (Germatronik, 2007) para la estimación de probabilidad de granizo. ZHAIL se basa en (Holleman, I., 2001) que fue desarrollado para radares de bandas X y S y de simple polarización (Holleman, I., 2001; Germatronik, 2007; Waldvogel, A. *et al.* 1979; Skripnikova, K., 2012); además de los datos del radar, necesita información de radio sondeos (Germatronik, 2007).

Las técnicas de DM permiten analizar grandes volúmenes de datos con el objetivo de encontrar relaciones impensadas y nuevas formas de resumir y presentar los datos que resulten útiles y comprensibles para el usuario (D. J. Hand, 2001).

En este contexto, ha sido de interés de INTA 1) examinar la utilidad del uso de técnicas de DM para el procesamiento de este tipo de información, 2) aprovechar las características de doble polarización

del radar, 3) generar herramientas que permitan minimizar el uso de software propietario para el manejo de los datos, 4) contribuir a la caracterización del comportamiento de las variables de doble polarización de banda C, para la identificación de granizo y daño en cultivos en la región de estudio, 5) generar herramientas que faciliten el acceso a los datos del radar para futuros estudios y 6) contribuir al entendimiento del uso de radares meteorológicos para aplicaciones agropecuarias.

En este trabajo presentamos al lector una breve introducción al granizo y su relación con el daño a la producción agropecuaria, los radares meteorológicos y las técnicas de DM para que cuente con los conceptos generales necesarios sobre sus funcionamientos. Finalmente se resumen los antecedentes de trabajos con los tres conceptos juntos: granizo, radar meteorológico y minería de datos.

MINERÍA DE DATOS

DM es la exploración y análisis por medios automáticos o semiautomáticos de grandes cantidades de datos con el fin de descubrir reglas y patrones significativos. Es parte del proceso de descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases - KDD) que se define como la extracción no trivial, desde los datos, de información implícita, previamente desconocida y potencialmente útil (Ale, J. M., 2006).

La creciente disponibilidad de información meteorológica y climática (ej.: registros históricos, simulaciones de modelos numéricos, datos de radares meteorológicos y satélites, etc.) hace necesario el uso de nuevas técnicas más eficientes y automáticas que las clásicas herramientas estándar de modelización estadística (Gutiérrez, J. M. *et al.* 2004), por lo que DM está emergiendo como un método adecuado para la extracción de patrones a partir de conjuntos amplios de datos heterogéneos relacionados con la predicción de fenómenos meteorológicos (Gutiérrez, J. M. *et al.* 2004; Bartok, J., 2010).

Estas herramientas deben funcionar bajo la restricción que los datos disponibles son observacionales (en contraposición a datos experimentales) debido que en la mayoría de los casos se trabaja con datos que han sido recolectados para un propósito diferente al análisis de DM. Esto implica que la

estrategia de recolección no tuvo en cuenta el objetivo del análisis; a diferencia del análisis estadístico, donde los datos son recolectados utilizando estrategias eficientes para contestar preguntas específicas (D. J. Hand, 2001).

Una segunda restricción es que los conjuntos de datos analizados con DM son grandes e involucran problemas relacionados con su manejo (almacenamiento, mantenimiento, accesibilidad, visualización, etc.) y con su análisis (determinar la representatividad de los datos, analizar los datos en un tiempo razonable, decidir si un patrón encontrado realmente refleja la realidad subyacente, etc.) (D. J. Hand, 2001).

El método seguido en el proceso de DM es una mezcla de los métodos matemáticos y científicos, existen diversas caracterizaciones del mismo, una de las más difundida es Cross-Industry Standard Process for Data Mining (CRISP-DM) y consta de seis fases concebidas como un proceso cíclico (Figura 2) (Olson D. L. y Delen, D., 2008; Nisbet, R. et al., 2008):

1) *Comprender “el negocio”*: el término “negocio” hace referencia al área de conocimiento en la cual se aplicará DM. Esta fase incluye determinar los

objetivos, evaluar la situación actual, establecer tareas de DM, y desarrollar un plan de proyecto (Bartok, J. et al., 2010; Olson D. L. y Delen, D., 2008; Nisbet, R. et al., 2008).

2) *Comprender los datos*: este paso incluye la recopilación, descripción y exploración de los datos. Se verifica la calidad de los mismos y se realizan los cálculos de las estadísticas básicas de los principales atributos y sus correlaciones. Se generan visualizaciones (Bartok, J. et al., 2010; Olson D. L. y Delen, D., 2008; Nisbet, R. et al., 2008).

3) *Preparar los datos*: es la fase más compleja y la que más tiempo insume (60% a 70% del tiempo total del proceso). Identificados los datos disponibles, es necesario seleccionar, limpiar y transformarlos para la fase de modelado. Incluye una exploración más profunda (Bartok, J. et al., 2010; Olson D. L. y Delen, D., 2008; Nisbet, R. et al., 2008).

4) *Modelar*: es el núcleo del proceso de DM. Se selecciona el método más adecuado de acuerdo a los datos disponibles y los objetivos del análisis. Cada método tiene sus requisitos para los datos de entrada, por lo tanto un cambio de método puede

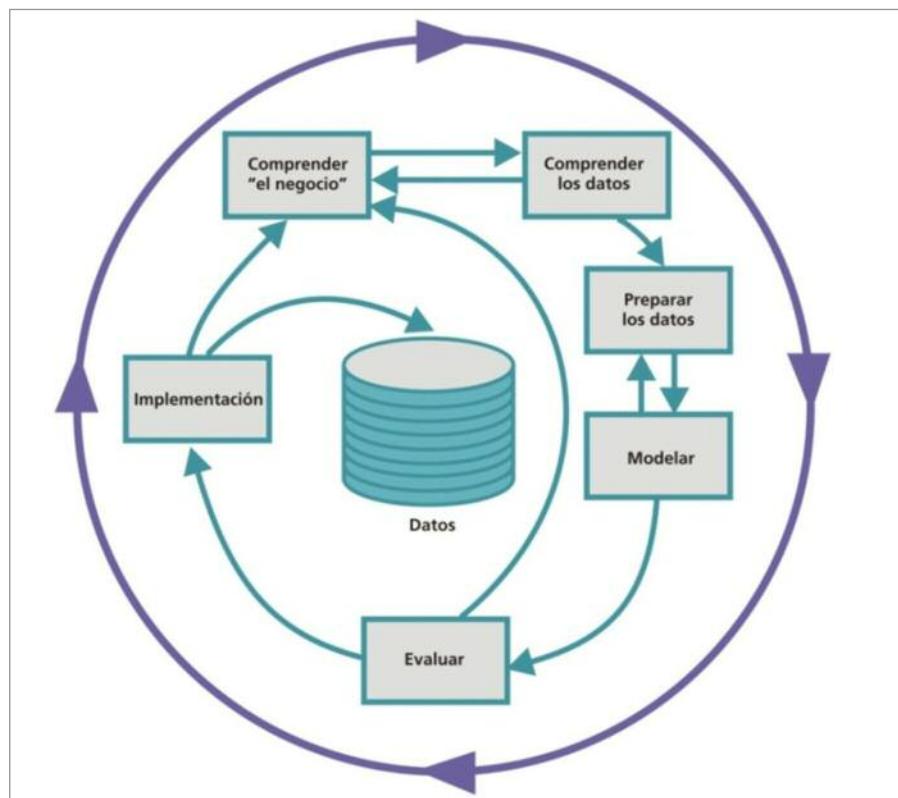


Figura 2. Fases del proceso de CRISP-DM (Adaptado de:Nisbet, R. et. al. 2008).

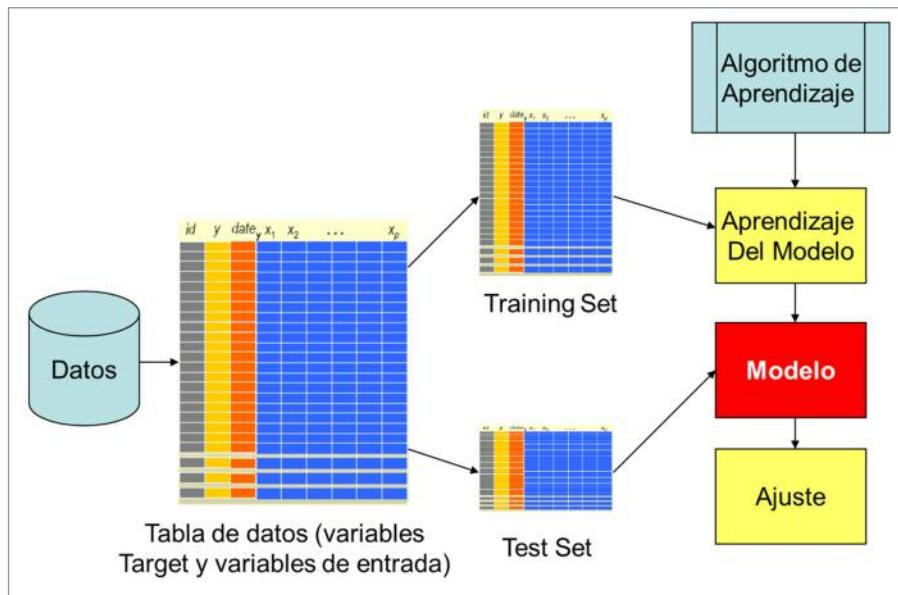


Figura 3. Esquema general de aprendizaje supervisado en minería de datos
(Fuente: elaboración propia)

conducir a la necesidad de repetir parte de la etapa de preparación de datos (Bartok, J. *et al.*, 2010; Olson D. L. y Delen, D., 2008; Nisbet, R. *et al.*, 2008).

5) Evaluar: los resultados del modelo son evaluados en el contexto de los objetivos establecidos en la primera fase y teniendo en cuenta el costo de realizar una clasificación incorrecta. La evaluación se basa en indicadores cuantitativos de la calidad de los modelos creados midiendo la tasa de error de los casos mal clasificados en un conjunto de datos con el cual el modelo trabaja por primera vez (Bartok, J. *et al.*, 2010; Olson D. L. y Delen, D., 2008; Nisbet, R. *et al.*, 2008; Witten, I. H., *et al.*, 2011).

6) Implementación: en esta fase se desarrolla un plan de implementación práctica para el modelo. También se determina la estrategia de supervisión y mantenimiento del mismo para asegurar su confiabilidad con el paso del tiempo (Bartok, J. *et al.*, 2010; Olson D. L. y Delen, D., 2008; Nisbet, R. *et al.*, 2008).

Proceso de aprendizaje automático

El modelado y la evaluación tienen diferentes técnicas de acuerdo al proceso de aprendizaje automático utilizado, existen dos formas básicas: aprendizaje supervisado cuando se cuenta con un conjunto de datos previamente etiquetado y aprendizaje no supervisado cuando en el set de datos no contamos con la información etiquetada o los valores de

la variable objetivo. La figura 3 presenta el esquema general utilizado para desarrollar un modelo utilizando aprendizaje supervisado. Los datos representan nuestra fuente de información del pasado, esto genera un conjunto o set de datos.

Ese set de datos tiene al menos tres variables: la variable objetivo o variable target (y en la figura 3), una variable de identificación (id en la figura 3) que permite identificar cada caso de forma única y las variables de entrada ($X_1 \dots X_n$ en la figura 3). En algunos casos la fecha también es una variable importante (date en la figura 3). Ese conjunto de datos, se divide en al menos dos partes: una para entrenar (training set) y otra para testear el modelo (test set) y asegurarse que no “aprendió” el conjunto de datos, si no que logró realizar una generalización de los patrones subyacentes.

Luego de la división de los datos, los datos para entrenar son utilizados por el algoritmo seleccionado (por ejemplo árboles de decisión, regresión lineal o redes neuronales) para construir un modelo, ese modelo se valida con el conjunto de entrenamiento y se mide su ajuste, volviendo a repetir todo el ciclo hasta lograr un modelo que tenga un buen rendimiento (figura 3). Una práctica que permite realizar una mejor validación es llamada cross-validation o validación cruzada, donde se crean un número de muestras al azar (en general 10) y se genera entrena y testeaa el modelo ese número de veces cambiando el set de testeoo y entrenamiento.



Figura 4. Diferentes tipos de granizo (Adaptado de: Eroski Consumer, 2012).

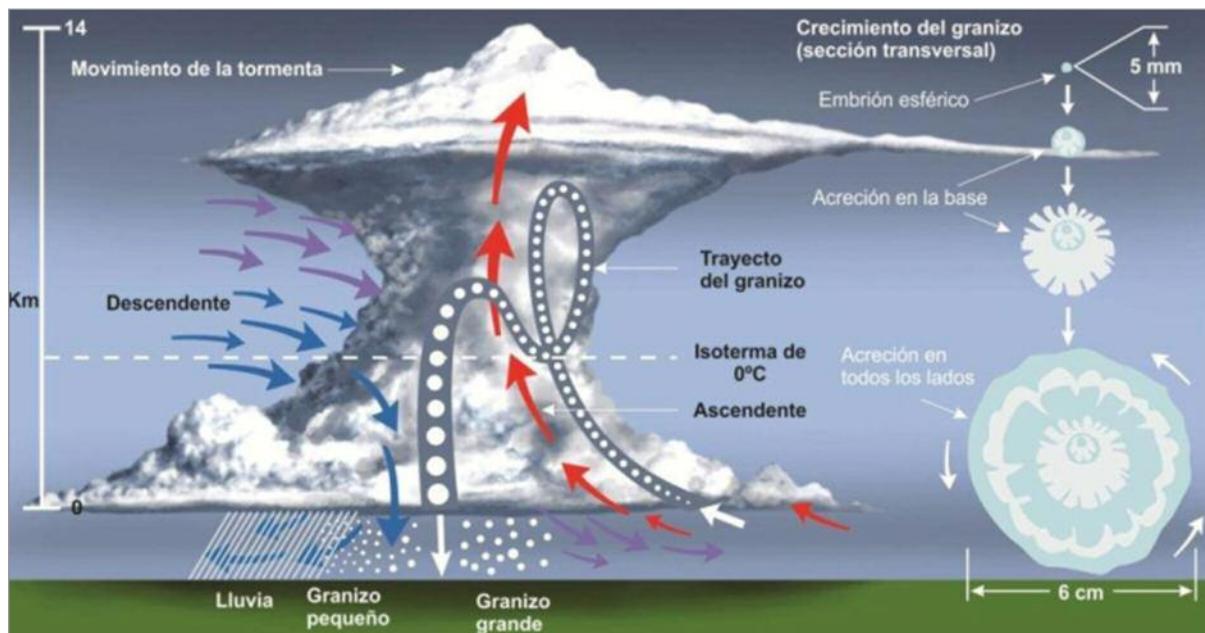


Figura 5. Crecimiento del granizo dentro de una tormenta (Adaptado de Encyclopedia Britannica, 2012).

EL GRANIZO

El granizo es un fenómeno meteorológico que consiste en un tipo de precipitación formado por partículas de hielo (conglomerado policristalinos) de diámetro igual o superior a 5mm (Ceperuelo Mallafré, M., 2007). Generalmente son esféricos y su superficie externa puede ser irregular, con lóbulos y protuberancias (figura 4) (Ponce de Leon, 1985).

Es considerado un riesgo agroclimático² y produce importantes pérdidas económicas en cultivos (Ceperuelo Mallafré, M., 2007; Mezher, R.N. y Mercuri, P.A. 2010; Katz R. W. y Garcia, R. R., 1981; Simeonov, P., 1996; Schiesser, H. H., 1990; Sánchez, J. L. et al. 1996). Más del 90% de los seguros agropecuarios en la Argentina son relativos a coberturas de granizo (SISN, 2013).

Las tormentas que pueden generar granizo son del tipo convectivo³ y pueden estar formadas por una o varias “celdas”. Las celdas son una unidad dinámica caracterizada por una fuerte corriente

ascendente (Ponce de Leon, 1985). Durante el transcurso de la tormenta, la celda pasa por diferentes estados. En la etapa de maduración, que dura entre 15 y 30 minutos, la nube se denomina Cúmulonimbus y se pueden dar condiciones para la formación de granizo ya que al ser una nube de gran desarrollo vertical se produce glaciación en su parte superior (Ponce de Leon, 1985; Maderey Rascón, L. E. et al. 2005). Se caracterizan por presentar una marcada actividad eléctrica y generalmente pueden generar intensas precipitaciones (Ponce de Leon,

² Se denomina riesgo agro-climático a la probabilidad de afectación del rendimiento o la calidad de los cultivos por efecto de un fenómeno climático adverso (Almada et al.. 2005).

³ Este tipo de tormenta tiene su origen en la inestabilidad de una masa de aire más caliente que las circundantes. La masa de aire caliente asciende, se enfria, se condensa y se forma la nubosidad denominada Cumuliforme que tienen gran desarrollo vertical. El ascenso de la masa de aire se debe, generalmente, a un mayor calentamiento en superficie (Maderey Rascón, Roman, y others 2005).

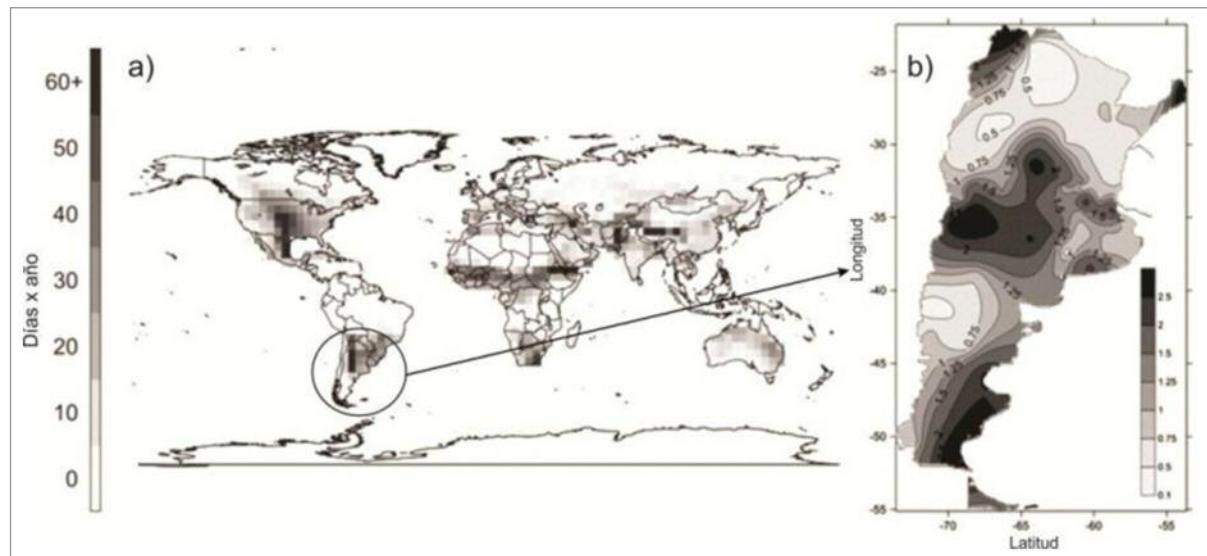


Figura 6a. Días por año con condiciones capaces de generar tormentas fuertes con granizo. (Fuente: Brooks, H. E. et al. 2003), 6b. Promedio anual de eventos de granizo en Argentina para la serie 1960-2008 (Fuente: Mezher, et al. 2012).

1985; Maderey Rascón, L. E. et al. 2005).

En el interior de la nube, el aire asciende y se enfriá. El vapor de agua contenido, se satura y se forman gotitas de agua que son arrastradas por las corrientes ascendentes. Al continuar el ascenso, en los niveles de temperaturas inferiores a cero grado centígrado, algunas gotas se congelan y forman cristales de hielo. Esta transformación no es abrupta; las gotas pequeñas, de agua pura, pueden permanecer en estado líquido hasta 30°C bajo cero en un estado metaestable⁴; se habla de gotas de agua sobreenfriadas. Como consecuencia, en las zonas en que la temperatura varía entre -4°C y -30°C , la nube está formada por una mezcla de gotas de agua y cristales de hielo (Ponce de Leon, 1985) (figura 5).

Mientras el aire asciende, algunas gotas de agua chocan con los cristales de hielo y se unen formando cristales y gotas de mayor tamaño. Las gotas más grandes se congelan antes que las de menor tamaño y por ser más pesadas comienzan a caer, llevando

consigo a las gotas más pequeñas durante el descenso. Comienza así el proceso de crecimiento del granizo por choque y congelación de gotas sobreenfriadas (Ponce de Leon, 1985).

Después de esta etapa, algunos granizos son suficientemente pesados como para iniciar el descenso, a pesar de la resistencia que les opone la corriente ascendente. Durante el pasaje por la parte inferior de la nube y en el trayecto hasta llegar al suelo, las piedras comienzan a fundirse. Las más pequeñas pueden transformarse totalmente en agua, dando lugar a la formación de grandes gotas de lluvia, mientras que las de mayor tamaño sólo funden su capa externa y llegan al suelo en estado sólido (Ponce de Leon, 1985) (figura 5).

Las condiciones más favorables para el desarrollo de estas tormentas se producen en horas de la tarde-noche, durante la primavera y el verano y sobre el centro de los continentes y latitudes medias, salvo en el continente africano (figura 6a) (Ponce de Leon, 1985; Brooks, H. E. et al. 2003). La caída de granizo se produce en regiones puntuales, localizadas y durante períodos de tiempo de corta duración (Ponce de Leon, 1985; Brooks, H. E. et al. 2003). En la Argentina la mayor frecuencia de granizo ocurre en las regiones montañosas, costeras y comprendidas entre las latitudes de 30° y 40° S (Mezher, et al. 2008; Mezher, et al. 2012) (figura 6b). La distribución anual, en las áreas orientales y costeras, presenta la mayor cantidad de eventos en los

⁴ La metaestabilidad es la propiedad que un sistema con varios estados de equilibrio, tiene que exhibir, durante un considerable espacio de tiempo. Bajo la acción de perturbaciones externas dichos sistemas exhiben una evolución temporal hacia un estado de equilibrio fuertemente estable. Como ejemplo de metaestabilidad física se presenta el agua en sobrefusión. Las gotas de agua pura en suspensión en un aire también muy puro no se congelan a los 0°C , sino que siguen en estado líquido hasta alcanzar los -30°C . Este estado de sobrefusión cesa bruscamente cuando la gota entra en contacto con un cuerpo externo (como un cristal de hielo) (Wikipedia contributors 2012b).

meses de primavera, pudiendo comenzar en invierno y continuar hasta el comienzo del verano (Mezher, et al. 2008; Mezher, et al. 2012). En el oeste y centro del país, los eventos también predominan en primavera, pero las frecuencias máximas se observan durante los meses de verano. (Figura 6b) (Mezher, et al. 2008; Mezher, et al. 2012). Sobre la región Pampeana, la máxima frecuencia de ocurrencia se encuentra en Córdoba (durante la primavera) y La Pampa (en primavera y verano) (Mezher, et al. 2008, Mezher, et al. 2012, Mezher, et al. 2008b). Esta frecuencia es mínima en otoño-invierno hasta julio (Mezher, et al. 2012).

EL RADAR METEOROLÓGICO

RADAR es el acrónimo de radio detection and ranging (Rinehart, R. E., 1997). En la figura 6a se presenta un diagrama con sus componentes básicos. En Rinehart, R. E., 1997 y Bringi y Chandrasekar, 2001 se explican el detalle de su funcionamiento y aplicaciones.

Los radares emiten a la atmósfera pulsos de energía electromagnética en el rango de frecuencias de las microondas. La mayoría de los radares meteorológicos operan en las longitudes de onda de 3 cm (banda X), 5 cm (banda C) y 10 cm (banda S)

(Rinehart, R. E., 1997). Cuando dichos pulsos encuentran una partícula, parte de esa energía es absorbida y el resto se dispersa en todas direcciones, devolviendo una fracción hacia la antena del radar que, normalmente, se usa tanto para transmitir como para recibir (figura 7b) (Rinehart, R. E., 1997; Peters, J. F. et al. 2003; UCAR, 2012).

Los pulsos se emiten mientras la antena gira 360 grados en forma horizontal o acimut. Estos giros (revoluciones) comienzan con una elevación cercana a los 0 grados y luego aumenta el ángulo de elevación entre 10 y 12 veces, hasta llegar aproximadamente a 20 grados (figura 7c). Esta exploración, que se ejecuta en aproximadamente 5 minutos como mínimo (UCAR, 2012), se llama barrido o escaneo del volumen porque explora el volumen completo que rodea al radar adquiriendo información sobre la estructura vertical y distribución de la tormenta (Ceperuelo Mallafré, M., 2007) (figura 7c).

El factor de reflectividad Z o reflectividad radar se calcula a partir de la medición de la energía del eco reflejado. Es una medida de la capacidad de los blancos de interceptar y devolver la energía del radar, depende de los parámetros físicos del objeto (ej.: forma, composición) y la relación entre la longitud de onda y el tamaño del objeto (Rinehart, R. E., 1997; UCAR, 2012; Aydin, K. et al. 1986). Mientras la energía se propaga por el aire o por la tormenta se

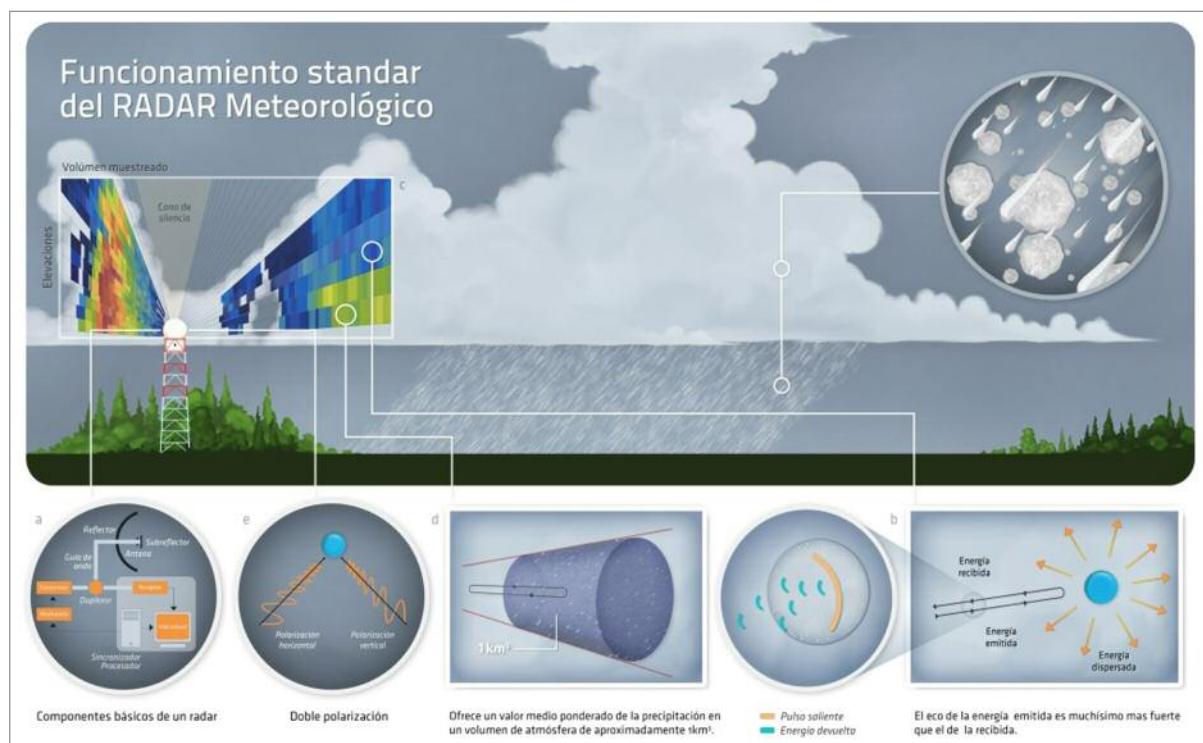


Figura 7. Funcionamiento estándar del RADAR Meteorológico (Adaptado de Rinehart, R.E., 1997 y Schuur, 2003)

produce una pérdida (atenuación) de la señal, que es más pronunciada para las longitudes de onda más cortas (Rinehart, R. E., 1997). La ecuación (1) muestra que Z es la suma de la sexta potencia de los diámetros (D) de todos los blancos que dispersan la energía transmitida en la unidad de volumen muestreada (Rinehart, R. E., 1997; UCAR, 2012).

$$Z = \sum_{i=1}^n D_i^6 \quad (1)$$

La dependencia de la sexta potencia significa que las partículas grandes predominan en el valor de Z calculado, por lo que la presencia de pocas gotas grandes producen el mismo valor de Z que muchas gotas pequeñas (Figura 8) (Rinehart, R. E., 1997; UCAR, 2012). Este factor se mide en decibelios (dBZ). Un alto valor de dBZ significa alta densidad de partículas por unidad de volumen (Peters, J. F. et al. 2003).

La radiación electromagnética de los radares tiene una orientación preferencial porque está polarizada (Rinehart, R. E., 1997). Los radares de simple polarización utilizan una sola polarización horizontal (Rinehart, R. E., 1997), en los de doble polarización, el haz electromagnético emitido, tiene orientación vertical y horizontal (figura 7e) (Ceperuelo Mallafré, M., 2007).

La comparación de diferentes maneras (proporciones, correlaciones, etc.) de estas dos potencias reflejadas permiten obtener información sobre el tamaño, forma y densidad de las partículas de las nubes (Bringi y Chandrasekar, 2001) y es por esta razón que los radares de doble polarización son más precisos en la detección de hidrometeoros (Foote, G.B., 2005; Bringi y Chandrasekar, 2001; Aydin, K. et al. 1986; Ryzhkov, A. et al. 2012; Anderson, M. E. et al. 2012; Heinselman P. L. y Ryzhkov, A. V. 2012; Brandes E. A. y Ryzhkov, A. V., 2004; Ryzhkov, A. V.

et al. 2005).

Las medidas polarimétricas que se utilizan para la detección de granizo y están disponibles en el radar de la EEA Anguil son:

- **Differential Reflectivity. Reflectividad diferencial (ZDR):** valores de -8 a +12 dB (decibelios).

Muestra la diferencia en la energía devuelta entre los impulsos horizontales y verticales del radar (Rinehart, R. E., 1997; Bringi y Chandrasekar, 2001; Aydin, K. et al. 1986; Schuur, 2003). Valores de ZDR cercanos a 0 indican que los hidrometeoros en el volumen tienen, en promedio, una forma esférica (su eje vertical es casi igual a su eje horizontal), si los valores son >0, los hidrometeoros en el volumen están orientados horizontalmente (su eje horizontal es más largo que su eje vertical), mientras que si ZDR < 0 significa que los hidrometeoros en el volumen están orientados verticalmente (su eje vertical es más largo que su eje horizontal) (figura 9).

- **Correlation Coefficient. Coeficiente de correlación co-polar (ρ_{HV} , RhoHV, CC):** valores entre 0 y 1.

Mide la similitud del comportamiento (características de retrodispersión) de los pulsos horizontales y verticales en un volumen (Rinehart, R. E., 1997; Bringi y Chandrasekar, 2001; Schuur, 2003). Su precisión disminuye con la distancia y con la mezcla de tipos de hidrometeoros en el volumen. Valores de RhoHV < 0.80 evidencian una dispersión compleja: los pulsos horizontales y verticales cambian de diferentes maneras de pulso a pulso. Valores de 0.80 a 0.97 muestran diferencias moderadas entre los pulsos y valores de RhoHV > 0.97 implican una dispersión ordenada, por lo tanto hay muy pequeñas diferencias entre los pulsos horizontales y verticales. Para esferas perfectas RhoHV = 1 (Rinehart, R. E., 1997; Schuur, 2003) (figura 8).

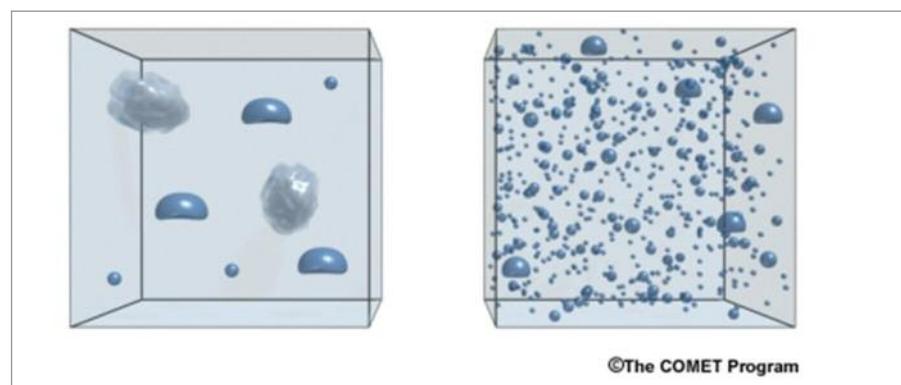


Figura 8. Muestras volumétricas con valores de Z equivalentes (Adaptado de UCAR, 2012).

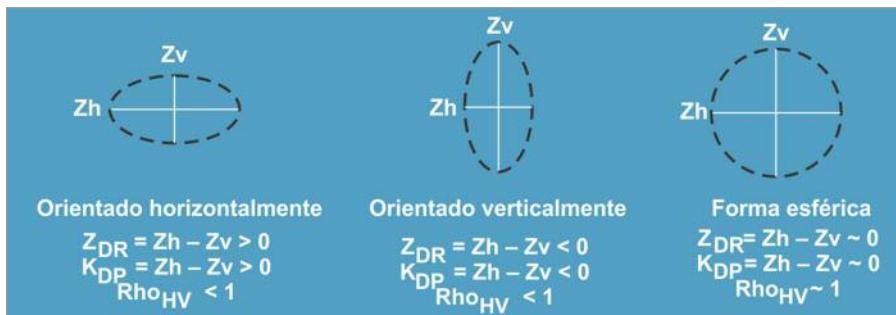


Figura 9. Relación entre los ejes verticales y horizontales de los hidrometeoros y las variables registradas por el radar (Fuente: Bellini Saibene, Y., 2015).

- **Differential Propagation Phase. Fase de propagación diferencial (ϕDP , ΦDP):** valores de 0° a 360° (grados). Las ondas electromagnéticas cambian de fase mientras pasan a través de material más denso (Rinehart, R. E., 1997). Dependiendo de la forma, orientación y concentración de los hidrometeoros, se producen ligeros cambios de fase entre las señales vertical y horizontal, retardándose una con respecto de la otra. El agua líquida es uno de los medios que más resistencia presenta (Rinehart, R. E., 1997; Bringi y Chandrasekar, 2001; Boettcher, J., 2014). Por ejemplo, si los hidrometeoros son oblatos (como las gotas de lluvia), ΦDP aumenta porque la fase horizontal se propaga más despacio que la vertical. Esta variable no dice nada por sí misma, pero su cambio en el tiempo y en el espacio sí, por eso se calcula la variable Fase diferencial específica (KDP).

- **Specific Differential Phase. Fase diferencial específica (KDP):** puede tomar valores de -18 a $+36$ $^\circ/km$ (grados por kilómetro). Deriva de ΦDP (Boettcher, J., 2014). Es la diferencia, en un rango, entre los impulsos horizontales y verticales del radar a medida que se propagan a través de un medio y son atenuadas (Rinehart, R. E., 1997; Schuur, 2003). La correspondencia de valores positivos (horizontal) y negativos (vertical) es análoga a ZDR. KDP es dependiente de la concentración de partículas, por lo que muchas gotas pequeñas pueden generar un valor positivo mayor que pocas gotas grandes (Rinehart, R. E., 1997; Schuur, 2003; Boettcher, J., 2014).

Como se ha mencionado, estudiar y analizar el granizo es importante para el sector agropecuario debido al alto impacto de los daños que ocasionan en los cultivos. La combinación de la minería de datos y los radares meteorológicos se presenta como una herramienta potencialmente útil para rea-

lizar estos análisis y estudios. A continuación se realiza una revisión de criterios de detección de granizo y daño tanto en cultivos como en infraestructura, basados en radares meteorológicos y minería de datos.

Existen numerosos estudios sobre la relación de las señales registradas por los radares y el granizo. Los primeros exploraron datos de simple polarización solos o combinados con datos de otros sensores (Heinselman P. L. y Ryzhkov, A. V. 2012). Con la aparición de los radares de doble polarización, los trabajos se concentraron en estudiar la relación entre ambas reflectividades que permite distinguir entre la forma esférica y rotativa de los granizos versus la forma no esférica de las gotas de lluvias (Holleman, I. et al., 2000). Finalmente, las investigaciones más recientes, estudian las diferencias de comportamiento de las señales de diferentes bandas (C vs. S) ante el granizo, ya que las longitudes de onda más cortas tienen mayor atenuación, especialmente ante la presencia de lluvia intensa (Sánchez J. L., et.al. 2013; Ryzhkov, A. et al. 2012; Anderson, M. E. et al. 2012; Kaltenboeck R. y Ryzhkov, A., 2012; Alberoni, P. P. et al. 2002).

En Holleman, I., 2001 y Holleman, I. et al., 2000 se realizan revisiones muy completas de métodos de simple polarización para detectar ocurrencia de granizo. En la tabla 1 se resumen los principales métodos y las variables que utilizan; en síntesis todos hacen referencia a la presencia de núcleos de Z por encima de un umbral (36 a 55 dBZ) y en algunos casos la combinan con información de temperatura y variables calculadas a partir de Z, como la energía cinética (E) (Holleman, I., 2001; Ceperuelo Mallafré, M., 2007; Foote, G.B., 2005; Germatronik, 2007; Ryzhkov, A. V. et al. 2005; Holleman, I. et al., 2000). Finalmente, en Skripnikova, K., 2012 se aplica un criterio de combinación de cuatro métodos y se clasifica cada pixel de acuerdo a la mayor coincidencia de al menos dos de los métodos, logrando

Método	Variables predictoras
CAPPI	Valor de Z a una altitud baja y constante. Se sugiere un umbral de 55 dBZ para distinguir entre granizo y lluvia utilizando este método (figura 10a) (Holleman, I. et. al., 2000).
Máxima reflectividad y persistencia de las máximas reflectividades.	Máximo valor de Z observado en una columna vertical. Umbral: 45 dBZ (figura 10b) (Holleman, I., 2001; Ceperuelo Mallafré, M., 2007).
Vertically Integrated Liquid (VIL)	Total de agua líquida presente en una columna sobre cierta posición. A mayor valor, mayor probabilidad de granizo. Discrepancia entre autores sobre el mejor umbral de VIL (Ceperuelo Mallafré, M., 2007; Holleman, I. et. al., 2000).
VIL Density	División de VIL por la altura del tope del eco del radar. Se propone como umbral universal para alarma de granizo 3.5 g/m³. Existen estudios que cuestionan este trabajo (Ceperuelo Mallafré, M., 2007; Holleman, I. et. al., 2000).
Reflectividad y temperatura de tope de nube	Combinación de Z (entre 36 y 53 dBZ) a baja altitud y temperatura en el tope de la nube (entre -11°C y -55°C.) (Holleman, I. et. al., 2000) (figura 10c).
Hail Detection Algorithm (HDA)	La primera versión tenía como parámetros: un núcleo de $Z \geq 50$ dBZ, ecos mayores a 55 dBZ entre los 5 y 12 km de altitud y presencia de ecos igual o más altos que 8 km (figura 10d) (Ryzhkov, A. V. et. al. 2005; Holleman, I. et. al., 2000). Las otras versiones corresponden a las de Waldvogel, Witt y Holleman (Holleman, I., 2001).
Severe Hail Index (SHI)	Se calcula por una integración vertical del producto del flujo de E del granizo con una función de temperatura y de Z (figura 10e) (Holleman, I., 2001; Ceperuelo Mallafré, M., 2007)
Waldvogel, Witt, Holleman	Núcleo de $Z \geq 45$ dBZ a 1.4 km o más por encima del nivel de congelamiento. Una diferencia de 1.6 Km corresponde a un 10% de probabilidad de granizo y una de 6 km es 100% de probabilidad (figura 10f) (Holleman, I., 2001; Ceperuelo Mallafré, M., 2007; Holleman, I. et. al., 2000). ZHAIL está, basado en estos algoritmos. Una explicación detallada del mismo se presenta en (Germatronik, 2007).

Tabla 1. Resumen de variables predictoras de granizo mediante métodos de simple polarización.

mejorar el rendimiento con respecto a cada método de forma individual.

En el caso de la doble polarización, la tabla 2 resume los valores que toma cada variable polarimétrica en presencia de granizo. En Anderson, M. E.

et al., 2012 y Brandes E. A. y Ryzhkov, A. V., 2004 se detallan las técnicas de doble polarización y en Kaltenboeck R. y Ryzhkov, A., 2012 se describen las diferencias de comportamiento encontradas entre las bandas S y C en presencia de granizo. La mayo-

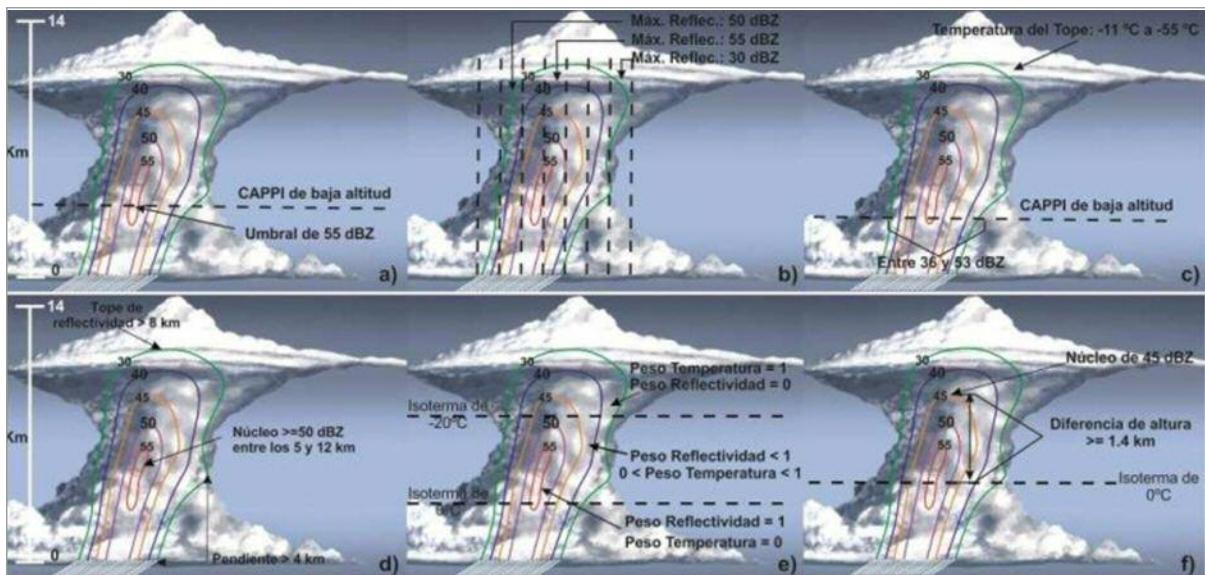


Figura 10. Esquema de variables utilizadas por el método: a) CAPPI (Bellini Saibene, Y., 2015), b) máxima reflectividad (Bellini Saibene, Y., 2015), c) Reflectividad y temperatura de tope de nube (Bellini Saibene, Y., 2015), d) HDA (Adaptado de: (Holleman 2001)), e) SHI (Adaptado de: (Holleman 2001)), f) Waldvogel, Witt, Holleman (Adaptado de: (Holleman 2001)).

ría de los algoritmos para banda S se basan en la anticorrelación que existe entre altos valores de Z y pequeños valores de ZDR cuando hay granizo. También se ha encontrado valores bajos de RhoHV ante la presencia de este hidrometeoro (Bringi y Chandrasekar, 2001; Anderson, M. E. et al. 2012).

La comparación de diferentes maneras (diversos autores sugieren que esta anticorrelación se puede

aplicar directamente en banda C (Anderson, M. E. et al. 2012), pero existen estudios con evidencias que refutan esta teoría (Tabary P., et al. 2009; Mezher, R.N., 2012b; Anderson, M. E. et al. 2012; Kaltenboeck R. y Ryzhkov, A., 2012). En banda C se encontraron valores elevados de Z, valores altos de ZDR, valores bajos y negativos de KDP y una abrupta caída de RhoHV (Mezher, R.N., 2012b; Anderson,

Variable	Valores para Banda S	Valores para Banda C
Z*	> 45 dBZ > 50 Dbz	Alto > 50 dBZ Granizo pequeño (< 2 cm): 50 a 60 dBZ Granizo grande (> 2 cm): 55 a 65 dBZ Granizo con lluvia: 45 a 80 dBZ
Rho _{HV}	0.90 < Rho _{HV} < 0.98	< 0.94, < 0.95 Granizo pequeño (< 2 cm): 0.92 a 0.95 Granizo grande (> 2 cm): 0.90 a 0.92 Granizo con lluvia: > 0.90
Z _{DR} *	< 1 dB granizo solo 1 dB <= Z _{DR} <= 4 dB granizo rodeado de lluvia.	Alto (3 – 8 dB) Granizo pequeño (< 2 cm): -0.5 a 0.5 dB Granizo grande (> 2 cm): -1 a 0.5 dB Granizo con lluvia: -1 a 6 dB > 4 dB, > 3 dB
K _{DP}		Alto > 4°/Km

* Estas variables son las más discriminantes de acuerdo a Alberoni, P.P. et.al. 2002 y Marzano, F. S. et.al. 2008.

Tabla 2. Listado de variables polarimétricas y sus valores ante la presencia de granizo para banda S (Tabary P., et al. 2009; Bringi y Chandrasekar, 2001; Anderson, M.E. et. al. 2012; Heinselman P.L. y Ryzhkov, A.V. 2012; Brandes E.A. y Ryzhkov, A.V., 2004) y banda C (Tabary P., et. al. 2009; Anderson, M. E. et. al. 2012).

M. E. et al. 2012). Algunos autores sugieren que la relación ZDR<1.0 dB y Z>50 dBZ se debe observar en radares de banda C cuando los granizos son grandes (12 a 15 mm de diámetro) y en mucha cantidad como para dominar Z y aparecer como esferas (Anderson, M. E. et al. 2012).

Se utiliza el enfoque de lógica difusa como metodología más difundida para la generación de algoritmos de clasificación (Marzano, F. S. et al. 2008), la cual que se explica en detalle en Marzano, F. S. et al., 2007 y Alberoni, P. P. et al. 2002. Brevemente, esta técnica se basa en una combinación de funciones miembro asociada con un tipo de hidrometeoro en particular. Estas funciones tiene dos argumentos: Z y una de las variables polarimétricas. La clasificación se hace tomando el valor más alto de la combinación de funciones para cada píxel de la imagen (Alberoni, P.P. et al. 2002). El uso de esta técnica se justifica en el hecho que en una tormenta, el granizo se superpone con otros hidrometeoros, presentando límites poco claros, por lo que se necesita buscar una solución sin límites rígidos (Alberoni, P. P. et al. 2002).

Las funciones obtenidas muestran un alto grado de flexibilidad y pueden ser empíricamente adaptadas a evidencias experimentales (Marzano, F. S. et al. 2008).

Los primeros trabajos en relacionar el granizo con el daño en cultivos⁵ utilizaron redes de granizómetros junto a información de compañías de seguros y redes de observadores a campo, determinando las variables que permiten predecir el nivel de daño en cultivos por granizo, siendo la más significativa la E alcanzada por el hidrometeoro. Como las redes de granizómetros son muy costosas, se realizaron estudios para relacionar las variables que caracterizan la caída del granizo por medio de la reflectividad medida por un radar meteorológico (Waldvogel, A. et al. 1978). Los resultados para banda S y C determinaron una relación empírica (Waldvogel, A. et al. 1978), de la forma:

$$E = a \times 10^{-6} \times 10^{b \times Z} \quad (2)$$

Para la cual se han definido diversos valores de a y b de acuerdo a la región de estudio (ej.: (Hohl R., et al. 2002; Makitov, V., 2007; Foote, G.B., 2005;

⁵ Daño realizado por granizo no simulado. Existen trabajos donde se simula el daño ocasionado por el granizo, pero estudios posteriores cuestionan la exactitud de dicha simulación (Katz R. W. y Garcia,R. R., 1981).

Waldvogel, A. et al. 1978; Schiesser,H. H., 1990), el más utilizado es a=5 y b=0,084 determinado por Waldvogel, A. et al. 1978. Con esto se define una función logística de la variable E con respecto del daño, para cada cultivo y estado fenológico del mismo (Katz R. W. y Garcia,R. R., 1981; Schiesser,H. H., 1990).

En nuestro país se estudió la relación de la probabilidad de granizo generada por el algoritmo ZHAIR de los radares de INTA Paraná e INTA Pergamino con el porcentaje de daño en cultivos y no se encontró una relación clara (Mezher, R.N. y Mercuri, P.A. 2010), generando muchas falsas alarmas (Mezher, R.N. y Mercuri, P.A. 2010; Mezher, R.N., 2011).

En cuanto a las variables polarimétricas de banda C, (Mezher, R.N., 2011) analizó la relación entre el parámetro Hail Differential Reflectivity (HDR), calculado con datos del radar de Paraná y el daño en cultivos, encontrando una buena relación, donde el mayor daño en cultivo coincide con el núcleo de 50 HDR, que puede indicar granizo de tamaño medio (Mezher, R.N., 2011).

Los trabajos que utilizan minería de datos para encontrar patrones entre las variables medidas por los radares y el granizo , en general, presentan cuatro tipos de objetivos: a) clasificar hidrometeoros donde una de las clases es “granizo” solo o combinado (ej.: (López L. y Sánchez J. L., 2009; Billet, J. et al., 1997; Holleman, I., 2001; Marzban C. y Witt A., 2001; Tsagalidis, E. G. et al. 2008), b) clasificar tormentas donde uno de los tipos indica granizo (ej.: Alexiuk,M. et al., 1999, Li, P. C. et al., 2000, Ramirez, L. et al. 2001, Gagne, D. J. et al. 2009, Peters, J. F. et al. 2003, (Li, X. et al. 2004), c) determinar el tamaño del granizo (ej.: (Billet, J. et al., 1997; Holleman, I., 2001; Marzban C. y Witt A., 2001; Tsagalidis, E. G. et al. 2008) y d) determinar el daño producido por el granizo (ej.: Hohl R., et al. 2002; Sánchez J. L., et.al. 2013; Katz R. W. y Garcia,R. R., 1981; Schiesser,H. H., 1990).

Las regresiones, y en particular la regresión logística, aparecen como el método más usado (ej.: Hohl R., et al. 2002; Sánchez J. L., et.al. 2013; Billet, J. et al., 1997; Katz R. W. y Garcia,R. R., 1981; Schiesser,H. H., 1990), especialmente para determinar el daño en cultivos. Otros métodos mencionados en la bibliografía son: a) las redes neuronales (ej.: Alexiuk, M. et al., 1999; Marzban C. y Witt A., 2001; Ramirez, L. et al. 2001), b) los vectores soporte (ej.:

Ramirez, L. *et al.* 2001; Suraj, Z. *et al.* 2002; Li, X. *et al.* 2004), c) los algoritmos genéticos (ej.: Li, P. C. *et al.*, 2000), d) los árboles de decisión (ej.: Gagne, D. J. *et al.* 2009; Tsagalisdis, E. G. *et al.* 2008), e) técnicas de clustering (ej.: Gagne, D. J. *et al.* 2009; Li, X. *et al.* 2004) f) clasificador Bayes (ej.: Tsagalisdis, E. G. *et al.* 2008; Li, X. *et al.* 2004) y g) métodos rough set (ej.: Peters, J. F. *et al.* 2003; Suraj, Z. *et al.* 2002).

Los trabajos analizados utilizan diversas variables de entrada donde se combinan datos del radar con información de otros sensores (ej.: radio sondes, satélites, etc.) (Ceperuelo Mallafré, M., 2007). En diversos trabajos (ej.: Alexiuk, M. *et al.*, 1999; Li, P. C. *et al.*, 2000; Ramirez, L. *et al.* 2001; Peters, J. F. *et al.* 2003; Suraj, Z. *et al.* 2002) se cuenta con un banco de datos de tormentas con variables previamente calculadas de las mismas.

Cuando se trabaja solo con datos de radar, las variables de los modelos son: a) variables que caracterizan la tormenta (ej.: tipo de tormenta, duración, etc.), b) diversas variables derivadas a partir de Z (ej.: VIL, VIL Density, E, SHI, etc.) y c) variables derivadas de las variables polarimétricas (ej.: HDR, Linear Depolarization Radio (LDR), etc.).

REFERENCIAS BIBLIOGRÁFICAS

- Alberoni, P. P., Zrnic, D. S., Ryzhkov, A. V. y Guerrieri, L. “Use of a fuzzy logic classification scheme with a C-band polarimetric radar: first results,” 2002, vol. 324, p. 327.
- Ale, J. M. “Data Mining y Cambio de Creencias: Aplicación de Lógicas Rebatibles al Problema de la Actualización de Reglas.” VI Congreso Internacional en Innovación Tecnológica Informática, CIITI, Capítulo Buenos Aires., 2006.
- Alexiuk, M., Li, P. C. Pizzi, N. y Pedrycz, W. “Classification of Hail and Tornado Storm Cells Using Neural Networks,” in 1999 IEEE Western Canada Conference and Exhibition, pp. 15–21.
- Almada, Natalia, Florencia Díaz, Leticia Osorio, Juan Blatter, Héctor Rodríguez, Juan José De Battista, Norma Arias, y Carlos Bocchio. 2005. «Evaluación del daño por granizo en soja». Boletín Técnico EEA Concepción del Uruguay., Producción Vegetal, 46: 3.
- Anderson, M. E., Carey, L. D., Petersen, W. A. y Knupp, K. R. “C-band Dual-polarimetric Radar Signatures of Hail,” Jul. 2012.
- Aydin, K., Seliga, T. A. y Balaji, V. “Remote sensing of hail with a dual linear polarization radar,” Journal of Climate and Applied Meteorology, vol. 25, no. 10, pp. 1475–1484, 1986.
- Bartok, J., Habala, O., Bednar, P., Gazak, M. y Hluchy, L. “Data mining and integration for predicting significant meteorological phenomena,” Procedia Computer Science, vol. 1, no. 1, pp. 37–46, 2010.
- Bellini Saibene, Y. “Estimación de ocurrencia de granizo en superficie y daño en cultivos utilizando datos de radar meteorológico y técnicas de minería de datos”. Tesis de Maestría. Octubre de 2015.
- Billet, J., DeLisi, M., Smith, B. G. y Gates, C. “Use of Regression Techniques to Predict Hail Size and the Probability of Large Hail,” Weather and Forecasting, vol. 12, no. 1, pp. 154–164, Mar. 1997.
- Boettcher, J. “Dual Polarization RADAR Operation Course,” Dual Polarization RADAR Operation Course. [Online]. Available: <http://www.wdtb.noaa.gov/courses/dualpol/RDA/Lesson1/player.html>. Accedido: 2014.
- Brandes, E. A. y Ryzhkov, A. V. “Hail detection with polarimetric radar,” 2004, vol. 5.
- Bringi V. N. y Chandrasekar, V. “Polarimetric Doppler weather radar: Principles and applications”. Cambridge Univ Pr, 2001.
- Brooks, H. E., Lee, J. W. y Craven, J. P. “The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data,” Atmospheric Research, vol. 67, pp. 73–94, 2003.
- Bustos C. y Videla, H. “Modelo estadístico de predicción de tormentas a corto plazo para la provincia de Mendoza,” en Anales del XI Congreso Argentino de Meteorología. Catuogno, GA, 1982.
- Casagrande, G., Vergara T., and Bellini Saibene Y., “Cartas agroclimáticas actuales de temperaturas, heladas y lluvia de la provincia de La Pampa [Argentina]. Recent agroclimatic maps of temperature, frost and rainfall in La Pampa [Argentina],” Revista de la Facultad de Agronomía, vol. 17, Nov. 2006.
- Ceperuelo Mallafré, M. “Identificación y caracterización del granizo mediante el radar meteorológico. Modelos de predicción del ciclo de vida de las células convectivas.” Tesis Doctoral., 2008.
- Foote, G. B., Krauss, T. W. y Makitov, V. “Hail metrics using conventional radar,” in Proc., 16th Conference on Planned and Inadvertent Weather Modification, 2005.
- Fundamentos de radar meteorológico. USA: University Corporation for Atmospheric Research (UCAR), 20120910.
- Gagne, D. J., McGovern, A. y Brotzge, J. “Classification of convective areas using decision trees,” Journal of Atmospheric and Oceanic Technology, vol. 26, no. 7, pp. 1341–1353, 2009.
- Gutiérrez, J. M., Cano, R., Cofi A. S. y Sordo, C. M. “Redes

- probabilísticas y neuronales aplicadas a las ciencias atmosféricas,” INM, Ministerio de Medio Ambiente, Madrid, 2004.
- Hand, D. J., Mannila, H. y Smyth, P. “Principles of data mining”. The MIT press, 2001.
 - Heinselman, P. L., Ryzhkov, A. V. “Validation of polarimetric hail detection,” 2010.
 - Hohl R., Schiesser, H.H. and Knepper, I. “The use of weather radars to estimate hail damage to automobiles: an exploratory study in Switzerland,” Atmospheric research, vol. 61, no. 3, pp. 215–238, 2002.
 - Holleman, I. Hail detection using single-polarization radar. Ministerie van Verkeer en Waterstaat, Koninklijk Nederlands Meteorologisch Instituut, 2001.
 - _____, Wessels, H. R., Onvlee, J. R. y Barlag, S. J. “Development of a hail-detection-product,” Physics and Chemistry of the Earth Part B Hydrology Oceans and Atmosphere, vol. 25, pp. 1293–1297, 2000.
 - “Infografía: El granizo | EROSKI Consumer,” Eroski Consumer, 27-Jul-2012. [Online]. Available: http://www.consumer.es/web/es/medio_ambiente/naturaleza/2009/06/07/185799.php. [Accessed: 27-Jul-2012].
 - “Instruction Manual. Rainbow 5. Products and Algorithms.,” Germatronik, 2007.
 - Kaltenboeck R. y Ryzhkov, A. “Comparison of polarimetric signatures of hail at S and C bands for different hail sizes,” Atmospheric Research, no. 0, Jul. 2012.
 - Katz R. W. y Garcia, R. R. “Statistical relationships between hailfall and damage to wheat,” Agricultural Meteorology, vol. 24, pp. 29–43, 1981.
 - Li, P. C, Pizzi, N., Pedrycz, W., Westmore, D. y Vivanco, R. “Severe storm cell classification using derived products optimized by genetic algorithms,” in Electrical and Computer Engineering, 2000 Canadian Conference on, 2000, vol. 1, pp. 445–448.
 - Li, X., Ramachandran, R., Rushing, J., Graves, S., Kelleher, K., Lakshmivarahan, S., Kennedy, D. y Levit, J. “Mining next-rad radar data: An investigative study,” in American Meteorology Society annual meeting, 2004.
 - Ligier, Daniel. 2009. «Documento Base del Programa Nacional: Ecorregiones». INTA. Disponible en: <http://inta.gob.ar/documentos/documento-base-del-programa-nacional-ecorregiones/>. Accedido: 2012.
 - López, L., García-Ortega, E. y Sánchez, J. L. “A short-term forecast model for hail,” Atmospheric Research, vol. 83, no. 2–4, pp. 176–184, Feb. 2007.
 - _____ y Sánchez, J. L. “Discriminant methods for radar detection of hail,” Atmospheric Research, vol. 93, no. 1, pp. 358–368, 2009.
 - Maderey Rascón, L. E., Roman, J. Y others, Principios de hidrogeografía. Estudio del ciclo hidrológico. Unam, 2005.
 - Marzano, F. S., Scaranari, D. y Vulpiani, G. “Supervised Fuzzy-Logic Classification of Hydrometeors Using C-Band Weather Radars,” IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 11, pp. 3784–3799, Nov. 2007.
 - _____, Scaranari, D., Montopoli, M. y Vulpiani, G. “Supervised classification and estimation of hydrometeors from C-band dual-polarized radars: A Bayesian approach,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 46, no. 1, pp. 85–98, 2008.
 - Marzban C. y Witt A. “A Bayesian neural network for severe-hail size prediction,” Weather and Forecasting, vol. 16, no. 5, pp. 600–610, 2001.
 - Makitov, V. “Radar measurements of integral parameters of hailstorms used on hail suppression projects,” Atmospheric Research, vol. 83, no. 2–4, pp. 380–388, Feb. 2007.
 - Mezher R. N. and Mercuri, P. A. “Análisis espacial y temporal de la ocurrencia de eventos de granizo sobre Argentina,” XV Congresso Brasileiro de Meterologia, agosto de 2008.
 - Mezher, R. N.; Doyle M. y Barros, V. “Climatology of hail in Argentina,” Atmospheric Research, vol. 114–115, pp. 70–82, Oct. 2012.
 - Mezher R. N., Mercuri P. A., and Gattinoni N. N., “Distribución espacio-temporal del granizo en Argentina,” Reunión Argentina de Agrometeorología. 12. 2008 10 o8-10, 8 al 10 de octubre de 2008. San Salvador de Jujuy. AR., 2008.
 - Mezher, R. N. Gattinoni, N. N. and Mercuri, P. A. “Variabilidad estacional de la ocurrencia de granizo en el centro, este y noreste de Argentina,” Reunión Argentina de Agrometeorología. 12. 2008 10 o8-10, 8 al 10 de octubre de 2008. San Salvador de Jujuy. AR., 2008.
 - Mezher, R. N., Banchero, S. y Bellini Saibene, Y. N. “Identificación de granizo con la utilización de variables polarimétricas de los radares de Paraná y Anguil, el radar de Pergamino y daño en cultivos” en Congreso Argentino de Meteorología. 11. 2012 05-06 28-01, 28 de mayo al 1 de junio de 2012. Mendoza. AR., 2012.
 - Mezher R. N. y Mercuri, P. A. “Uso de la red de radares de INTA para la detección de granizo,” XIII Reunión Argentina y VI Latinoamericana de Agrometeorología., Oct. 2010.
 - Mezher, R. N., Vidal, L. y Salio, P. “Hailstorms Analysis using Polarimetric Weather Radars and Microwave Sensors in Argentina,” 6th European Conference on Severe Storms (ECSS 2011), 26082011.
 - Nisbet, R., Elder, J., Elder, J. F. y Miner, G. “Handbook of statistical analysis and data mining applications”. Academic Press, 2009.
 - Olson D. L. y Delen, D. “Advanced data mining techni-

ques". Springer Verlag, 2008.

- Pérez, R. C. "Caracterización de las celdas graniceras que producen daño severo en la provincia de Mendoza utilizando mediciones del radar meteorológico de Banda C," 2006.
- Peters, J. F., Suraj, Z., Shan, S., Ramanna, S., Pedrycz, W. y Pizzi, N. "Classification of meteorological volumetric radar data using rough set methods," *Pattern Recognition Letters*, vol. 24, no. 6, pp. 911–920, Mar. 2003.
- Ponce de Leon, E. "Granizo." Servicio Meteorológico Nacional, 1985.
- Ramirez, L., Pedrycz, W. y Pizzi, N. "Severe storm cell classification using support vector machines and radial basis function approaches," in *Electrical and Computer Engineering*, 2001. Canadian Conference on, 2001, vol. 1, pp. 87–91.
- "Resultados de la 'Encuesta Censal sobre los Seguros en el Sector Agropecuario y Forestal,'" Superintendencia de Seguros de la Nación, 3545, Apr. 2013.
- Rinehart, R. E. "Radar for meteorologists". Dept. of Atmospheric Sciences, Center for Aerospace Sciences, University of North Dakota, 1997.
- Ryzhkov, A., Zrnic, D., Krause, J., Kumjian, M. y Ganson, S. "Small Hail Dry Hail? Wet Hail? Giant Hail," Jul. 2012.
- _____, Schuur, T. J., Burgess, D. W., Heinselman, P. L., Giangrande, S. E. y Zrnic, D. S. "The joint polarization experiment. polarimetric Rainfall Measurement and Hidrometeor Classification," *Bull. Amer. Meteor. Soc.*, vol. 86, pp. 809–824, 2005.
- Sánchez, J. L., Fraile, R., De La Madrid, J. L., De La Fuente, M. T., Rodríguez, P. y Castro, A. "Crop damage: The hail size factor," *Journal of Applied Meteorology*, vol. 35, no. 9, pp. 1535–1541, 1996.
- _____, López, L., García-Ortega, E. y Gil, B. "Nowcasting of kinetic energy of hail precipitation using radar," *Atmospheric Research*, vol. 123, pp. 48–60, 2013.
- Schiesser, H. H. "Hailfall: the relationship between radar measurements and crop damage," *Atmospheric Research*, vol. 25, no. 6, pp. 559–582, 1990.
- Schuur, T. "Radar: Frequently Asked Questions - Summary." National Severe Storms Laboratory. Polarimetric Radar Research., 17-Feb-2003.
- Simonelli, S. C. y Nicolini, M. "Modelo estadístico de Pronóstico de Convección para la zona norte de la Provincia de Mendoza," 2000.
- Simeonov, P. "An Overview of Crop Hail Damage and Evaluation of Hail Suppression Efficiency in Bulgaria," *J. Appl. Meteor.*, vol. 35, no. 9, pp. 1574–1581, Sep. 1996.
- Skripnikova, K., Rezacova, D. y Skripnikova, K. "Testing radar-based hail detection criteria." ERAD 2012. The seventh european conference on Radar in meteorology and hydrology.2012.
- Suraj, Z., Peters, J. F. y Rzasa, W. "A Comparison of Different Decision Algorithms Used in Volumetric Storm Cells Classification," *Fundamenta Informaticae*, vol. 51, no. 1, pp. 201–214, Enero 2002.
- Tabary, P., Fradon, B., Illingworth, A. J. y Vulpiani, G. "Hail detection and quantification with a C-band polarimetric radar: Challenges and promises," in 34th Conference on Radar Meteorology, 2009, pp. 5–9.
- "Thunderstorm: hail production -- Britannica Online Encyclopedia," Encyclopedia Britannica, 27-Jul-2012. [Online]. Available: <http://www.britannica.com/EBchecked/media/161933/Hail-producing-thunderstorm-in-cross-section>. [Accessed: 27-Jul-2012].
- Tsagalidis, E. G.; Tsitouridis, K. G.; Evangelidis, G. y Dervos, D. A. "Hail Size Estimation and Prediction using Data Mining Techniques.". Proceedings 5th European Conference on Radar in Meteorology and Hydrology. 2008.
- Waldvogel, A., Schmid, W. y Federer, B. "The Kinetic Energy of Hailfalls. Part I: Hailstone Spectra," *J. Appl. Meteor.*, vol. 17, no. 4, pp. 515–520, Abril 1978.
- _____, Federer, B. y Grimm, P. "Criteria for the detection of hail cells," *Journal of Applied Meteorology*, vol. 18, no. 12, pp. 1521–1525, 1979.
- Wikipedia contributors. 2012a. «Meteoro (meteorología)». Wikimedia Foundation, Inc. [http://es.wikipedia.org/w/index.php?title=Meteoro_\(meteorolog%C3%ADA\)&oldid=56864014](http://es.wikipedia.org/w/index.php?title=Meteoro_(meteorolog%C3%ADA)&oldid=56864014).
- _____, 2012b. «Metaestabilidad». Wikimedia Foundation, Inc. <http://es.wikipedia.org/w/index.php?title=Metaestabilidad&oldid=57130092>.
- Witten, I. H., Frank, E. y Hall, M. A. "Data Mining: Practical machine learning tools and techniques". Morgan Kaufmann, 2011.

Identificación de granizo y daño en cultivos con la utilización de variables polarimétricas de los radares de Paraná y Anguil

5.

Mezher, Romina²; Banchero, Santiago²; Bellini Saibene, Yanina¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA

² Instituto de Clima y Agua del INTA

mezher.romina@inta.gob.ar

RESUMEN

El objetivo de este estudio es identificar y caracterizar las zonas con daños ocasionados por granizo en campos cultivados de soja, maíz y trigo. Para ello se analizaron datos de intensidad de daño en cultivos sobre la región Pampeana desde febrero de 2009 hasta mayo de 2011 y la relación con las variables polarimétricas reflectividad horizontal (Z_h), reflectividad diferencial (Z_{dr}), desplazamiento de fase diferencial (Φ_{DP}), desplazamiento de fase diferencial específico (KDP) y coeficiente de correlación polarimétrico ($RhoHV$) obtenidas por medio de los radares meteorológicos de banda-C INTA-Paraná e INTA-Anguil. Los datos de los radares fueron almacenados en bases de datos y se utilizaron las elevaciones inferiores para describir las regiones con mayor daño y cortes verticales para analizar las zonas donde se observan caídas de hidrometeoros. Se encontraron valores característicos para las regiones afectadas por granizo para radares de banda C, elevados valores de Z_h , valores altos de Z_{dr} , valores bajos y negativos de KDP y una abrupta caída de $RhoHV$ coincidiendo con los trabajos producidos para radares de este tipo. Este trabajo es una primera aproximación a una clasificación de hidrometeoros en Argentina y a un mejor entendimiento del uso de radares meteorológicos para aplicaciones agropecuarias.

Palabras clave: radar meteorológico, doble polarización, riesgo agropecuario

ABSTRACT

The aim of this study is to identify and characterize areas with hail damage in cultivated fields of soybeans, corn and wheat. We analyze crop damage intensity data in Pampa's region from February 2009 until May 2011 and the relationship with polarimetric variables such us horizontal reflectivity (Z_h), differential reflectivity (Z_{dr}), differential phase shift (Φ_{DP}), specific differential phase shift (KDP) and polarimetric correlation coefficient ($RhoHV$) obtained by C-Band weather radars INTA-Paraná and INTA-Anguil. The radar data were stored in a database environment and we used the lower elevations to describe regions with the greatest damages and vertical cross sections to analyze areas of falling hydrometeors. We found typical values for C-band radars for regions affected by hail, high values of Z_h , high values of ZDR , low and negative values of KDP and a sharp drop of $RhoHV$ in agreement with previous works of C-Band weather radars. This work is a first approach to a hydrometeors classification in Argentina and a better understanding of the use of weather radar for agricultural applications.

Keywords: hail, weather radar, crop damage, dual polarization

INTRODUCCIÓN

Los radares polarimétricos tienen grandes beneficios ya que con ellos se puede mejorar la medición de precipitación, se puede discriminar entre los distintos tipos de hidrometeoros y hasta se pueden

Tabla 1. Listado de fechas y eventos analizados.

Fecha	Radar	Cantidad de eventos y horas (UTC)
24 de Octubre de 2009	Paraná	1 (00:20)
29 de Noviembre de 2009	Paraná	2 (14:30 y 16:10 a 16:50)
28 de Diciembre de 2009	Paraná	2 (11:20 y 14:00)
20 de Abril de 2010	Paraná	1 (12:10)
15 de Enero de 2011	Anguil	2 (20:30 y 21:30 a 21:40)

identificar ecos no meteorológicos (Zrnic y Ryzhkov 1999). Numerosos trabajos se enfocan en la determinación de hidrometeoros a partir de radares de banda S (con longitud de onda aproximada de 10 cm), en particular sobre Estados Unidos (Balakrishnan y Zrnic 1990). En los últimos años se comenzó a investigar la problemática en los radares de banda C, en longitudes de onda de los 5.6 cm aproximadamente, desarrollando clasificaciones (Keenan 2003) y analizando atenuaciones y sensibilidades en lluvias y granizo (Gu y otros 2011).

El objetivo de este trabajo es investigar mediante el uso de los radares de INTA Paraná e INTA Anguil la relación entre variables polarimétricas en casos de eventos de granizo que produjeron daño en campos con cultivos de soja, trigo y maíz.

METODOLOGÍA

Se utilizaron datos de las campañas agrícolas 2009-2010 y 2010-2011 de daño en campo de cultivos de soja, trigo y maíz. Se buscaron fechas que

coincidieran con áreas de los radares y que tuvieran al menos 50 puntos con daño. Se tuvo en cuenta la distancia desde el centro del radar hasta cada tormenta ya que se querían analizar los niveles más bajos de las mismas, quedando el nivel más alto en 3 km (tabla 1).

Los datos de radar usados fueron Z (Reflectividad), Zdr (Reflectividad diferencial), Phidp (Diferencial de cambio de fase), KDP (diferencial de cambio de fase específica) y RhoHV (Coeficiente de correlación) con una resolución espacial de 0.5 km y una resolución temporal de 10 minutos.

Variables del Radar

Reflectividad Diferencial (Zdr): se define como el cociente entre la potencia reflejada entre la polarización horizontal y la vertical (ecuación 1). Relaciona la proporción entre el eje vertical y horizontal de los hidrometeoros.

$$Zdr (dB) = 10 \log \left(\frac{Zh}{Zv} \right) \quad (1)$$

Diferencial de cambio de fase (Φ_{DP} - Φ_{IDP}) y diferencial de cambio de fase específica (KDP): cuando la señal del radar pasa por diferentes medios, se retrasa una con respecto de la otra, dependiendo del blanco. Φ_{DP} se debe a la superposición de las fases entre la onda dispersada hacia adelante por el blanco (P en la figura 1) y de la onda transmitida por el radar, las cuales tienen diferentes ángulos de fase (ecuación 2). KDP es Φ_{DP} pero acotada a un rango (ecuación 3).

$$\Phi_{DP} = \Phi_H - \Phi_V \quad (2)$$

$$KDP = \frac{1}{2} \frac{\theta dp(r2) - \theta dp(r1)}{r2 - r1} \quad (3)$$

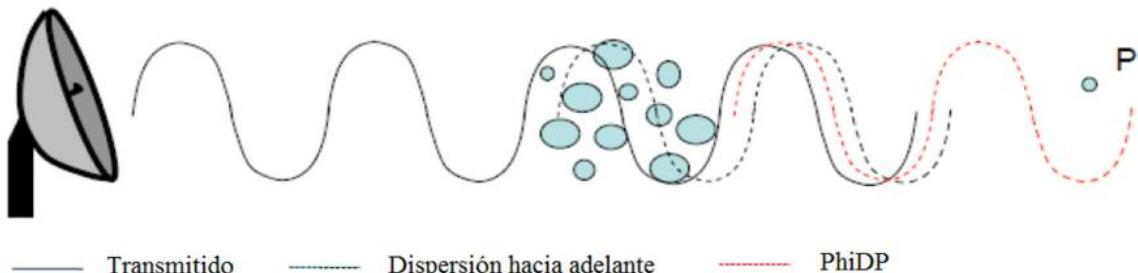


Figura 1. Esquema del comportamiento de las señales del radar cuando atraviesan diferentes medios y cálculo de Φ_{DP} .

Coeficiente de correlación copolar (ρ_{HV} - Rho_{HV}): se define como el cociente entre la potencia reflejada entre la polarización horizontal y la vertical. Relaciona la proporción entre el eje vertical y horizontal de los hidrometeoros.

RESULTADOS

Las figuras 2 y 3 presentan los mapas de Indicadores de Posición de Plano (Plan Position Indicator-PPI) de las variables Z, KDP, RhoHV y ZDR del RADAR de Anguil correspondientes al 15 de enero de 2011. Los puntos de campo están marcados como círculos negros.

La figura 4 muestra los cortes verticales de la tormenta del 15 de Enero de 2011 de las 21:30 hs a 21:40 hs registrada por el RADAR de Anguil. Las líneas blancas marcan la altura en kilómetros: la prime-

ra línea corresponde a 2 km de altura y se marcan cada 2 kilómetros hasta llegar a 8 km de altura en la última línea marcada. El corte se realizó en el espacio marcado con línea negra gruesa en la variable Z de la imagen 3 y la escala de las variables también es la misma que la utilizada en las imágenes 2 y 3.

En estos cortes verticales se aprecia el gran desarrollo vertical de la tormenta (por encima de los 10 km de altura) y los núcleos de Z con valores mayores a 70 dBZ (color blanco en la figura 3 y 4), mayores a 60 dBZ (color rojo oscuro en la figura 3 y 4).

Las figuras 5 y 6 presentan los mapas de PPI de las variables Z, KDP, RhoHV y ZDR del RADAR de Paraná para las tormentas del 24 de Octubre de 2009 y la del 28 de Diciembre de 2009 a las 14:00 hs. Nuevamente, los puntos de campo están marcados como círculos negros.

Finalmente la figura 7 y 8 presenta las relaciones entre las variables polarimétricas por medio de grá-

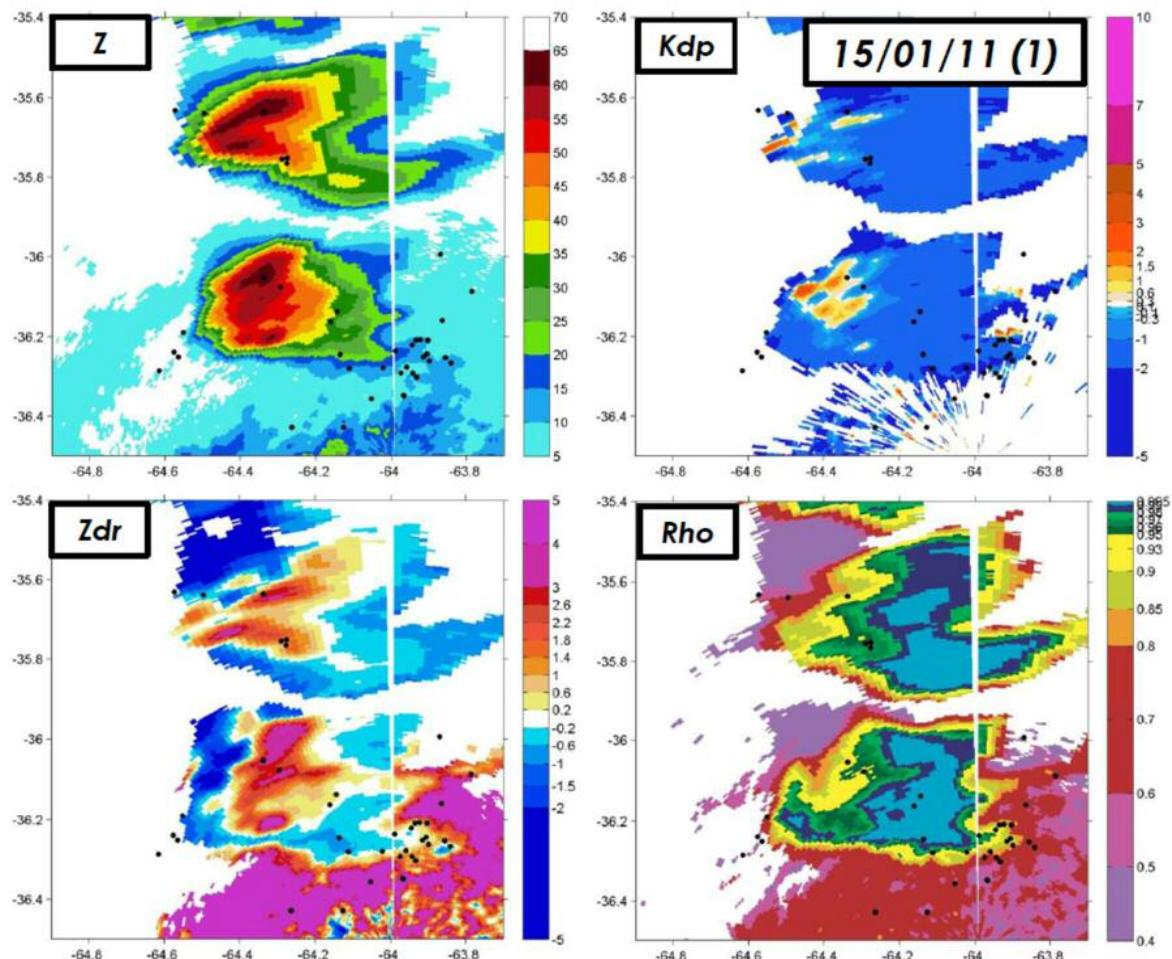
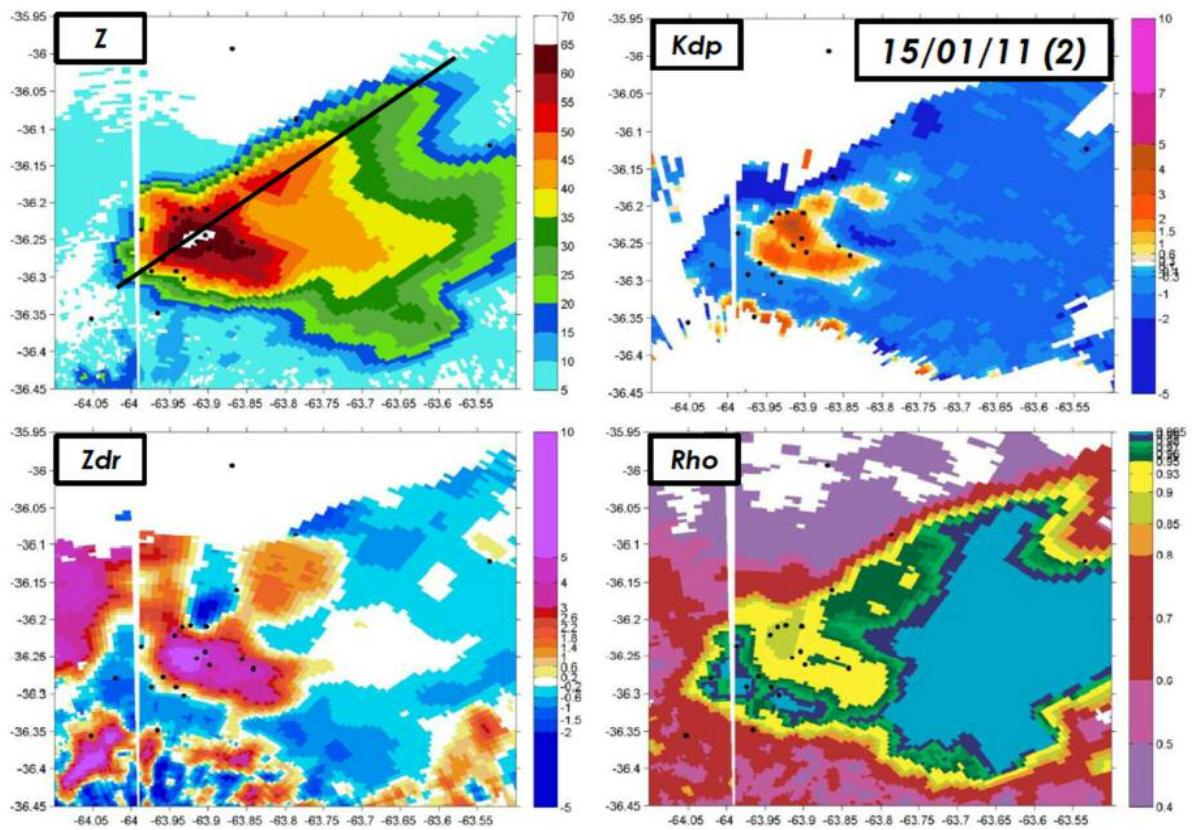
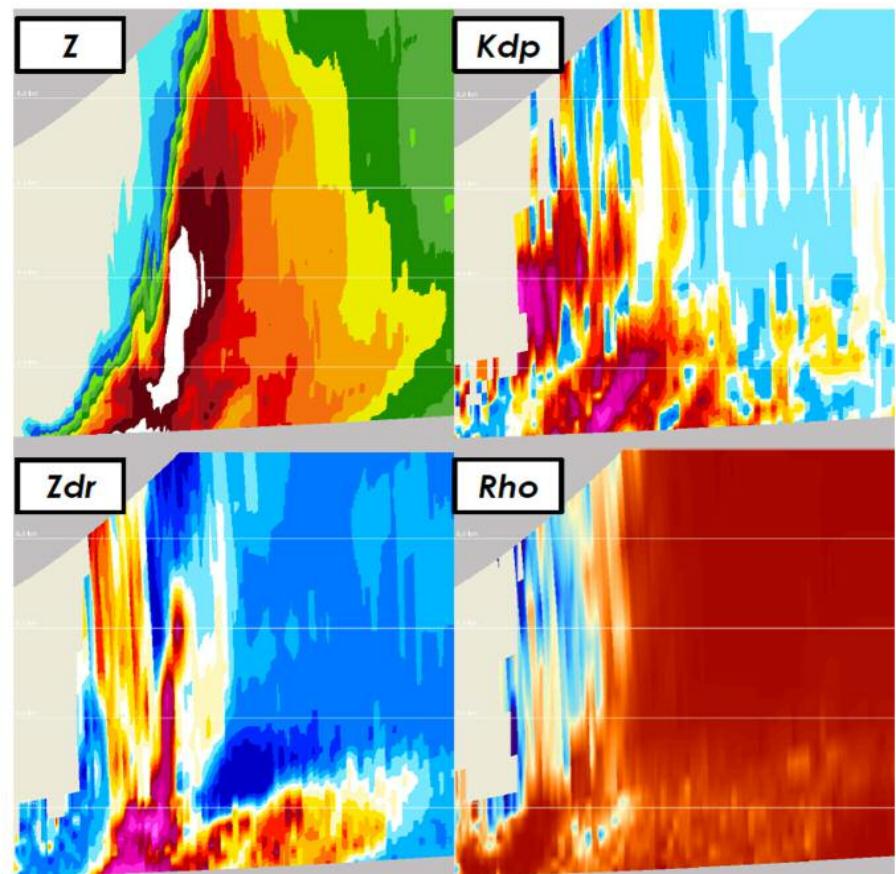


Figura 2. PPI de las variables Z, KDP, ZDR y RhoHV registradas por el RADAR de Anguil de la tormenta del 15 de Enero de 2011 a las 20:30hs.



(ARRIBA) Figura 3. PPI de las variables Z, KDP, ZDR y RhoHV registradas por el RADAR de Anguil de la tormenta del 15 de Enero de 2011 de las 21:30hs a 21:40hs.



(DERECHA) Figura 4. Corte vertical de la tormenta del 15 de Enero de 2011 de las 21:30hs a 21:40hs registrada por el RADAR de Anguil para las variables Z, ZDR, KDP y RhoHV.

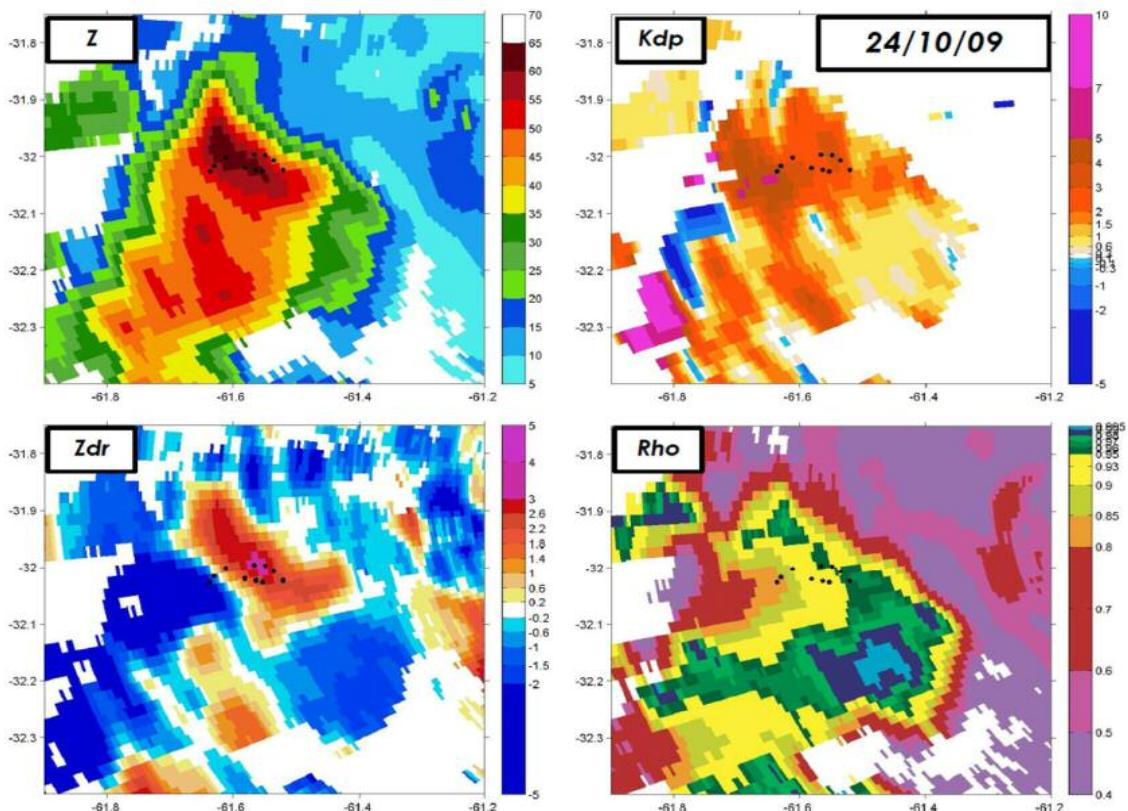


Figura 5. PPI de las variables Z, KDP, ZDR y RhoHV registradas por el radar de Paraná de la tormenta del 24 de Octubre de 2009

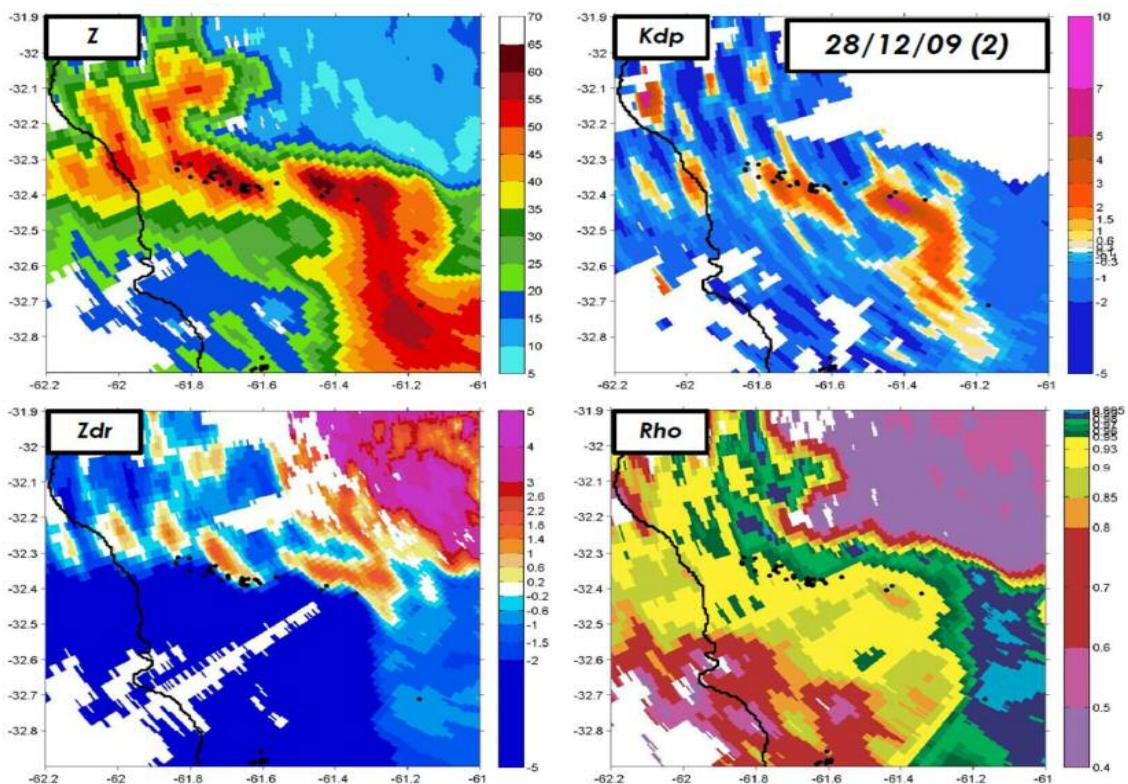


Figura 6. PPI de las variables Z, KDP, ZDR y RhoHV registradas por el radar de Paraná de la tormenta del 28 de Diciembre de 2009 a las 14:00hs.

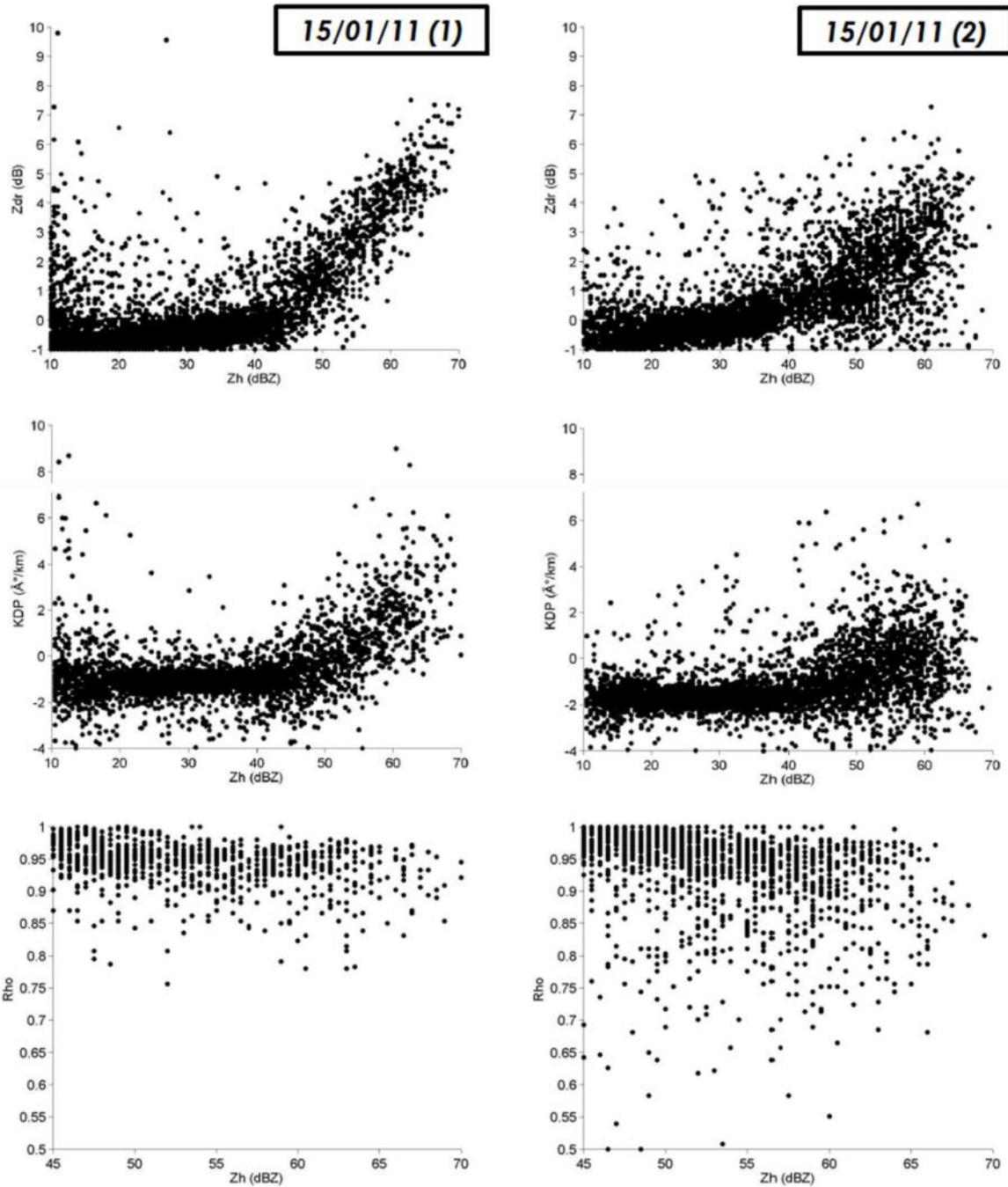


Figura 7. Relaciones entre variables polarimétricas del RADAR de Anguil para las tormentas del 15 de Enero de 2011.

ficos de dispersión para las dos tormentas registradas por el radar de Anguil y las tormentas del 24/10/2009 y 28/12/2009 registradas por el radar de Paraná.

Cuando se analizan los valores de Z que superan los 60 dBZ se observan numerosos eventos con $Z\text{dr} > 4 \text{ dB}$.

Valores máximos de Zdr pueden estar asociados con la presencia de fases líquida y sólida, en particular con granizos pequeños y grandes gotas (Picca y Ryzhkov 2012)

Con $\text{RhoHV} < 0.75$ y valores grandes de Z es posible que se estén observando granizos de tamaño mediado.

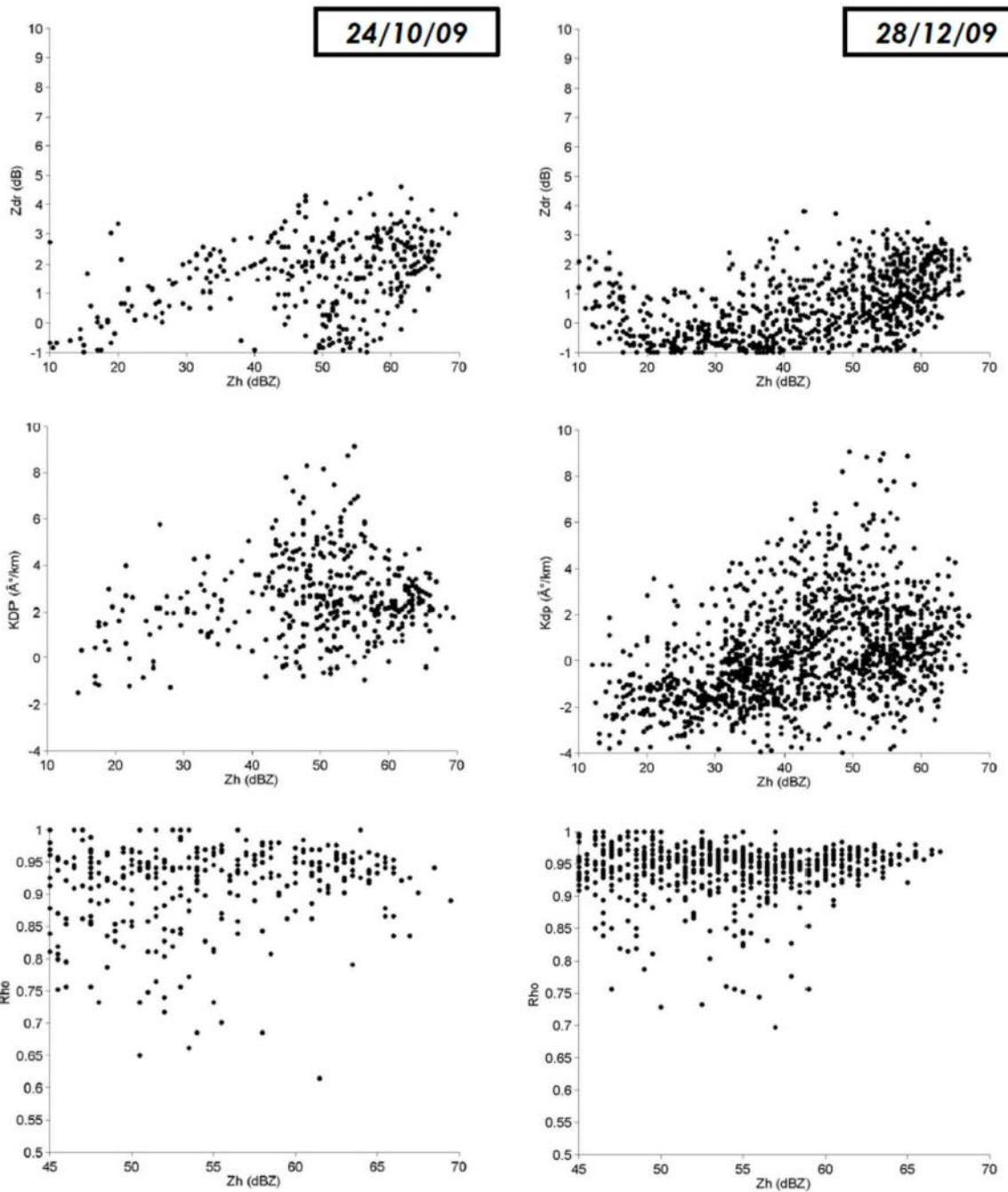


Figura 8. Relaciones entre variables polarimétricas registradas por el RADAR de Paraná para las tormentas del 24 de Octubre de 2009 y 28 de Diciembre de 2009.

CONCLUSIONES

Se analizaron 8 eventos de daño por granizo con los datos registrados por los radares meteorológicos ubicados en las localidades de Paraná y Anguil.

Se encontró una buena relación entre las varia-

bles Z y Zdr. Tanto el Kdp, PhiDP y RhoHV muestran campos muy variables y con “ruido” lo cual dificulta la correcta interpretación de los datos.

Para el radar de Paraná, el día 20 de abril de 2010 se observaron valores bajos de Zdr (cercaos a 0) en zonas con elevado Z coincidiendo con Ryzhkov y

otros (2009) que sugiere que estos datos se registran cuando el granizo es numeroso y aparece en el Zdr como una esfera casi perfecta.

El radar de Anguil mostró en el día 15 de enero de 2011 una buena coherencia espacial con los daños en campo. Los altos valores de Zdr coincidiendo con Z elevados y con Kdp entre 2 y 3 °/km demuestran una zona con gotas grandes mezcladas con granizo.

Las regiones donde RhoHV alcanza valores menores a los 0.90 muestra una coherencia con lo que Balakrishnan and Zrnic (1990) encontraron.

El efecto de resonancia en la banda C provoca que aumenten los valores de Zdr, aumente el Kdp y disminuya el coeficiente de correlación en la zona de fusión de granizo.

La problemática de este efecto fue y sigue siendo estudiada en los últimos años y junto a la atenuación producida en los radares de banda C son los grandes desafíos de los próximos estudios en los radares de Argentina.

Agradecimientos

Los autores quieren agradecer a La Dulce Cooperativa de Seguros y Sancor Seguros por los datos de daño en campo provistos. Este trabajo fue financiado por el proyecto de área estratégica AERN 2944 GEOINTA 2.0

REFERENCIAS BIBLIOGRÁFICAS

- Balakrishnan, N. and D. Zrnic, 1990. Use of Polarization to Characterize Precipitation and Disrdiminate Large Hail. JAS, vol. 47, 13, 1525-1540.
- Gu J., A. Et al, 2011. Polarimetric Attenuation Correction in Heavy Rain at C Band. J. Appl. Met. Clim. 50. 39-58.
- Keenan T., 2003. Hydrometeor classification with a C-Band Polarimetric Radar. Aust. Met. Mag. 52. 23-31
- Picca J. and A. Ryzhkov, 2012. A Dual Wavelengh Polarimetric Analysis of the 16 May of 2010 Oklahoma City Extreme Hailstorm. Month. Wea. Rev. 140. 1385-1403.
- Ryzhkov, A. S. et al, 2009: Polarimetric characteristics of melting hail at S and C bands. 34th Conference on Radar Meteorology, Williamsburg, VA, Amer. Meteor. Soc., 4A.6
- Zrnic, D. and A. Ryzhkov, 1999. Polarimetry for Weather Surveillance Radars. BAMS, 80, 3

Estimación de ocurrencia de granizo en superficie mediante datos de radar meteorológico utilizando técnicas de Data Mining

6.

Bellini Saibene, Yanina¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
bellini.yanina@inta.gob.ar

RESUMEN

El granizo es capaz de infligir cuantiosos daños y el estudio de su frecuencia e impacto económico es de interés para la industria de los seguros y el sector agroindustrial. Con el objetivo de estimar la ocurrencia de granizo en superficie se utilizó Gene Expression Programming Regresión Logística usando datos del radar polarimétrico de banda C (INTA Anguil, La Pampa) desde Marzo de 2009 a Marzo de 2013. La ocurrencia de granizo en superficie se modeló como un problema binario, utilizando solo variables derivadas del radar. A pesar de la simplicidad del modelo obtenido, una comparación entre medidas de performance (Probabilidad de Detección (POD), Falsas Alarmas (FAR) y Porcentaje Correcto (PC)) entre 30 modelos internacionales publicados, que usan diferentes técnicas y set de datos, ubicó a nuestro modelo entre los tres primeros. Confirmando que los modelos con variables polarimétricas funcionan mejor que aquellos que usan variables de simple polarización y que la presencia de granizo aumenta con mayores valores de Reflectividad (Z), menores valores del Coeficiente de correlación co-polar (RhoHV) y valores extremos de Reflectividad Diferencial (ZDR). La implementación del modelo estableció problemas con el uso de técnicas tradicionales de manejo de datos, alentando enfoques adicionales como bases de datos no estructuradas y técnicas de procesamiento paralelos para la operación de los mismos.

Palabras clave: Gene Expression Programming, Riesgos agrícolas, doble polarización, banda C, tormentas severas

ABSTRACT

Hail is capable of inflicting considerable damage and the study of their frequency and financial impact is useful for the insurance industry and agribusiness. In order to calculate the probability of hail on the ground Gene Expression Programming Logistic Regression was used with data from polarimetric C-band radar (INTA Anguil La Pampa) from March 2009 to March 2013. The hailfall was modeled as a binary problem, using only variables derived from radar. Despite the simplicity of the model obtained, a comparison between the measurements of performance Probability of Detection (POD), False Alarm Ratio (FAR) and Percent Correct (PC) among 30 reported international models, using different techniques and data sets, showed our analysis within the three models with higher fitting. Confirming that models with polarimetric variables showed higher performance than those with single polarization and the presence of hail increase with higher of Reflectivity (Z) values, lower Correlation Coefficient (RhoHV) values and extreme values of Differential Reflectivity (ZDR). The implementation of the model using traditional data management techniques established problems, encouraging additional approaches like unstructured databases and parallel processing techniques for radar data processing and model operation.

Keywords: Gene Expression Programming, Agricultural risks, dual polarization, C band, severe storms

INTRODUCCIÓN

El granizo se considera un riesgo agropecuario y los daños que provoca este fenómeno meteorológico extremo causan graves pérdidas en algunas regiones argentinas y afecta a diferentes sectores económicos, tanto en las zonas urbanas como rurales (Ponce de Leon, 1985). En la región Pampeana Argentina, la máxima frecuencia de ocurrencia de granizo se encuentra en Córdoba y La Pampa (Mezher, et. al. 2008, Mezher, et. al. 2012, Mezher, et. al. 2008b, Mezher, et. al. 2010).

El estudio del granizo, su frecuencia e impacto son de utilidad para la industria de los seguros (estimación de daño, mitigación y transferencia, riesgo forense) y la comunidad agropecuaria (protección, mitigación y transferencia del riesgo, caracterización de riesgos agropecuarios, riesgo forense, lucha anti-granizo) (Hohl et al. 2002, Casellas et al. 2012), entre otros ámbitos. La precipitación de granizo se caracteriza por tener una alta variabilidad espacial y temporal lo que representa un gran desafío para su detección de ocurrencia en superficie, convirtiendo ésta última en una tarea costosa debido al instrumental de campo necesario (Bustos et al. 1982, Sánchez et al. 2013, Tabary et al. 2009, López et al. 2009, López et al. 2007). En este contexto, los sensores remotos como los radares meteorológicos son una alternativa a las redes terrestres de mediciones, con la ventaja de abarcar una gran superficie y disponer de una resolución unificada en tiempo y espacio (R. Hohl et al. 2002).

Los radares generan un importante volumen de información diario (Bellini Saibene, Y. 2015), lo que hace necesario el uso de nuevas técnicas para su explotación, como la Minería de datos (Data Mining-DM), que está emergiendo como un método adecuado para la extracción de patrones a partir de conjuntos amplios de datos heterogéneos relacionados a la predicción de fenómenos meteorológicos (Bartok et al. 2010).

Existen numerosos estudios que exploran la relación de las variables medidas por los radares con el granizo generando diversos modelos para estimar su ocurrencia en superficie, algunos de los cuales utilizan técnicas de DM (ej: Alexiuk, et. al. 1999, Holleman, 2001, Gagne, et. al. 2009, Marzano, et. al. 2007). Estos modelos deben ser ajustados a la región de estudio y a los instrumentos de medición utilizados (Bustos et al. 1982; Marzano, et. al. 2007).

Los estudios que usan radares de banda C en Argentina, analizan variables de simple polarización en Mendoza (ej.: Bustos et al. 1982; . Foote et. al. 2005) y variables de doble polarización en las provincias de Buenos Aires, Entre Ríos y La Pampa (ej.: Mezher, et. al. 2012b, Mezher, et. al. 2010, Mezher, et. al. 2011). Las técnicas de DM utilizadas son Regresión Logística (RL), RL con Gene Expression Programming (GEP), Maquinas de Vectores Soporte (SVM), Árboles de Decisión (DT), y Naive Bayes (Bellini Saibene, Y. et. al. 2015). En la mayoría de los trabajos realizados con datos de los radares de INTA se analiza solo la primera elevación de datos (ej: Bellini Saibene, Y. et. al, 2015, Mezher, et. al. 2012b, Mezher, et. al. 2011).

El objetivo de este trabajo es el desarrollo de un modelo de clasificación de ocurrencia de granizo en superficie ajustado a la región que comprende el este de La Pampa, sur de Córdoba y oeste de Buenos Aires, por medio de técnicas de DM y datos generados por el radar meteorológico de banda C de la Estación Experimental Agropecuaria (EEA) Anguil “Ing. Agr. Guillermo Covas” del INTA en La Pampa.

METODOLOGÍA

El método utilizado en este trabajo es el Cross-Industry Standard Process for Data Mining (CRISP-DM).

Técnica de DM

Se utilizaron técnicas de aprendizaje supervisado: el conjunto de datos tiene una etiqueta (target) que indica la precipitación de granizo posterior a una tormenta, se trató de forma binaria, dándole valor cero cuando no precipitó granizo y valor uno cuando si cayó granizo. Se modeló utilizando GEP, seleccionando los métodos de regresión logística y clasificación con las funciones de rendimiento Máxima Verosimilitud y ROC Measure respectivamente. Estas técnicas ya mostraron buenos rendimientos clasificando granizo con datos de RM (Bellini Saibene, Y. et. al. 2015, Bellini Saibene, Y. et. al, 2015b, Bellini Saibene, Y. et. al. 2014). A continuación se introduce brevemente los conceptos del método GEP y de Regresión Logística.

Método GEP

Es un algoritmo evolutivo que genera automáti-

a) Expresión algebraica: $\sqrt{(a+b)\times(c-d)}$

b) Cromosoma: 01234567
Q*+-abcd

c) Árbol de Expresión:

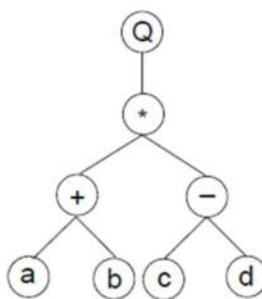


Figura 1. Ejemplo de cromosoma de una función algebraica y su representación por medio de un árbol de expresión utilizado en GEP (Adaptado de: Ferreira C. 2001)

icamente programas de computadora para modelar las relaciones entre las variables analizadas (Ferreira, C. 2001). Se puede utilizar para construir diferentes tipos de modelos como regresiones logísticas, arboles de decisión y redes neuronales (Ferreira, C 2001; Ferreira, C 2006a). Como todo algoritmo evolutivo se rige por el principio de la selección natural: los individuos (en este caso modelos o soluciones) mejor adaptados, tienen mayores posibilidades de sobrevivir y reproducirse en las próximas generaciones.

Tiene dos componentes principales: los cromosomas (genotipo) que son cadenas de caracteres de longitud fija para representar a los programas y los árboles de expresión (fenotipo) de diferentes formas y tamaños, por medio de los cuales se expresan los cromosomas (figura 1). Los cromosomas son los que están sujetos a variaciones genéticas utilizando uno o más operadores genéticos, permitiendo la evolución (Ferreira C. 2001).

El algoritmo GEP inicia con una población aleatoria de cromosomas que representan posibles soluciones al problema en estudio.

Estos cromosomas se asignan a los árboles de expresión, la aptitud de cada individuo se evalúa en base a una función predefinida y los mejores indivi-

duos se seleccionan para su reproducción y modificación genética a través de operadores genéticos de recombinación, mutación y reproducción. Esta nueva generación se somete al mismo proceso hasta que se encuentra una solución o se llega a un número definido de generaciones. El individuo más apto, de acuerdo a una o varias medidas de rendimiento, es la solución final (Ferreira C. 2001). Para más detalles sobre este algoritmo ver Ferreira C. 2006b.

GEP cuenta con una implementación en la herramienta GeneXProTools¹, utilizada en este trabajo.

Este software tiene plataformas para trabajar con clasificación y regresión logística que cuentan con plantillas, con una configuración previa para el modelado, de acuerdo a la cantidad de variables independientes presentes en los datos; en ambos casos los modelos se generan de forma evolutiva (Ferreira C. 2013a; Ferreira C. 2013b), armando automáticamente, nuevas variables a partir de las originales y relaciones entre las mismas, por medio de constantes, cálculos y funciones.

Los algoritmos de reproducción, cruce y selección están debidamente explicados en Ferreira C. 2006b y este trabajo se concentró en la aplicación práctica del mismo y los resultados obtenidos, asumiendo como correcta su implementación en el software utilizado.

Regresión Logística

Se utiliza para modelar problemas de clasificación cuya variable objetivo toma la forma $y \in \{0,1\}$, donde el cero corresponde a la clase negativa y el uno a la clase positiva. En el modelo de regresión logística se trata de calcular la probabilidad en la que una de las opciones de la variable ocurra a partir de los valores que tomen una serie de variables independientes x . La hipótesis se representa por (7) (Hosmer y Lemeshow 2000).

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n \quad (7)$$

Para que la función cumpla con la restricción de $0 \leq h_{\theta}(x) \leq 1$ se utiliza la función sigmoidal, por lo que la función de hipótesis queda como (8) (Hosmer y Lemeshow 2000).

$$\frac{1}{1 + e^{-\theta_0 + \theta_1 x_1 + \dots + \theta_n x_n}} \quad (8)$$

¹ <http://www.gepsoft.com>.

Para obtener los valores de los coeficientes (θ) de los predictores que mejor los relacionan a la variable objetivo, la regresión logística utiliza la estimación de la máxima verosimilitud. La función de costo a minimizar queda dada por (9) (Hosmer y Lemeshow 2000).

$$(9) J(\theta) = -\frac{1}{n} \left[\sum_{i=1}^n (y_i \log h_\theta(x_i) + (1 - y_i) \log(1 - h_\theta(x_i))) \right]$$

Una vez obtenidos los valores de θ , el resultado que arroja la función $h_\theta(x)$ determina la probabilidad que tiene el caso estudiado de pertenecer a la clase positiva, en general se predice $y=1$ si > 0.5 e $y=0$ si < 0.5 .

Datos

Como GEP es una técnica de aprendizaje supervisado necesita un conjunto de datos previamente etiquetado para aprender y generar el modelo que posteriormente permita identificar la presencia de granizo en una nueva tormenta (Bellini Saibene, Y. 2015). Este conjunto de datos etiquetado se armó con datos de campo y de radar correspondientes a 18 fechas con tormentas convectivas del período Marzo de 2009 a Marzo de 2013 en el área de influencia de 240 kilómetros de radio con centro en el radar localizado en la EEA Anguil, La Pampa (figura 2).

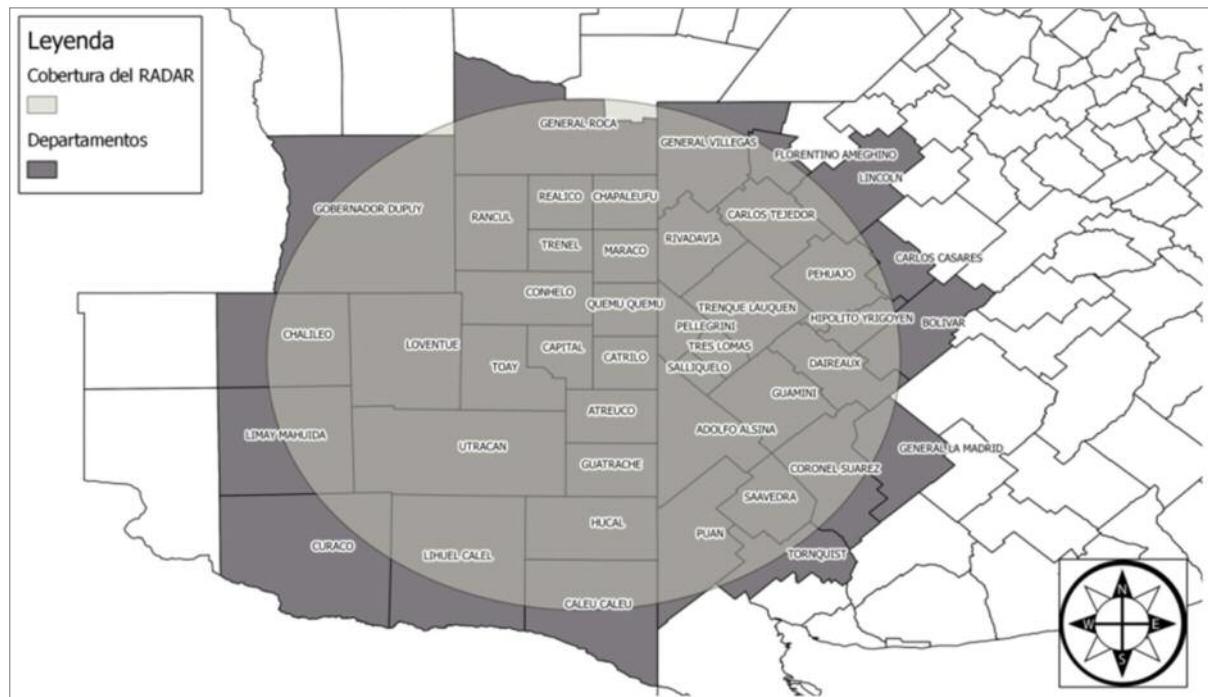


Figura 2. Mapa del área de influencia de 240 km de radio del radar de la EEA Anguil (Fuente: Bellini Saibene, Y. 2015).

Datos del radar

El radar está configurado para completar una serie de giros a 360° que se repiten para 12 ángulos de elevación entre $0,5^\circ$ de base y $15,1^\circ$ de tope con una resolución espacial de 1 km^2 . Este escaneo completo está programado cada 10 minutos, totalizando 144 adquisiciones diarias, bajo funcionamiento normal. El día del radar se extiende de las 00:00 h a las 23:50 h. En este barrido se registran una serie de variables *crudas*, que son útiles para la detección de granizo:

- Factor de reflectividad (Z) (Rinehart 1997)
- Differential Reflectivity - Reflectividad diferencial (ZDR) (Rinehart 1997)
- Correlation Coefficient - Coeficiente de correlación co-polar (ρ_{HV} , RhoHV, CC) (Rinehart 1997)
- Differential Propagation Phase - Fase de propagación diferencial (Φ_{DP} , PhiDP) (Rinehart 1997)
- Specific Differential Phase - Fase diferencial específica (KDP) (Rinehart 1997)

Para este trabajo se utilizaron las variables Z, ZDR, RhoHV las cuales pueden ser procesadas por el software libre RadINTA (Bellini Saibene, Y. et. al. 2014; Bellini Saibene, Y. et. al. 2016). La figura 3 presenta la relación de las variables con la forma de los hidrometeoros.

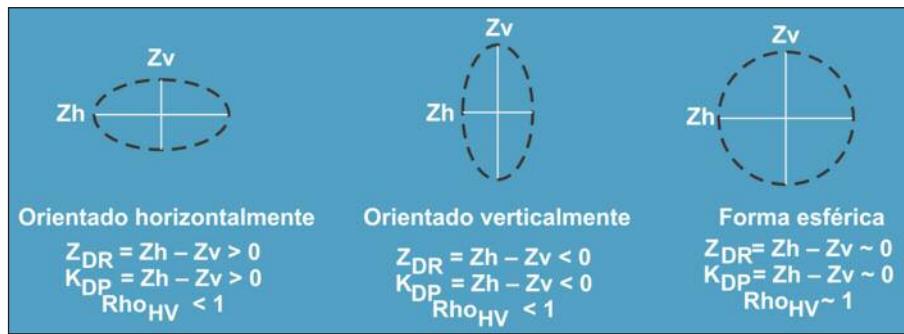


Figura 3. Relación entre los ejes verticales y horizontales de los hidrometeoros y las variables registradas por el radar
(Fuente: Belini Saibene, Y. 2015).

A partir de estas variables y utilizando RadINTA, se realizó el cálculo de un serie de variables derivadas de Z y ZDR que son utilizadas en trabajos previos de detección de granizo: a) Hail Differential Reflectivity (HDR) (Rinehart 1997), b) valor máximo de Z vertical, c) Z media vertical, d) elevación más alta con Z diferente de 0, e) valor de Z de la elevación más alta donde Z es diferente de 0, f) número de elevaciones de Z igual o superior a los umbrales 45, 50, 55 y 60 dBZ, g) Número de elevaciones con Z diferente de 0 y h) Energía Cinética (E). También se agregaron otros cálculos como por ejemplo mínimos, máximos, promedios, totales por cada elevación y de todas las elevaciones juntas para las variables Z, ZDR y RhoHV. Como GEP no permite trabajar con valores perdidos, los cálculos se hicieron por un lado descartando los lotes a los cuales les faltaban datos del radar y por otro reemplazando los valores perdidos por “-99.00”, ya que es un número muy alejado del rango válido que pueden tomar las variables, permite utilizar todos los casos y representar la falta de tormenta en ese pixel.

Datos de campo

A partir del análisis de antecedentes se determinó que las variables de terreno necesarias para el armado de los modelos son: 1) fecha del evento (formato: dd-mm-aaaa) y hora del evento (hora de inicio y fin en formato hh:mm), 2) ocurrencia de granizo (si, no), 3) posición geográfica (en coordenadas geográficas, latitud y longitud).

No existe a nivel nacional un sistema oficial que concentre y gestione el registro de la información de ocurrencia e intensidad de eventos de granizo de forma exhaustiva; por lo que esta información de campo se tuvo que relevar por medio de diferentes

fuentes, con el objetivo de obtener la mayor cantidad de puntos posibles para las fechas registradas.

Las 5 fuentes de información utilizadas son:

- *Compañías de seguros agrícolas:* proveen reportes que se utilizan ampliamente como información para la validación de métodos de estimación de ocurrencia y tamaño de granizo (Holleman 2001, Mezher, et al. 2012, Mezher et al. 2010, Mezher, et al. 2011, Holleman et al. 2000, Changnon Jr 1971, Tuovinen et al. 2009, Changnon et al. 1997). Se contactaron vía correo electrónico 29² empresas, de las cuales Sancor Seguros, La Segunda y La Dulce Cooperativa de Seguros enviaron datos; si bien los registros recibidos tienen diferente formato (.xls, .kml y .txt) y diferentes datos, todos comparten dos atributos: 1) georeferenciación del establecimiento donde ocurrió el evento y 2) una calificación del daño ocasionado por el granizo en el cultivo.

- *Medios de comunicación y redes sociales:* aparecen como segunda fuente mencionada en otros estudios de granizo (ej.: Aydin, et. al. 1986, Boodoo et al. 2009, Collino, et al. 2009, Tuovinen et al. 2009, Saltikoff et al. 2010, Bonelli et al. 2011, Prieto et al. 2009). Los eventos severos son un tema sensible capturado por los medios locales en el área de estudio. Las notas sobre las tormentas y sus consecuencias son acompañadas por evidencia fotográfica o filmica, además de testimonios de entidades oficiales como la policía, los bomberos, cooperativas, asociaciones de productores o gobiernos municipales; sumado a los contenidos de fotos y videos que comparten los usuarios en la web 2.0, permitió obtener 223 reportes en 63 medios de comunicación y redes sociales que se presentan en formato digital (diarios, radios, televisión, blogs, videos y sitios oficiales). Del análisis de estos reportes se extrajo información diversa como: localidades afectadas, superficie del daño, horario, duración, tipo de even-

² Esta cantidad de empresas se obtuvo de «Resultados de la “Encuesta Censal sobre los Seguros en el Sector Agropecuario y Forestal”» 2013). Los datos de contacto se buscaron en los sitios web de cada empresa.

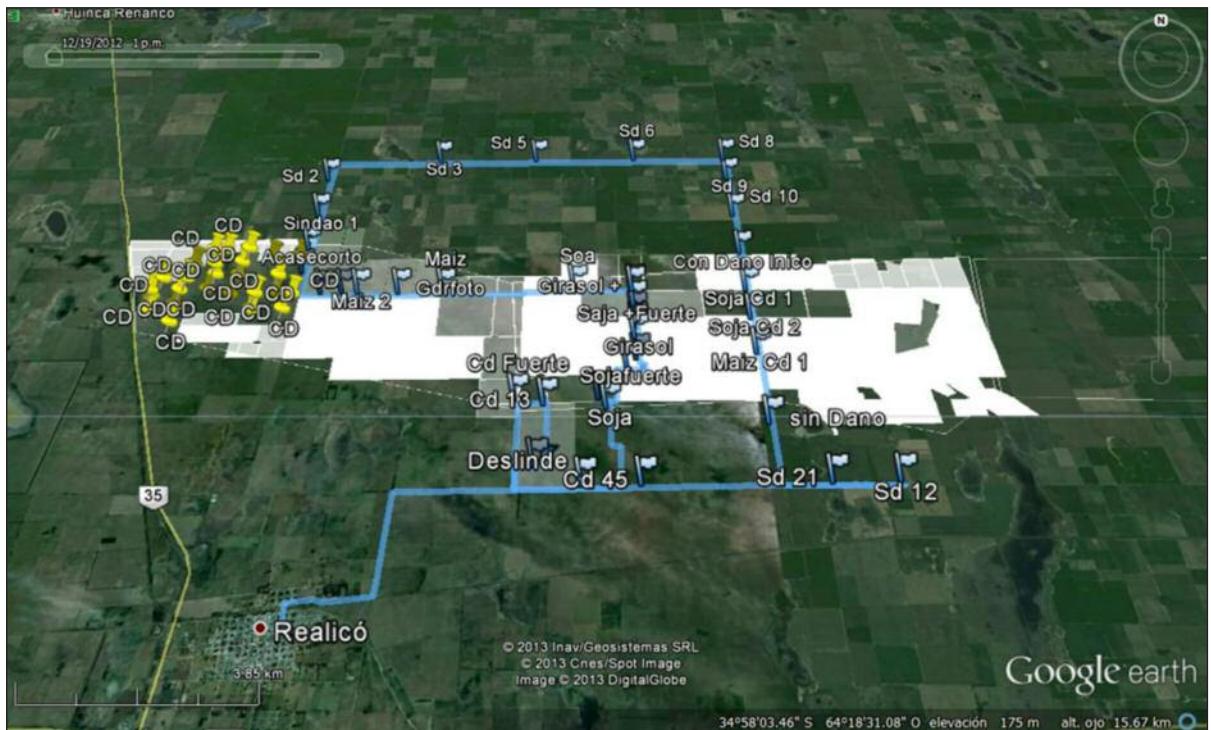


Figura 4. Primer tramo de la recorrida de campo correspondiente al evento 10-12-2012 (Bellini Saibene, Y. 2015).

to (lluvia, viento, granizo o sus combinaciones) y efecto causado en terreno por las tormentas. Para la base de datos de verdad de campo se registró: la fecha del evento, la georeferenciación de las localidades afectadas y el horario de inicio y fin del evento, en caso que figurara. Si las localidades mencionadas en los reportes, cuentan con un medidor en alguna red de observación meteorológica a la cual se tenía acceso, esa localidad solo se agregaba si en la red se informaba precipitación.

• **Redes de información:** son utilizadas por diversos trabajos (ej: Mezher, et al. 2012, Aydin, et al. 1986; WW2010; Alexiuk, et al. 1999; Tuovinen et al. 2009; Saltikoff et al. 2010; Mezher R.N. et al. 2010). En este trabajo se utilizaron datos de: 1) Red de informantes del Servicio Meteorológico Nacional (SMN), los cuales reportan eventos extremos a este organismo que se publican en <http://www.smn.gov.ar/?mod=voluntarios&id=1>, 2) Red de pluviómetros de la policía de La Pampa que informa en <http://www.policia.lapampa.gov.ar/lluvias.php>, 3) Red termo pluviométrica de la Red de Información Agropecuaria Nacional (RIAN), que releva información diaria de precipitaciones en aproximadamente 100 puntos en el área de estudio (<http://ian.inata.gov.ar/agua/informes.aspx>) y 4) Sistema Integrado de Información Agropecuaria (SIIA) que genera informes semanales de estimaciones de superficie sembrada por delegación provincial, indicando para cada cultivo el estado general, avances de labores, estado fenológico y principales adversidades (<http://www.siiia.gob.ar/informes>).

ta.gov.ar/agua/informes.aspx) y 4) Sistema Integrado de Información Agropecuaria (SIIA) que genera informes semanales de estimaciones de superficie sembrada por delegación provincial, indicando para cada cultivo el estado general, avances de labores, estado fenológico y principales adversidades (<http://www.siiia.gob.ar/informes>).

- **Informantes calificados:** por ejemplo, López, L. 2009, Billet et al. 1997, Li, P. C. et al. 2000, Ceperuelo Mallafré 2008, Anderson et al. 2012, Alexiuk, et al. 1999, Tuovinen et al. 2009, Saltikoff et al. 2010, A. Ferreyra 2011, Ortega et al. 2009, utilizan informantes calificados para obtener información de campo. En este trabajo se contactaron a extensionistas de INTA, productores y contratistas locales. Con Google Earth se ubicaron y digitalizaron los lotes informados y se cargó la información de ocurrencia de granizo (si, no), cultivo afectado y porcentaje de daño (figura 5).

- **Recorridas de campo posteriores a una tormenta:** en trabajos previos (Aydin, et. al 1986, Heinselman, et. al 2010, A. Ferreyra 2011; Bringi, et. al 1986) se realizan salidas ex pos a la ocurrencia de una tormenta granizera. Se aprovecharon recorridas



Figura 5. Lotes de soja con daño por granizo (tormentas del 10-12-2012 y 01-03-2013).

mensuales de la RIAN donde se registran datos referentes al estado y evolución de los cultivos y se incluyen las principales adversidades de origen ambiental dentro de las cuales se encuentra el granizo. (Yanina Bellini Saibene, 2015). También se realizaron dos salidas posteriores a las tormentas del 10-12-2012 y el 24-12-2012. Ambas recorridas se realizaron por rutas y caminos vecinales, acompañados por productores y contratistas locales, los cuales guiaron en el recorrido y permitieron el acceso a los lotes con daño en sus establecimientos agropecuarios. Durante el recorrido se obtuvo información de georeferenciación (latitud y longitud, con un GPS Garmin eTrex Legend HCx), cultivo dañado, porcentaje de daño y tareas a realizar en los lotes con daño por granizo (ej.: resiembra, picado). También se registraron las coordenadas geográficas de lotes sin daño de granizo presentes en el área de la tormenta. En la figura 4 se muestra un mapa con las recorridas realizadas para el relevamiento correspondiente al evento del 10-12-2012 y en la figura 5 se presentan fotos de lotes dañados por granizo.

Generación de los conjuntos de datos

Para gestionar y unificar toda la información de radar y de campo recolectada para una fecha, se utilizó HAR (Bellini Saibene, Y. et. al. 2015b). Se relevaron en total 3018 lotes (41% positivos y 59% negativos). Las 79 variables generadas se agruparon de diferente manera con los objetivos de: 1) analizar el poder clasificadorio y el comportamiento de las variables de doble polarización ante la presencia de granizo; 2) analizar si se puede generar un modelo donde solo se necesite Z como variable de entrada; 3) conocer cómo influye en la clasificación representar en los datos la falta de tormenta en un momento y lugar específico. Se realizaron análisis estadísticos utilizando el software R para examinar la relación de las diversas variables con la variable objetivo (target), lo cual ayudó a generar los diferentes set de datos.

Esta agrupación dio como resultado seis conjuntos de datos. La figura 6 presenta el detalle de la distribución de los casos en los seis dataset generados y la tabla 1, el detalle del agrupamiento.

Tabla 1. Listado de combinaciones de variables en los diferentes conjuntos de datos.

Tipo de Variable	Elevaciones	Datos
Polarimétrica	1ra Elevación	Sin datos faltantes
		Con datos faltantes
	Todas las elevaciones	Sin datos faltantes
		Con datos faltantes
Derivadas de Z	1ra Elevación	Sin datos faltantes
		Con datos faltantes
	Todas las elevaciones	Sin datos faltantes
		Con datos faltantes

Modelado

La figura 7 presenta el esquema de modelado con la configuración de GEP (correspondiente a los valores por defecto de la plantilla para regresión logística); la configuración de los set de datos y la estrategia de entrenamiento y testeо. Las condiciones de parada del algoritmo fueron: llegar al valor

1000 en el “Max Fitness” o llegar a la generación 30.000, lo que ocurra primero.

Para evaluar y comparar los modelos obtenidos se utilizan indicadores basados en la matriz de confusión (tabla 2) sobre un conjunto de casos que los modelos generados “ven” por primera vez; ese conjunto de datos se conoce como testing. En este trabajo se detallan los tres indicadores más utilizados en los trabajos de modelos de eventos severos generados con DM y usando datos de radares meteorológicos (Bellini Saibene Y., 2015).

False Alarm Ratio (FAR): número de falsas alarmas positivas dividido por el número total de positivos predichos. Para un modelo perfecto FAR=0. Se calcula:

$$\text{FAR} = \text{FP}/(\text{VP}+\text{FP})$$

Probability Of Detection (POD): capacidad para identificar resultados positivos correctamente. Modelo perfecto POD=1. También se denomina: sensibilidad, recall, exhaustividad. Se calcula:

$$\text{POD} = \text{VP}/(\text{VP}+\text{FN})$$

Percent Correct (PC): indica la proporción de los resultados reales (VP y VN) en la población. También se denomina Accuracy. Se calcula:

$$\text{PC} = (\text{VP}+\text{VN})/(\text{VP}+\text{VN}+\text{FP}+\text{FN})$$

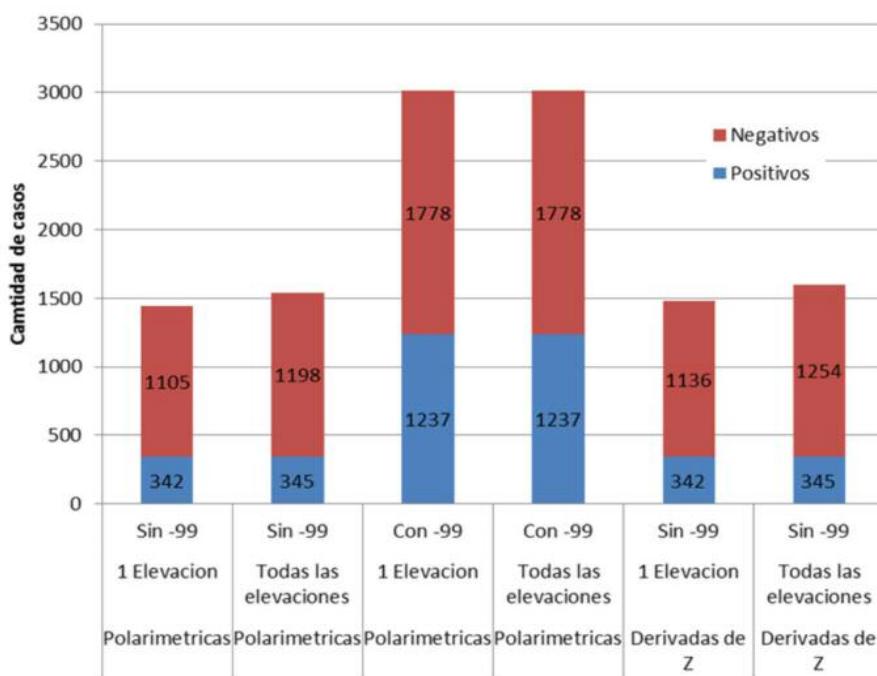


Figura 6. Distribución de los casos en los seis dataset para modelar el target Granizo.

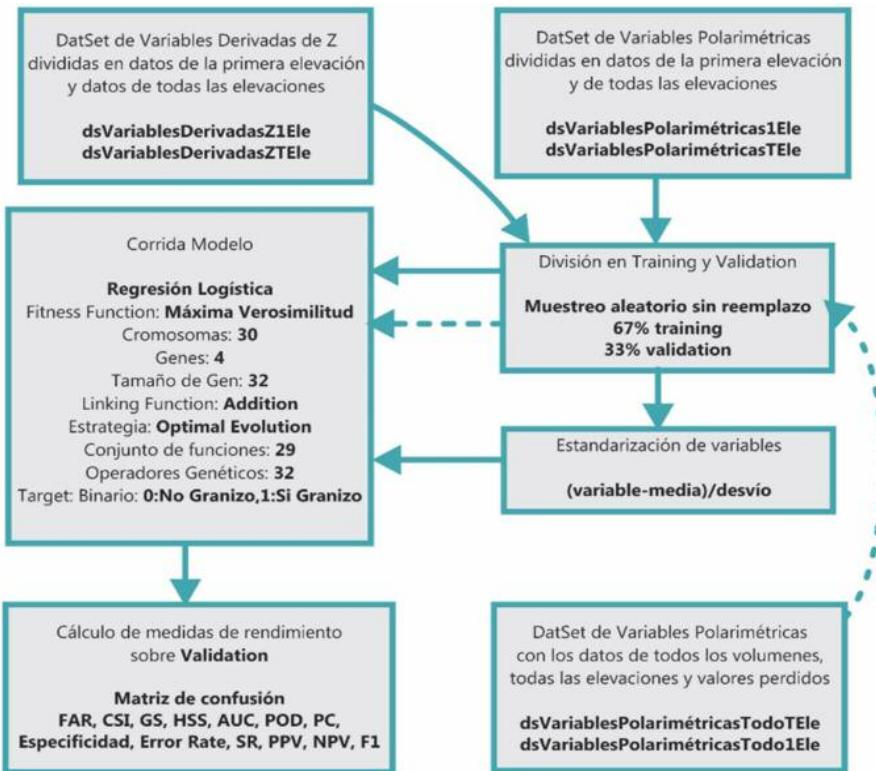


Figura 7. Esquema de tareas realizadas en la etapa de modelado del target Granizo (Bellini Saibene, Y., 2015)

Tabla 2. Matriz de Confusión (Contingencia)

		Valores Observados	
		Si	No
Valores predichos	Si	VP (Verdaderos Positivos) Identificados Correctamente	FP (Falso Positivo) Identificados incorrectamente
	No	FN (Falso Negativo) Incorrectamente rechazados	VN (Verdadero Negativo) Correctamente rechazados

Implementación

La implementación del modelo se programó con Python, para continuar con el mismo lenguaje utilizado por RADINTA. Todo el procesamiento de los datos crudos del radar se realiza con dicho software, generando las imágenes raster correspondientes a las variables seleccionadas por los modelos generados. Estas imágenes son los datos de entrada de los modelos generados, que pueden clasificar una serie de lotes indicados por medio de latitud y longitud o bien una imagen completa del área de influencia de 240 km del radar meteorológico.

RESULTADOS Y DISCUSIÓN

Análisis Estadísticos

Conjunto de datos

La figura 8 muestra los diagramas de caja donde se aprecia la diferencia en la distribución de las variables Z (dBZ), RhoHV y ZDR, entre los casos de las fechas que tienen los volúmenes completos con respecto a las que no. Además el ANOVA de un fac-

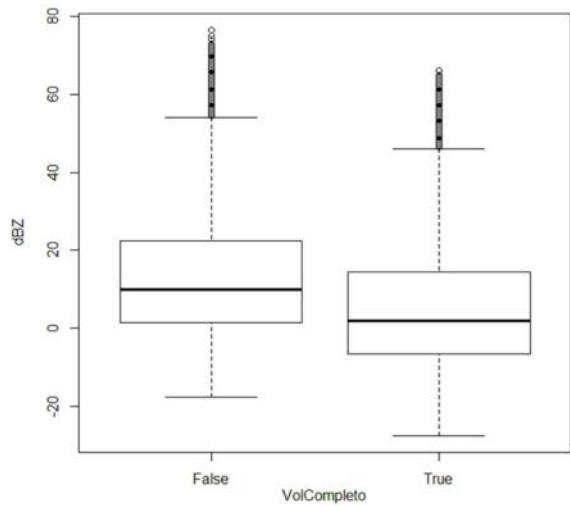


Figura 8.a. Diagramas de Caja de la variable Z

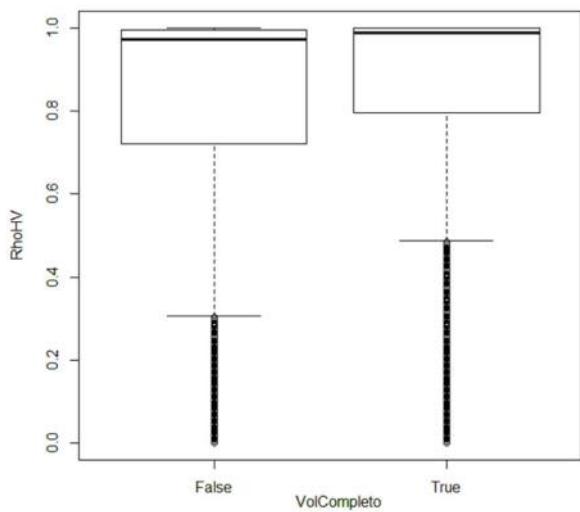


Figura 8.b. Diagramas de Caja de la variable Rho_{HV}

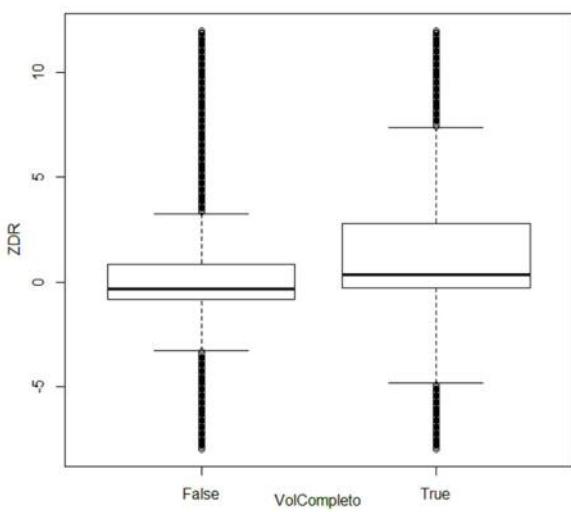


Figura 8.c. Diagramas de Caja de la variable Z_{DR}

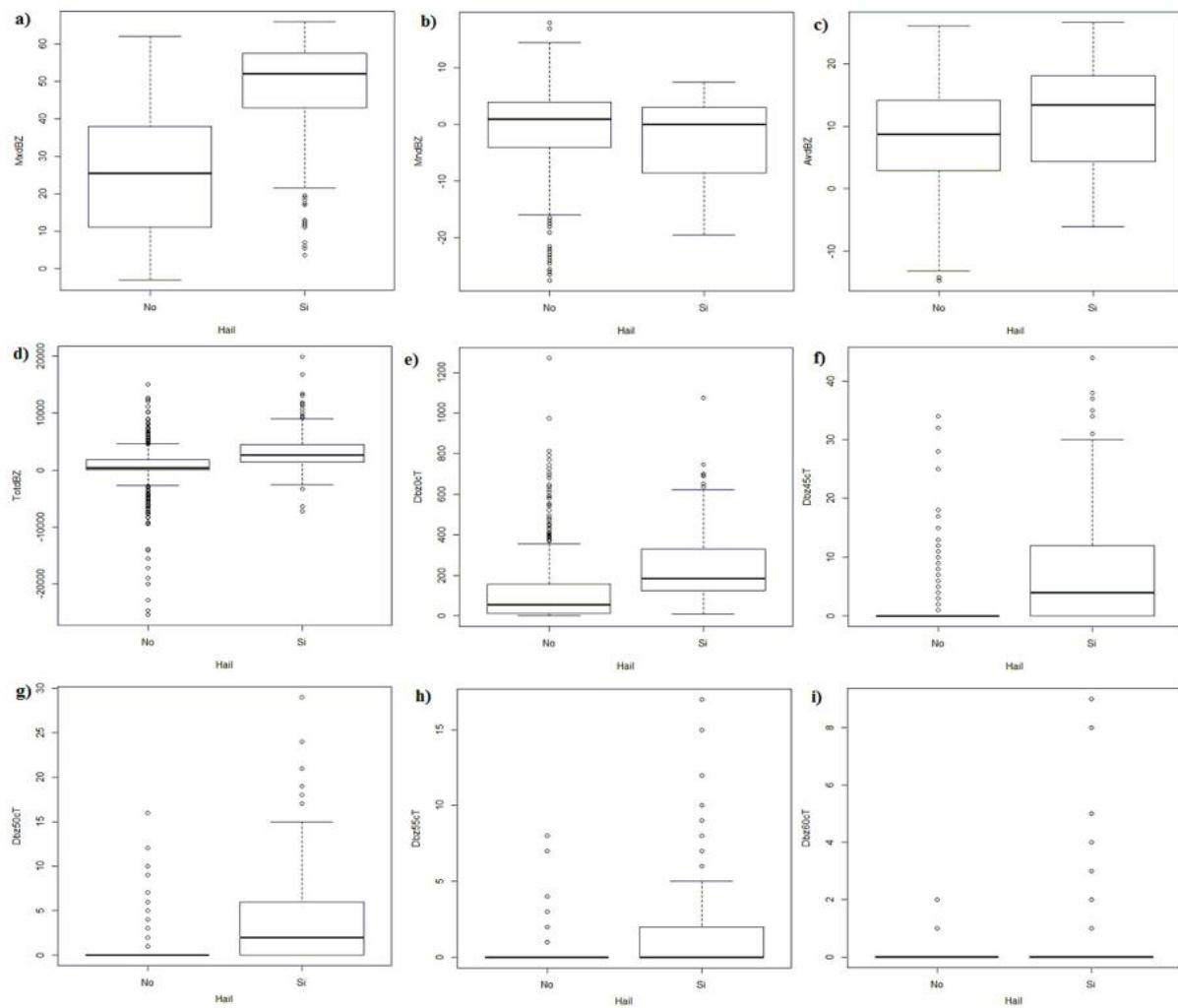


Figura 9. Diagramas de Caja de las variables: a) $MxdBZ$, b) $MndBZ$ y c) $AvdBZ$, d) $TotdBZ$, e) $DbzocT$, f) $Dbz45cT$, g) $Dbz50cT$, h) $Dbz55cT$, i) $Dbz60cT$.

tor mostró diferencias significativas en los valores de dBZ ($\text{Pr}(\text{>F}) = <2\text{e-16}$), ZDR ($\text{Pr}(\text{>F}) = <2\text{e-16}$) y RhoHV ($\text{Pr}(\text{>F}) = <2\text{e-16}$) entre los volúmenes completos y los incompletos. Esta evidencia apoya la decisión de construir una serie de datasets solo con los casos cuyas fechas tenga todos los volúmenes presentes y otros con todos los casos (tengan o no todos los volúmenes presentes).

Reflectividad

La figura 9 presenta los diagramas de caja de nueve variables generadas a partir de Z (para todas las elevaciones) y su distribución con respecto del target granizo. En todas las variables, los casos positivos presentan mayores valores de Z y mayor cantidad de presencia de ecos de reflectividad ($>0 \text{ dBZ}$) y ecos fuertes (45, 50, 55 y 60 dBZ), lo cual coincide

con los valores propuestos por Tabary *et al.* 2009; Anderson *et al.* 2012; Zrnic 2012 y May *et al.* 2003. El promedio de los valores máximos de Z para los casos positivos es de 47.5 dBZ (figura 9.a), lo que sugiere que se debería aumentar el umbral de 45 dBZ propuesto en la configuración del radar de INTA Anguil (Gematronik, 2007); se propone el uso de 50 dBZ para la zona bajo estudio, utilizando el mismo valor que Mezher, R. N *et. al.* 2010 y Mezher, R. N *et. al.* 2011 usan con el radar de INTA Pergamino. Este valor es menor al sugerido por los mismos autores para el radar de INTA Paraná (60-65 dBZ).

RhoHV

La figura 10 muestra los diagramas de caja de las variables de resúmenes de RhoHV para todas las elevaciones. Con excepción del máximo (MxRho) el

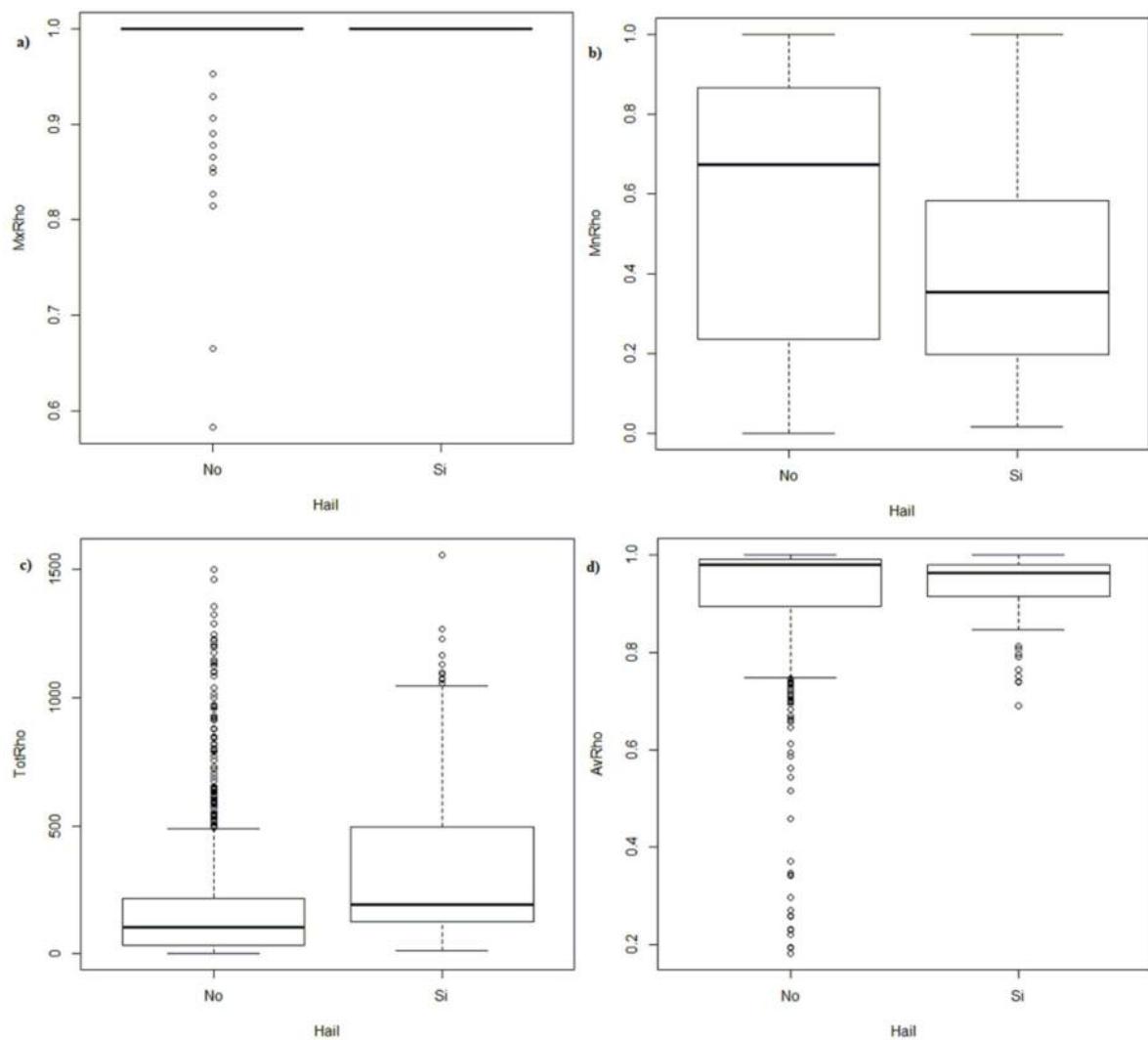


Figura 10. Diagramas de Caja de las variables: a) MxRho, b) MnRho y c) TotRho, d) AvRho

resto de las variables muestran diferencia entre las clases del target; para la variable de mínimos (MnZDR) los valores se alejan marcadamente de 1 en ambas clases, siendo los casos positivos los que menor valor presentan. La variable promedio (AvZDR) muestra una mayor dispersión en los casos negativos y también una caída en la media de AvZDR para los casos positivos. Estos comportamientos coinciden con lo mencionado en la bibliografía para la banda C (Tabary *et al.* 2009; Anderson *et al.* 2012; Zrnic 2012; May *et al.* 2003 y para Argentina Mezher, R.N. *et. al.* 2012). Los valores menores a 0,95 presentes en este trabajo se pueden deber a áreas de granizo mezclado con grandes gotas de lluvia, que generan una gran variedad de formas y maneras de caer de los hidrometeoros de acuerdo a Anderson *et al.* 2012.

ZDR

Los valores máximos y mínimos de ZDR, presentan diferencias importantes entre las clases del target. El promedio de los valores mínimos es de -5.14 dB (figura 11.b), el promedio de los valores máximos es de 5.69 dB (figura 11.a) y el promedio de las medias es de 0.54 dB (figura 11.d). Estos promedios de los máximos y los mínimos coinciden con los valores observados por Anderson *et al.* 2012; Zrnic 2012; May *et al.* 2003 y Keenan 2003 para banda C, donde los valores son elevados (> 4 dB); mientras que el promedio de las medias coincide con el comportamiento observado para banda S, donde los valores son cercanos a cero (Bringi *et al.* 2001; Anderson *et al.* 2012; Heinselman *et al.* 2010; Brandes *et al.* 2004; Ryzhkov *et al.* 2010). Ésta marcada variación en los valores de ZDR fue observada

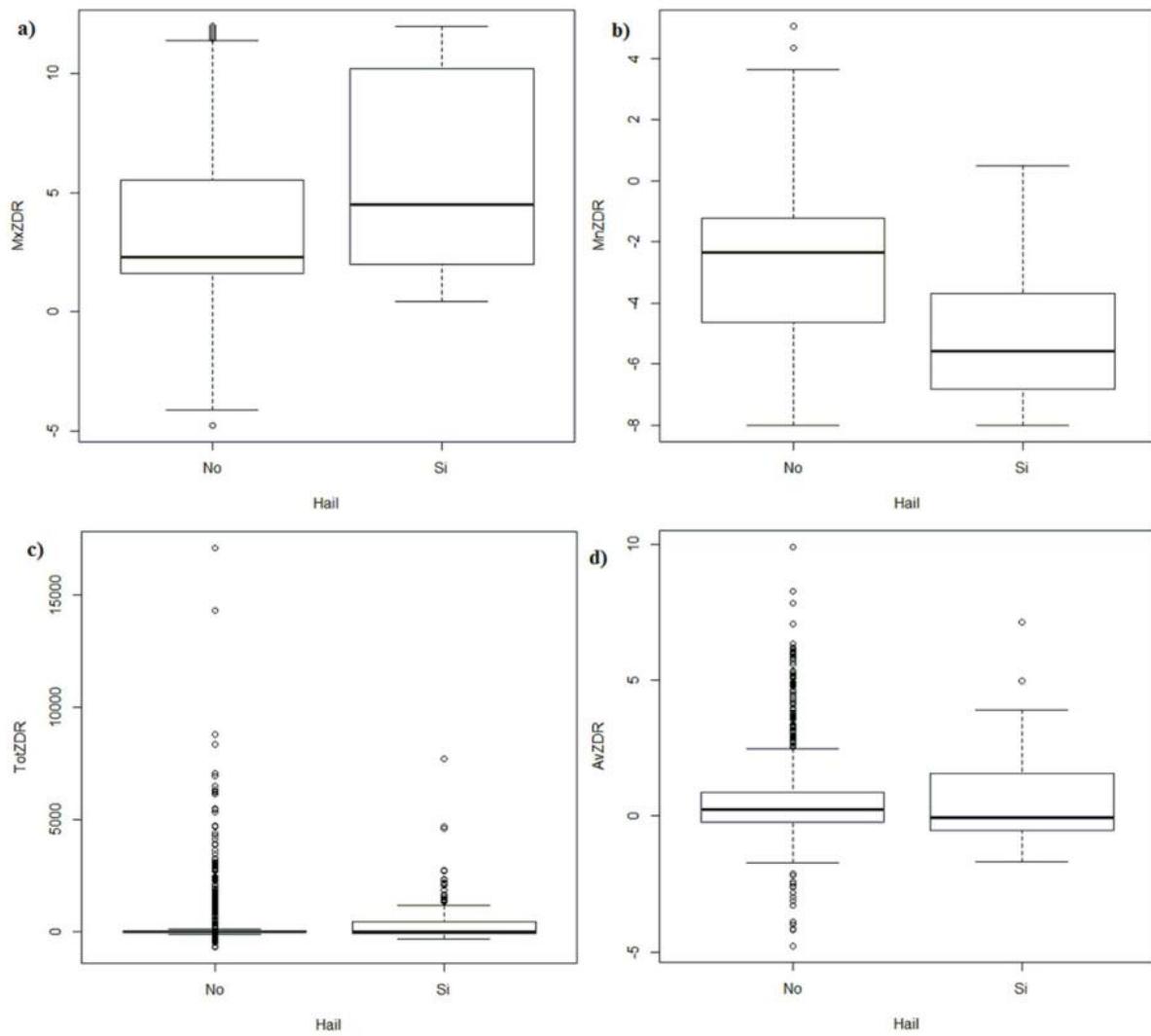


Figura 11. Diagramas de Caja de las variables: a) MxZDR, b) MnZDR y c) TotZDR, d) AvZDR

Tabla 3. Detalle de los tres mejores modelos

	Simple polarización (Z)	Doble polarización	
Variables que entraron en el modelo	Dbz45cT Dbz55cT Dbz1T Dbz45T Dbz55T MxdBZ MndBZ TotEWt MnEWt	MxdBZ MxRho MnRho TotRho MxZDR MnZDR AvZDR	MxdBZ MndBZ TotdBZ AvdBZ MxRho TotRho AvRho MxZDR MnZDR TotZDR AvZDR
Medidas	Resultados dataset de Testing		
FAR	0,3413	0,2353	0,1441
POD	0,7217	0,8426	0,8437
PC	0,8593	0,9125	0,8739

por Tabary *et al.* 2009, que encontró valores de -5 dB hasta 5 dB pasando por los valores cercanos a 0, ante la presencia de diferentes formas de granizo (grande, derritiéndose, mezclado con lluvia, etc.). Este comportamiento también lo observaron

Mezher, R. N. *et. al.* 2012 y Mezher, R. N. *et. al.* 2011 en el radar de INTA Paraná ante la presencia de granizo.

Modelos Obtenidos

Todos los modelos presentan rendimientos muy buenos en la clasificación de los casos en el dataset de testing. Los modelos con variables polarimétricas son mejores que aquellos que usan las variables derivadas de Z, lo cual coincide con los antecedentes (Bringi *et al.* 2001, Aydin, et. al 1986, Ryzhkov *et al.* 2012, Anderson *et al.* 2012, Heinselman *et al.* 2010, Brandes *et al.* 2004). Los dos mejores modelos de doble polarización corresponden a los datos de todas las elevaciones y sin estandarizar, mientras que el mejor modelo basado en Z utiliza variables de todas las elevaciones, estandarizadas. La tabla 3 presenta los tres mejores modelos.

Del análisis de las variables incluidas en todos los modelos se destacan para clasificar el target: a) el valor máximo de Z (MxdBZ, seleccionada en 9 modelos), b) La presencia de los ecos de 45 dBZ, elegida en los cuatro modelos de Z. Los ecos de 50,

Función logística						
P(Y=1) =	$\frac{1}{1 + \exp(0.112107575724697*y + -6.45746730193226)}$					
Armado de y						
$y = (\text{gepGOE2A}((\text{AvZDR} * \text{gepGOE2C}(-1.6825800347911, 8.18572069460128)), -7.17642750328074) + (3.55021820734275 - \text{gepLT2C}(\text{MxRho}-\text{MxdBZ}, -1.6825800347911)))$ $y = y + (9.3725394451735 + (((\text{gepGOE2G}(\text{gepLT2A}(\text{MnZDR}, \text{TotRho}), \text{dMnRho}) + (\text{dMxZDR}-\text{dMnZDR})/2.0)) - (\text{gepAND1}(-19.3568254646443, \text{dMxZDR}) - \text{dAvZDR})))$ $y = y + (((\text{gepAND1}(-9.27060762352367, \text{MxRho}) + (-6.30753501998962 + 7.53898739585559)/2.0))/2.0) - \text{gepGOE2G}(\text{gepAND1}(\text{MxZDR}, \text{dMnZDR}), \text{TotRho}) * \text{gepAND2}((-5.29526657918027 + \text{MxdBZ})/2.0), \text{AvZDR}))$ $y = y + (\text{MnRho} - (\text{gepGOE2E}(8.75606555375835, \exp(\text{MnRho})) * \text{MnRho}))$						
Funciones lógicas para el armado de y						
gepAND1	gepAND2	gepLT2A	gepLT2C			
Recibe dos valores x e y: Si (x < 0) y (y < 0) devuelve 1, si no devuelve 0	Recibe dos valores x e y: Si (x >= 0) y (y >= 0) devuelve 1, si no devuelve 0.	Recibe dos valores x e y: Si (x < y) devuelve x, si no devuelve y.	Recibe dos valores x e y: Si (x < y) devuelve (x+y), si no devuelve (x-y)			
gepGOE2A	gepGOE2C	gepGOE2E	gepGOE2G			
Recibe dos valores x e y: Si (x >= y): devuelve x, si no devuelve y	Recibe dos valores x e y: Si (x >= y): devuelve (x+y), si no (x-y)	Recibe dos valores x e y: Si (x >= y): devuelve (x+y), si no devuelve (x*y)	Recibe dos valores x e y: Si (x >= y): devuelve (x+y), si no devuelve atan(x*y)			

Figura 12. Función logística para la probabilidad de granizo positivo del Modelo Kurá y detalles de su construcción.

Función logística						
P(Y=1) =	$\frac{1}{1 + \exp(6.48084911389186E-05*y - 4.63926726110945)}$					
Armado de y						
$y = \text{gepLT2C}((\text{gepLT2G}(9.64666513565478*(1.0-\text{TotdBZ})),(1.0-\text{gepLT2G}(\text{MxZDR},\text{MnZDR}))*\text{AvRho}),\text{AvdBZ})$						
$y = y + \text{pow}(\exp(\exp(\text{gepGOE2G}(\text{gepLT2C}(\text{gepLT2B}(\text{TotdBZ},\text{TotZDR}),\exp(-4.10822324594867)),\text{gepAND2}(\text{gepLT2E}(\text{TotRho},\text{TotZDR}),\text{pow}(\text{MxdBZ},3.0)))),4.0)$						
$y = y + \text{gepGOE2C}(\text{pow}((\text{MnZDR}/\text{gepLT2G}(\text{TotRho},\text{TotZDR})),2.0),\text{gepLT2E}(\text{gepLT2E}(\text{gepLT2G}(\text{MxZDR},\text{MxdBZ}),(\text{MxRho}-\text{MxZDR})),(1.0-\text{MxdBZ}))$						
$y = y + \text{gepGOE2G}(((\text{gepGOE2G}(\text{MxZDR},\text{AvRho})*\text{MndBZ})+\text{pow}(\text{MxdBZ},3.0))/2.0)+(\text{gepLT2G}(\text{AvRho}, 5.10116885891293)*\text{gepLT2G}(\text{AvZDR},\text{AvRho})),\text{AvZDR})$						
Funciones lógicas para el armado de y						
gepAND2	gepLT2B	gepLT2C	gepLT2E			
Recibe dos valores x e y: Si ($x \geq 0$) y ($y \geq 0$): devuelve 1 si no 0	Recibe dos valores x e y: Si ($x < y$): entonces devuelve 1, si no devuelve 0	Recibe dos valores x e y: Si ($x < y$) devuelve ($x+y$), si no ($x-y$)	Recibe dos valores x e y: Si ($x < y$) devuelve ($x+y$), si no devuelve ($x*y$)			
gepLT2G	gepGOE2C	gepGOE2G				
Recibe dos valores x e y: Si ($x < y$): devuelve ($x+y$), si no devuelve $\text{atan}(x*y)$	Recibe dos valores x e y: Si ($x \geq y$) devuelve ($x+y$), si no devuelve ($x-y$)	Recibe dos valores x e y: Si ($x \geq y$): devuelve ($x+y$), si no devuelve $\text{atan}(x*y)$				

55 y 60 dBZ entraron en la mitad de los modelos.; c) el máximo y mínimo de ZDR, seleccionado en los seis modelos polarimétricos; d) el mínimo de RhoHV seleccionado en cinco de los seis modelos y e) el total de E, elegida en todos los modelos de Z.

Los dos modelos que tienen la mejor performance se seleccionaron y se nombraron con los vocablos mapuches Kurá que significa Piedra y Pire que significa Nieve-Granizo, con el objetivo de darles una identidad territorial asociada al área de estudio y al lugar donde se desarrollaron; permitiendo además que puedan ser identificados y recordados fácilmente por los usuarios.

Cuando se compara el rendimiento de estos dos modelos con los reportados por otros 30 modelos implementados a nivel internacional, Kurá y Pire ocupan el tercer puesto (de acuerdo a las medidas PC, POD y FAR que son reportadas en las publicaciones correspondientes) (Bellini Saibene, Y., 2015).

Las figuras 12 y 13 presentan la función logística para la probabilidad de granizo positivo para los dos mejores modelos y los detalles de su construcción.

Figura 13. Función logística para la probabilidad de granizo positivo del Modelo Pire y detalles de su construcción.

Implementación

Se probaron dos implementaciones, una utilizando bases de datos relacionales para realizar los cálculos de las variables de entrada y la otra realizando el procesamiento por medio de imágenes compuestas, utilizando el software RadINTA. Este último resultó mucho más rápido tardando 1,3 días en procesar una imagen completa (Bellini Saibene, Y. 2015). El cuello de botella se encuentra en la transformación de los datos crudos a imágenes, aunque es una tarea que se puede paralelizar. La figura 14 muestra el resultado de la implementación de uno de los modelos.

CONCLUSIONES

Se lograron desarrollar dos modelos de clasificación de granizo en superficie con información de un radar meteorológico de banda C con una alta performance. Los dos modelos seleccionados se nombraron como Kurá y Pire, por las palabras mapuches correspondientes a Piedra y Granizo/Nieve. Los

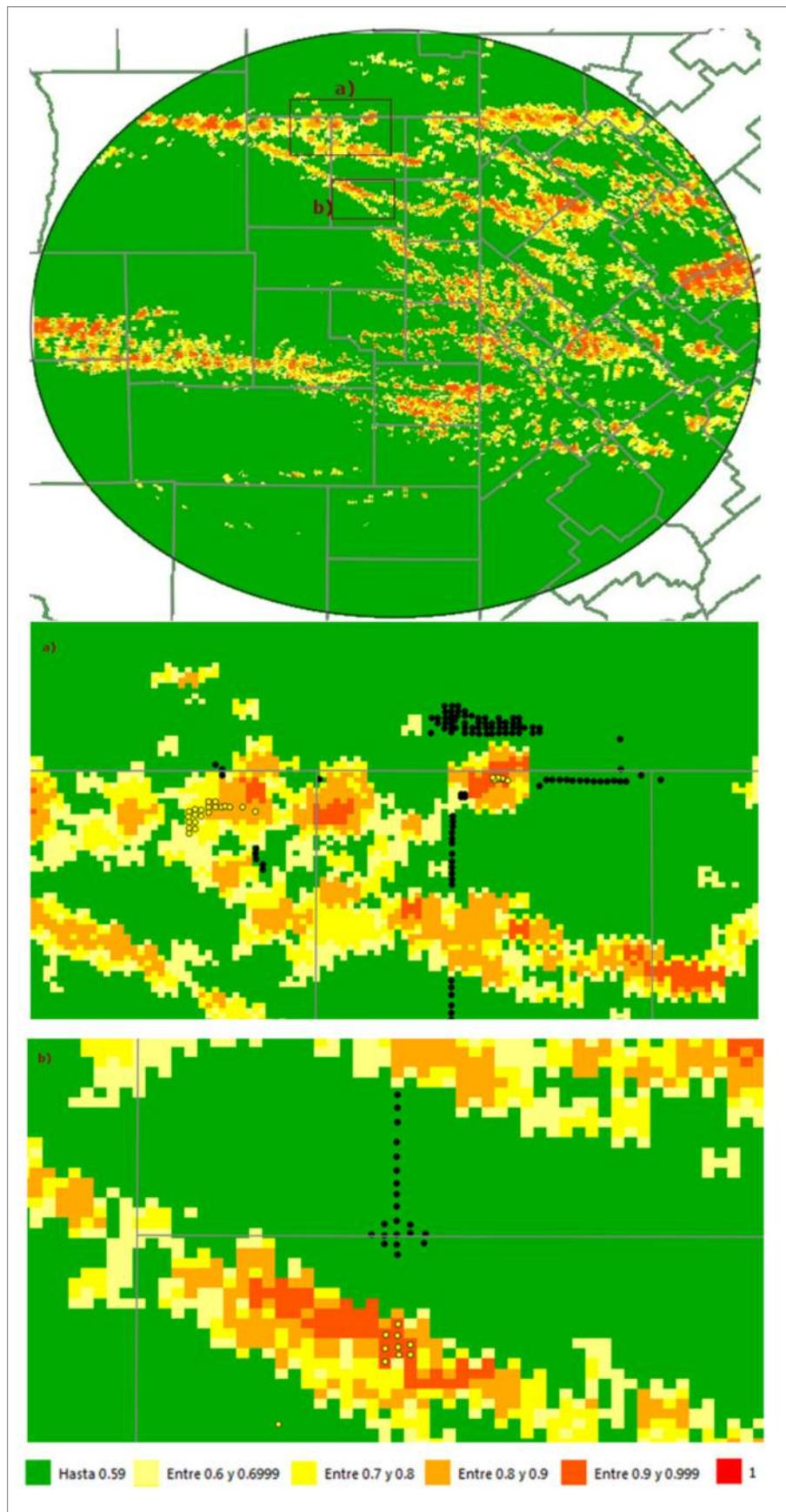


Figura 14. Resultados de la implementación de la clasificación de granizo: imagen completa del 10-12-2012 modelo 3 de granizo. a) y b) Detalle de la clasificación de la imagen y los puntos de verdad de campo: negro sin registro de granizo, amarillo con registro de granizo.

valores de POD (84%), FAR (23% y 14%) y PC (86% y 84%) de estos dos modelos fueron superados sólo por dos algoritmos (de un total de 30 modelos ya existentes internacionalmente).

Se pudieron utilizar como variables de entrada para los modelos, resúmenes (máximos, mínimos, promedios, totales y cantidades) de las variables Z, ZDR y RhoHV. También se pudieron calcular las variables E y HDR derivadas de Z y Z_{DR}. Para la identificación de granizo, los modelos con variables polarimétricas funcionaron mejor que aquellos que usan las variables de simple polarización. Se observó que, a mayor valor de Z, menores valores de Rho_{HV} y valores extremos de Z_{DR}, mayor es la probabilidad de presencia de granizo. Se establece que el umbral de Z para determinar presencia de granizo cuando se trabaje con simple polarización, se debe fijar entre 50 y 55 dBZ para este radar.

El problema planteado cuenta con la características de una problemática a resolver con DM debido: a) al gran volumen de información generada por el radar, b) la dificultad de realizar un diseño experimental en la recolección de los datos de campo (ocurrencia del granizo y daño en los cultivos), c) la necesidad de buscar relaciones no lineales como soluciones al problema y d) la baja frecuencia del caso positivo con relación al negativo. La aplicación de CRISP-DM fue adecuada para el problema a resolver y se cumplió que la etapa de preparación de los datos insumió el 70% del tiempo del proceso.

El método GEP (regresión logística) resultó ser una muy buena herramienta para clasificar la ocurrencia de granizo en superficie.

Se pudieron implementar tres modelos, logrando clasificar una serie de lotes e imágenes completas. Estas implementaciones demostraron inconvenientes con el almacenamiento de los datos del radar y su posterior procesamiento, tardando como mínimo 1,3 días en obtener una clasificación de una imagen completa.

RECOMENDACIONES

A partir de los resultados de este primer análisis exploratorio, sería conveniente en futuros trabajos, poder aislar el momento de ocurrencia de la tormenta y trabajar solo con los datos correspondientes a ese momento, evitando de esta manera suavizar los valores de las variables del radar. Otra característica

a determinar, para evitar la subespecificación, son las elevaciones que realmente aportan a la identificación de la caída en superficie. También sería muy importante darle continuidad a la actualización y ampliación de la información de verdad de campo utilizando el sistema de información desarrollado para tal fin. Además, sería interesante realizar una comparación de los resultados del algoritmo ZHAIL del software Rainbow, con los resultados de los modelos generados en este trabajo para las mismas fechas y zona de estudio.

A medida que aumenta la distancia desde el radar, el haz emitido toma datos a mayor altura (pudiendo perder datos de la tormenta) y se ensancha (aumentando el volumen de muestreo). En la estimación de otras variables como la lluvia, la distancia influye en la precisión de la misma, por lo que el análisis de este aspecto se debería abarcar en futuros estudios.

La distribución espacial del granizo puede ser una variable de entrada importante, en futuros trabajos se debería estudiar la forma de incorporar la misma a los modelos y analizar su influencia. Esto incluye la posibilidad de estudiar el valor que toman las variables del radar no solo en el pixel correspondiente a la localización de verdad de campo, si no a aquellos píxeles circundantes a la misma.

Si bien en este trabajo el objetivo era utilizar datos de un solo sensor remoto, otra variable de interés es la temperatura provista por radio sondos, para poder determinar la isoterma de cero grado que puede ayudar a mejorar las clasificaciones. Un estudio futuro de múltiples sensores sería conveniente.

Las variables PhiDP y KDP se deberían incluir en futuros modelos para analizar su aporte a los target buscados, una vez que se pueda programar su lectura y procesamiento.

Para que el modelo sea más útil es necesario encontrar una manera eficiente de almacenar y catalogar los datos disponibles del radar; el uso de bases de datos no estructuradas con productos como Mongo DB o Hadoop se presentan como una alternativa a estudiar. La velocidad de procesamiento también es una característica que se debe mejorar; como estos problemas se pueden parallelizar, es conveniente analizar el uso de metodologías como Map Reduce. Se debería trabajar en lograr una implementación completamente automática de los modelos obtenidos.

Finalmente, sería provechoso generar los modelos con otras técnicas de DM que se mencionan en los antecedentes, con los objetivos de buscar la más eficiente para este tipo de información ya sea por una mejora en la performance de los modelos o por el uso de menos cantidad de variables o información, permitiendo que la clasificación se procese más rápidamente.

REFERENCIAS BIBLIOGRÁFICAS

- Alexiuk, M., P. C. Li, N. Pizzi, and W. Pedrycz, "Classification of Hail and Tornado Storm Cells Using Neural Networks," in 1999 IEEE Western Canada Conference and Exhibition, pp. 15–21.
- Anderson, M. E., L. D. Carey, W. A. Petersen, y K. R. Knupp. 2012. «C-band Dual-polarimetric Radar Signatures of Hail», julio. <http://www.nwas.org/ej/2011-EJ2/>.
- Aydin, K., T. A. Seliga, y V. Balaji. 1986. «Remote sensing of hail with a dual linear polarization radar». Journal of Climate and Applied Meteorology 25 (10): 1475–1484.
- Bartok, J., O. Habala, P. Bednar, M. Gazak, y L. Hluchy. 2010. «Data mining and integration for predicting significant meteorological phenomena». Procedia Computer Science 1 (1): 37–46.
- Bellini Saibene, Y. 2015. "Estimación de ocurrencia de granizo en superficie y daño en cultivos utilizando datos de radar meteorológico y técnicas de minería de datos". Tesis de Maestría. Octubre.
- ——, M. Volpacchio. 2015. "Clasificación de granizo en superficie usando técnicas de minería de datos y datos de radar meteorológico". 16º Argentine Symposium on Artificial Intelligence. ISSN: 1850-2784: 137-141. Rosario, Santa Fe, Argentina. 31 de Agosto.
- ——, J.M. Caldera, M. Volpacchio. 2015b. "HAR (HAIL-Archive): desarrollo de un sistema de información y base de datos sobre granizo en la región semiárida pampeana central". XII Congreso Argentino de Meteorología. CONGREMET XII. 26 al 29 de Mayo. Mar del Plata, Argentina.
- ——, S. Banchero, R.N. Mezher. 2014. "Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA". Anales del 6º Congreso Argentino de Agroinformática: 74-86. Septiembre. Argentina.
- ——, S. Banchero. 2016. "RadINTA: Facilitando el uso de la red de radares de INTA mediante el desarrollo de herramientas libres." 3º Jornadas de Tecnología de Información Geográfica del Sur Argentino. Bahía Blanca, Argentina. 24 al 26 de Agosto.
- Billet, John, Mark DeLisi, Brian G. Smith, y Cory Gates. 1997. «Use of Regression Techniques to Predict Hail Size and the Probability of Large Hail». Weather and Forecasting 12 (1): 154-64.
- Bonelli, P., P. Marcacci, E. Bertolotti, E. Collino, y G. Stella. 2011. «Nowcasting and assessing thunderstorm risk on the Lombardy region (Italy)». Atmospheric Research 100 (4): 503-10. doi:10.1016/j.atmosres.2010.08.009.
- Boodoo, S., D. Hudak, M. Leduc, A. V. Ryzhkov, N. Donaldson, y D. Hassan. 2009. «Hail detection with a C-band dual polarization radar in the Canadian Great Lakes region.» 34th Conference on Radar Meteorology. <http://ams.confex.com/ams/pdffiles/156032.pdf>.
- Brandes, E. A., y A. V. Ryzhkov. 2004. «Hail detection with polarimetric radar». En . Vol. 5. http://ams.confex.com/ams/11aram22sls/techprogram/paper_82033.htm.
- Bringi, V. N., y V. Chandrasekar. 2001. "Polarimetric Doppler weather radar: Principles and applications." Cambridge Univ Pr. <http://books.google.com/books?hl=es&lr=&id=4r7e6assEHWc&oi=fnd&pg=PP1&dq=Polarimetric+Doppler+weather+radar:+principles+and+application&ots=JlmyM5yYR&sig=R3hLvXP6MEbh7w75U4aKl-buFbLg>.
- Bringi, V. N., J. Vivekanandan, y J. D. Tuttle. 1986. «Multiparameter Radar Measurements in Colorado Convective Storms. Part II: Hail Detection Studies». Journal of the Atmospheric Sciences 43 (22): 2564-77.
- Bustos, Carlos, y Hugo Videla. 1982. «Modelo estadístico de predicción de tormentas a corto plazo para la provincia de Mendoza». En Anales del XI Congreso Argentino de Meteorología. Catuogno, GA. <http://www.congremet.prm-arg.org/upload/bustoscarlos2.pdf>.
- Casellas, Karina, Gabriel Parellada, Lucia Longo, José Portillo, Patricio Calonge, y Estela Cristeche. 2012. «Memoria de La XXX Jornada de Perspectivas Agropecuarias "Gestión del riesgo agropecuario: herramientas para mitigar y transferir el riesgo climático"». Ciudad Autónoma de Buenos Aires.
- Ceperuelo Mallafré, Manuel. 2008. «Identificación y caracterización del granizo mediante el radar meteorológico. Modelos de predicción del ciclo de vida de las células convectivas.» Tesis Doctoral.
- Changnon, David, y Stanley A. Changnon. 1997. «Surrogate data to estimate crop-hail loss». Journal of applied meteorology 36 (9): 1202–1210.
- Changnon Jr, Stanley A. 1971. «Hailfall characteristics related to crop damage». Journal of Applied Meteorology 10 (2): 270–274.
- Collino, Elena, Paolo Bonelli, y Luciano Gilli. 2009. «ST-AR (STorm-ARchive): A project developed to assess the ground

- effects of severe convective storms in the Po Valley». *Atmospheric Research* 93 (1–3): 483–89. doi:10.1016/j.atmosres.2008.10.021.
- Ferreira, Candida. 2001. «Gene expression programming: a new adaptive algorithm for solving problems». arXiv preprint cs/0102027. <http://arxiv.org/abs/cs/0102027>.
 - ——. 2006a. «Designing neural networks using gene expression programming». En *Applied Soft Computing Technologies: The Challenge of Complexity*, 517–535. Springer. http://link.springer.com/chapter/10.1007/3-540-31662-0_40.
 - ——. 2006b. *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (Studies in Computational Intelligence)*. Springer-Verlag New York, Inc., Secaucus, NJ.
 - ——. 2013a. «Getting Started with Classification». Tutorial. GeneXproTools Tutorials – A Gepsoft Web Resource. octubre 9. <http://www.gepsoft.com/tutorials/GettingStartedWithClassification.htm>.
 - ——. 2013b. «Getting Started with Logistic Regression». Tutorial. GeneXproTools Tutorials – A Gepsoft Web Resource. octubre 11. <http://www.gepsoft.com/tutorials/GettingStartedWithLogisticRegression.htm>.
 - Ferreyra, Adriana. 2011. «Determinación de zona afectada por granizo en el partido de Chacabuco (Pcia. de Buenos Aires) el día 20 de noviembre de 2011, mediante imágenes satelitales.» Pergamino, Buenos Aires: INTA.
 - Foote, G. B., T. W. Krauss, and V. Makitov, 2005. “Hail metrics using conventional radar,” in Proc., 16th Conference on Planned and Inadvertent Weather Modification.
 - Gagne, D. J., A. McGovern, and J. Brotzge, “Classification of convective areas using decision trees,” *Journal of Atmospheric and Oceanic Technology*, vol. 26, no. 7, pp. 1341–1353, 2009.
 - Gematronik. 2007. «Instruction Manual. Rainbow 5. Products and Algorithms.» En *Instruction Manual. Rainbow 5*.
 - Heinselman, P. L., & Ryzhkov, A. V. 2006. “Validation of polarimetric hail detection”. *Weather and forecasting*, 21(5), 839–850.
 - Hohl, R., H. H. Schiesser, y D. Aller. 2002. «Hailfall: the relationship between radar-derived hail kinetic energy and hail damage to buildings». *Atmospheric Research* 63 (3): 177–207.
 - Holleman, I., Hail detection using single-polarization radar. Ministerie van Verkeer en Waterstaat, Koninklijk Nederlands Meteorologisch Instituut, 2001.
 - Holleman, I., H. R. Wessels, J. R. Onvlee, y S. J. Barlag. 2000. «Development of a hail-detection-product». *Physics and Chemistry of the Earth Part B Hydrology Oceans and Atmosphere* 25: 1293–1297.
 - Hosmer, D. W., y S. Lemeshow. 2000. *Applied logistic regression*. Second. Vol. 354. Wiley-Interscience.
 - Keenan, T. 2003. «Hydrometeor classification with a C-band polarimetric radar». *Australian Meteorological Magazine* 52 (1): 23–31.
 - Li, Pak C., Nicolino Pizzi, Witold Pedrycz, Dave Westmore, y Rodrigo Vivanco. 2000. «Severe storm cell classification using derived products optimized by genetic algorithms». En *Electrical and Computer Engineering, 2000 Canadian Conference on*, 1:445–448. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=849748.
 - López, L., y J. L. Sánchez. 2009. «Discriminant methods for radar detection of hail». *Atmospheric Research* 93 (1): 358–368.
 - ——, E. García-Ortega, y J. L. Sánchez. 2007. «A short-term forecast model for hail». *Atmospheric Research, European Conference on Severe Storms 2004 ECSS 2004 European Conference on Severe Storms 2004*, 83 (2–4): 176–84. doi:10.1016/j.atmosres.2005.10.014.
 - Marzano, F. S., D. Scarpani, and G. Vulpiani, “Supervised Fuzzy-Logic Classification of Hydrometeors Using C-Band Weather Radars,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 11, pp. 3784–3799, Nov. 2007.
 - May, P. T., y T. D. Keenan. 2003. «Four-dimensional microphysical data from Darwin». En *13th ARM Science Team Meeting Proceedings*. https://education.arm.gov/publications/proceedings/conf13/extended_abs/may2-pt.pdf.
 - Mezher, R. N, V Barros, y P. A Mercuri. 2010. «Climatología de eventos de granizo en la Región Pampeana». www.inta.gov.ar/cya.
 - ——, N. N Gattinoni, y P. A Mercuri. 2008. «Variabilidad estacional de la ocurrencia de granizo en el centro, este y noreste de argentina». Reunión Argentina de Agrometeorología. 12. 2008 10 o8-10, 8 al 10 de octubre de 2008. San Salvador de Jujuy. AR.
 - ——, P. A Mercuri. 2008. «Análisis espacial y temporal de la ocurrencia de eventos de granizo sobre Argentina». XV Congresso Brasileiro de Meterologia, Agosto.
 - ——. P. A Mercuri. 2010b. «Uso de la red de radares de INTA para la detección de granizo». XIII Reunión Argentina y VI Latinoamericana de Agrometeorología., Octubre.
 - ——, P. A Mercuri, y N. N Gattinoni. 2008. «Distribución espacio-temporal del granizo en Argentina». Reunión Argentina de Agrometeorología. 12. 2008 10 o8-10, 8 al 10 de octubre de 2008. San Salvador de Jujuy. AR.
 - ——, M. Doyle, y V. Barros. 2012. «Climatology of hail in

- Argentina». *Atmospheric Research* 114–115 (octubre): 70-82. doi:10.1016/j.atmosres.2012.05.020.
- ——, Luciano Vidal, y Paola Salio. 2011. «Hailstorms Analysis using Polarimetric Weather Radars and Microwave Sensors in Argentina». 6th European Conference on Severe Storms (ECSS 2011), Agosto.
 - ——, Santiago Banchero, y Yanina Noemí Bellini Saibene. 2012b. «Identificación de granizo con la utilización de variables polarimétricas de los radares de Paraná y Anguil, el radar de Pergamino y daño en cultivos.» En Congreso Argentino de Meteorología. 11., 28 de mayo al 1 de junio de 2012. Mendoza. AR. <http://orton.catie.ac.cr/cgi-bin/wxis.exe/?IsisScript=INTA2.xis&method=post&formato=2&cantidad=1&expresion=mfn=067676>.
 - Ortega, Kiel L., Travis M. Smith, Kevin L. Manross, Angelyn G. Kolodziej, Kevin A. Scharfenberg, Arthur Witt, y Jonathan J. Gourley. 2009. «The Severe Hazards Analysis and Verification Experiment». *Bulletin of the American Meteorological Society* 90 (10): 1519-30. doi:10.1175/2009BAMS2815.1.
 - Ponce de Leon, E. “Granizo.” Servicio Meteorológico Nacional, 1985.
 - Prieto, R., R. Herrera, P. Doussel, L. Gimeno, P. Ribera, R. García, y E. Hernández. 2009. «Looking for Periodicities in the Hail Intensity in the Andes Region». *ATMOSFERA* 14 (2). <http://revistas.unam.mx/index.php/atm/article/view/8477>
 - Rinehart, R. E. 1997. “Radar for meteorologists.” Dept. of Atmospheric Sciences, Center for Aerospace Sciences, University of North Dakota.
 - Ryzhkov, A., D. Zrnic, J. Krause, M. Kumjian, y S. Ganson. 2010. «Discrimination Between Large And Small Hail. Final Report». NSSL.
 - Saltikoff, Elena, Jari-Petteri Tuovinen, Janne Kotro, Timo Kuitunen, y Harri Hohti. 2010. «A Climatological Comparison of Radar and Ground Observations of Hail in Finland». *Journal of Applied Meteorology and Climatology* 49 (1): 101-14. doi:10.1175/2009JAMC2116.1.
 - Sánchez, J. L., L. López, E. García-Ortega, y B. Gil. 2013. «Nowcasting of kinetic energy of hail precipitation using radar». *Atmospheric Research* 123: 48–60.
 - Tabary, Pierre, B. Fradon, A. J. Illingworth, y G. Vulpiani. 2009. «Hail detection and quantification with a C-band polarimetric radar: Challenges and promises». En 34th Conference on Radar Meteorology, 5–9. https://ams.confex.com/ams/34Radar/techprogram/paper_155530.htm.
 - Tuovinen, Jari-Petteri, Ari-Juhani Punkka, Jenni Rauhala, Harri Hohti, y David M. Schultz. 2009. «Climatology of Severe Hail in Finland: 1930–2006». *Monthly Weather Review* 137 (7): 2238-49. doi:10.1175/2008MWR2707.1.
 - Zrnic, D. 2012. «Dual Polarization for Weather Observations, Clasification of Hydrometeors and Measurement of Rain.» presentado en Simposio de Radarización y Sistema de Alertas Hidrometeorológicos del Cono Sur, Buenos Aires, Argentina, septiembre 11. <http://pampero.cima.fcen.uba.ar/SimposioAlertas/>.
-

Clasificación de daño en cultivos por granizo utilizando técnicas de Data Mining

7.

Bellini Saibene, Yanina¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
bellini.yanina@inta.gob.ar

RESUMEN

Es conocido que el granizo puede ocasionar serios daños a los cultivos, ocasionando importantes pérdidas económicas, por lo que su estudio es de interés para el sector agropecuario y agroindustrial. Con el objetivo de estimar el posible daño ocasionado a los cultivos se utilizó Gene Expression Programming (GEP) Regresión Logística usando datos del radar polarimétrico de banda C (INTA Anguil, La Pampa) desde Marzo de 2009 a Marzo de 2013.

Para determinar el daño se usaron variables de radar y de cultivo para clasificarlo en cuatro problemas binarios de acuerdo a los porcentajes de destrucción: leve (1-25%)/no leve, moderado (25-50%)/no moderado, severo (50-75%)/no severo y grave (75-100%)/no grave. A pesar que los modelos obtenidos no son robustos en diferenciar estas clases debido a: la simplificación del problema, la baja disponibilidad de datos y la subespecificación de las variables de alto impacto. A pesar de esto, la correlación encontrada, sugiere que estas herramientas se pueden usar para análisis futuros en un conjunto de datos más grande y completo. Análisis adicionales reduciendo los niveles a tres: sin daño, <50% y >50% aumentan la correlación, reforzando la idea que estas herramientas son adecuadas para generar modelos sobre el daño en cultivos.

Palabras clave: Gene Expression Programming, riesgo agropecuario, eventos severos, agrometeorología, teledetección

ABSTRACT

It is known that hail can cause serious damages to crops, causing significant economic losses. Their

study is useful for the agribusiness and insurance industry. In order to calculate the possible crops damage Gene Expression Programming (GEP) Logistic Regression was used with data from polarimetric C-band radar (INTA Anguil La Pampa) from March 2009 to March 2013.

The radar and crop variables were used to classify into four binary problems according the percentages of destruction: slight (1-25%)/no slight, moderate (25-50%)/no moderate, severe (50-75%)/no severe and grave (75-100%)/no grave. Although the models fail to differentiate these four classes due to the simplification of the problem, the lower data availability and a under specification of the high impact variable, a correlation found suggested that these tools can be used for future analysis on larger and more complete dataset. Further analysis reducing the levels to three categories: without damage, <50% and >50% increased the correlation, reinforcing the idea that these tools are useful for generating models that classified crops losses.

Key words: Gene Expression Programming, Agricultural risk, severe weather, agrometeorology, remote sensing

INTRODUCCIÓN

Es bien conocido que el granizo puede ocasionar serios daños a los cultivos (Katz *et. al.* 1981). Por esta razón se lo considera un riesgo agroclimático, que se define como la probabilidad de afectación del rendimiento o la calidad de los cultivos por efecto de un fenómeno climático adverso (Almada *et al.* 2005). Las pérdidas en cultivos de verano (Soja, Maíz, Girasol) en los últimos cinco años, para la provincia de La Pampa, promedian las 10.000 hectáreas.

Conocer la frecuencia de eventos climáticos severos, su comportamiento promedio y sus condiciones extremas resultan primordiales para diseñar evaluaciones de riesgo adecuadas como insumo a las tasas de seguros, para evaluar la variabilidad climática y contar con opciones y alternativas de manejo agropecuario (Katz *et. al.* 1981, Bellini Saibene, Y. 2015).

Generar funciones que relacionen el daño realizado por granizo en los cultivos es complejo porque esta relación depende de varios factores entre los cuales se pueden mencionar el tipo de cultivo, su estado fenológico, el tipo de tormenta (por ejemplo, si el granizo estuvo acompañado por fuertes vientos causa un daño mayor). (Katz *et. al.* 1981).

A esto se debe sumar que, en general al inicio de estos estudios, no se cuenta con registros de caída de granizo y daño ocasionado a los cultivos sistematizados, históricos y disponibles abiertamente (Katz *et. al.* 1981; Changnon, *et. al.* 1997), realidad que se repite para Argentina y La Pampa (Bellini Saibene *et. al.* 2015).

Los primeros trabajos en relacionar el granizo con el daño en cultivos utilizaron redes de granizómetros junto a información de compañías de seguros y redes de observadores a campo, determinando funciones y variables que permiten predecir el nivel de daño en cultivos por granizo. Estas variables son: a) energía cinética (E) (Sánchez, J. L. *et. al.* 2013, Katz *et. al.* 1981 Changnon *et. al.* 1971, Morgan *et. al.* 1976), b) impulso (Katz *et. al.* 1981), c) masa (Katz *et. al.* 1981), d) número de granizos con diámetro mayor a un determinado tamaño (ej: 6.4 mm Changnon *et. al.* 1971, 12.7 mm Katz *et. al.* 1981 y Morgan *et. al.* 1976), e) tamaño del granizo (Sánchez, J. L. *et. al.* 2013, Sánchez, J. L. *et. al.* 1996) f) viento (Katz *et. al.* 1981, Morgan *et. al.* 1976) g) cultivo (Changnon *et. al.* 1971, Morgan *et. al.* 1976) y h) estado fenológico del cultivo (Katz *et. al.* 1981, Changnon *et. al.* 1971, Morgan *et. al.* 1976).

Como las redes de granizómetros son muy costosas, se realizaron estudios para relacionar las variables que caracterizan la caída del granizo por medio de la reflectividad medida por un radar meteorológico (Waldvogel *et. al.* 1978, Seino, H. 1980). Los resultados para banda S y C determinaron una relación empírica, de la forma (Waldvogel *et. al.* 1978):

$$E = a \times 10^{-6} \times 10^{b \times Z} \quad (1)$$

Para la cual se han definido diversos valores de a y b de acuerdo a la región de estudio (ej.: Hohl *et. al.* 2002b, Makitov, V. 2007, Foote, G. *et. al.* 2005, Waldvogel *et. al.* 1978, Schiesser, H. H. 1990), el más utilizado es $a=5$ y $b=0,084$ determinado por Waldvogel *et. al.* 1978. Para utilizar la relación (2), se generaron diversas técnicas para decidir en qué zonas tiene sentido calcular E (Makitov, V. 2007): a) “Cutting method”: se fijan umbrales de Z (ej.: de 49 a 65 dBZ (Visser, P. *et. al.* 2000, Makitov, V. 2007) y de la altura de los ecos de 45 dBZ (H45) (las cuales dependen de la región de estudio (Makitov, V. 2007, Waldvogel *et. al.* 1978, Schiesser, H. H. 1990), b) Función de peso de Z: donde a (2) se le agrega un multiplicando $W(Z)$, el cual se define como (Ceperuelo Mallafré, M. 2008, Williamson, S. P. 2006):

$$W(Z) = \begin{cases} 0 & \text{para } Z \leq Z_{min} \\ \frac{Z - Z_{min}}{Z_{max} - Z_{min}} & \text{para } Z_{min} < Z < Z_{max} \\ 1 & \text{para } Z \geq Z_{max} \end{cases} \quad (2)$$

Los valores Z_{min} y Z_{max} también dependen de la región de estudio, algunos de los valores utilizados son: 40-50, 45-55, 50-60 (Ceperuelo Mallafré, M. 2008, Skripnikova, K. *et. al.* 2014, Williamson, S. P. 2006, Witt, A. *et. al.* 1998). Finalmente, para muchas aplicaciones, el valor de E se debe integrar en el tiempo, durante la duración de la tormenta y para cada pixel del radar (Hohl, R *et. al.* 2001). La forma de relación más común es la definición de una regresión logística de la variable E, para cada cultivo y estado fenológico del mismo (Katz *et. al.* 1981, Schiesser, H. H. 1990, Hohl, R *et. al.* 2002).

En nuestro país se estudió la relación de la probabilidad de granizo generada por el algoritmo ZHAIL de los radares de INTA Paraná e INTA Pergamino con el porcentaje de daño en cultivos y no se encontró una relación clara (Mezher, R. N. *et. al.* 2010), generando muchas falsas alarmas (Mezher, R. N. *et. al.* 2010, Mezher, R. N. *et. al.* 2011).

En cuanto a las variables polarimétricas de banda C, Mezher, R. N. *et. al.* 2011 analizó la relación entre el parámetro Hail Differential Reflectivity (HDR), calculado con datos del radar de Paraná y el daño en cultivos, encontrando una buena relación, donde el mayor daño en cultivo coincide con el núcleo de 50 HDR, que puede indicar granizo de tamaño medio (Mezher, R. N. *et. al.* 2011).

El parámetro HDR fue desarrollado por Aydin, K. *et. al.* 1986, para distinguir granizo de lluvia utilizando-

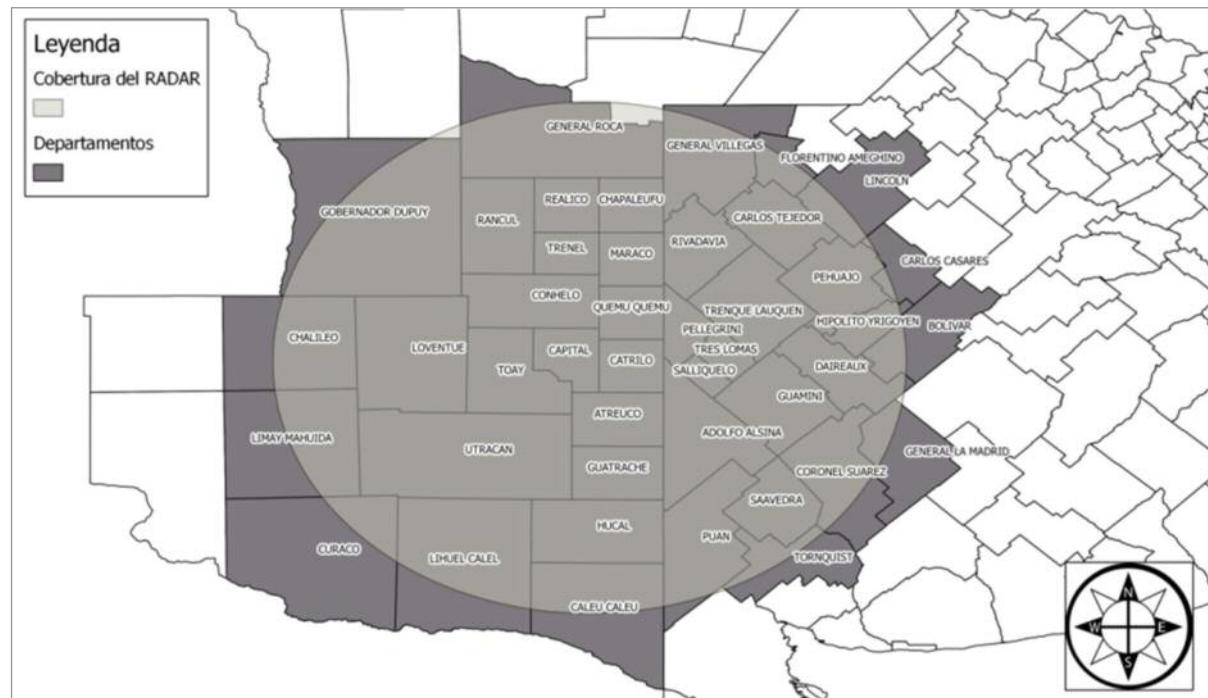


Figura 1. Mapa del área de influencia de 240 km de radio del radar de la EEA Anguil (Bellini Saibene, y. 2015).

do Z y ZDR, por medio de la siguiente relación (Mezher, R. N. et al. 2011):

$$\text{HDR} = Z - f(ZDR) \text{ donde:}$$

$$f(ZDR) = \begin{cases} 21 \text{ dB} & \text{para } ZDR \leq 0 \text{ dB} \\ 19 ZDR + 27 \text{ dB} & \text{para } 0 < ZDR \leq 1.74 \text{ dB} \\ 60 \text{ dB} & \text{para } ZDR > 1.74 \text{ dB} \end{cases} \quad (3)$$

HDR es un buen indicador de presencia de granizo, para las elevaciones más bajas y que se encuentran por debajo del nivel de congelamiento. Mayor valor de HDR indica mayor probabilidad que las medidas de Z no son debido a gotas de lluvia (Mezher, R. N. et al. 2011).

El objetivo de este trabajo es utilizar métodos supervisados de aprendizaje automático con información del RADAR meteorológico de INTA en La Pampa para analizar el comportamiento y poder de clasificación de las variables de doble polarización para identificar el daño realizado por el granizo en los cultivos presentes en el área de influencia del RADAR.

METODOLOGÍA

El área de estudio corresponde a la zona de influencia del radar situado en la Estación Experimental Agropecuaria (EEA) Anguil en el rango

de los 240 km de radio con centro en el sensor remoto (figura 1). El período temporal se fijó de Marzo de 2009 (puesta en funcionamiento del radar (Bellini Saibene, Y. et. al. 2010)) a Marzo de 2013 donde se registraron 8 fechas con tormentas que generaron daño de granizo en cultivos en la zona bajo estudio y sobre las cuales se obtuvieron datos de campo.

Se tomaron dos tipos de datos: 1) los datos de campo que detallan el daño en cultivos y 2) los registrados por el radar.

Datos de campo

A partir del análisis de antecedentes se determinó que las variables de terreno necesarias para el armado de los modelos son:

- 1) *fecha del evento* (formato: dd-mm-aaaa) y hora del evento (hora de inicio y fin en formato hh:mm),
- 2) *ocurrencia de granizo* (si, no),
- 3) *posición geográfica* (en coordenadas geográficas, latitud y longitud, WGS84),
- 4) *cultivo* (en caso que corresponda a una localización no urbana se registra el uso del suelo, si es un cultivo, se registra su nombre vulgar, por ejemplo: Girasol, Soja, Maíz, Trigo, etc.),
- 5) *estado fenológico* (dato mínimo: Vegetativo,



Figura 2. Esquema de relevamiento de datos de campo (Bellini Saibene, Y. 2015)

Reproductivo, Madurez, de ser posible nombre correspondientes de acuerdo a escalas fenológicas definidas para cada cultivo: Trigo: escala de Zadoks, Chang y Konzak, 1974; Girasol: escala de Schneiter y Miller, 1974; Soja: escala de Fehr y Caviness, 1977; Maíz: escala de Ritchie y Hanway, 1982),

6) daño (Sin Daño, Leve, Moderado, Severo y Grave),

7) área afectada (en hectáreas).

En la figura 2 se resume el esquema de recolección de información. Todos estos datos, juntos con los recolectados y procesados del radar, se gestionaron utilizando el sistema HAR (Bellini Saibene, Y. et. al. 2015)

Datos del radar

Para este trabajo se utilizaron las variables Factor de reflectividad (Z) (Rinehart 1997), Differential Reflectivity-Reflectividad diferencial (ZDR) (Rinehart 1997) y Correlation Coefficient-

Coeficiente de correlación co-polar (ρ_{HV} , RhoHV, CC) (Rinehart 1997), las cuales se toman 144 veces en un día (cada 10 minutos) y para 12 ángulos de elevación ($0,5^\circ$; $0,9^\circ$; $1,3^\circ$; $1,9^\circ$; $2,3^\circ$; $3,0^\circ$; $3,5^\circ$; $5,0^\circ$; $6,9^\circ$; $9,1^\circ$; $11,8^\circ$ y $15,1^\circ$) con una unidad de muestreo de 1 km² y 1° . Estas variables se procesaron con el software libre RadINTA (Bellini Saibene, Y. et. al. 2014; Bellini Saibene, Y. et. al. 2016).

Armado de los conjuntos de datos

Los datos de campo tienen una escala temporal diaria, por lo tanto las variables del radar fueron resumidas en valores de 24 horas por medio de diferentes cálculos. Estos resúmenes permiten utilizar juntos los datos de campo y los de radar al llevarlos a la misma escala temporal (Holleman et al. 2000, Collino et. al. 2009). Los cálculos realizados por medio de consultas Transact SQL (Rose, R. 2000) son: máximo (MAX) mínimo (MIN), promedio (AVG) y total (SUM) de Z, RhoHV y ZDR. Estos cálculos se hacen sobre los valores de todas las elevaciones y sobre los valores de la primera elevación solamente.

Esta configuración de los datos permitirá analizar si es factible obtener una buena predicción del target utilizando una sola elevación, haciendo que el modelo sea más rápido y sencillo al necesitar menos datos. Se selecciona la primera elevación porque al ser la más cercana a la superficie es la que mejor representa lo que puede precipitar a nivel del suelo (R. Hohl et. al. 2002, Mezher, et. al. 2012, Mezher, et. al. 2011, Schiesser 1990, Aydin et. al 1986, Ryzhkov A. V. et al. 2005).

En el caso de las variables HDR y E estos cálculos se aplicaron al resultado de usar la fórmula 3 y 4 respectivamente.

$$E = \left(\sum_{i=\text{base}}^{i=\text{tops}} 5 \times 10^{-6} \times 10^{0.084 \times Z_i} \times W(Z_i) \right) \cdot t \quad (4)$$

Dónde:

$i = 1 \text{ a } 12$ elevaciones

$t = 600$ segundos¹

$$W(Z) = \begin{cases} 0 & \text{para } Z \leq 40 \\ \frac{Z - 40}{40 - 55} & \text{para } 40 < Z < 55 \\ 1 & \text{para } Z \geq 55 \end{cases}$$

También se generaron una serie de variables que intentan capturar la presencia y altura de ecos fuertes. Se calcularon utilizando las sentencia cuenta (COUNT) y CASE de T-SQL sobre la variable Z.

Debido a la forma de la toma de datos y de las tormentas, la aparición del valor missing (perdido) en las variables del radar, ocurre cuando no hay tormenta. Esta particularidad hace imposible reemplazar ese valor por algún otro, ya que en ese lugar y momento, no había que medir. Para representar esta particularidad se representó el valor perdido con el número “-99.00”². Los cálculos de máximo, mínimo, promedio y total de Z, ZDR y RhoHV se calcularon con todos los casos por un lado y descartando los casos que tuvieran valores perdidos por otro.

¹ Es la cantidad de segundos contenidos en los 10 minutos que transcurren entre cada toma de datos del radar.

² El -99.00 es un valor estándar para representar el “Sin Dato” o el “NULL” en los sistemas de información agrometeorológica de INTA. Cumple con el requisito de estar muy alejado de los valores válidos de las diferentes medidas agroclimáticas, incluyendo los datos de radar.

De esta manera ninguna variable presenta valores perdidos y por ende se puede usar con las técnicas de minería de datos (DM) seleccionadas. Aquellas calculadas sin los “-99.00” también permiten comparar su comportamiento con la bibliografía aportando al objetivo de contribuir con la caracterización de las variables de doble polarización de banda C, para la estimación del daño en cultivos en la región de estudio. Las variables calculadas con el “-99.00” permite reflejar en los datos el momento y lugar donde la tormenta no registró valores y estudiar cómo influye esta particularidad en el rendimiento de los modelos.

En total se generaron 82 variables, para un detalle de las mismas se puede consultar Bellini Saibene, Y. 2015.

La variable objetivo (target) en este trabajo es el daño, el cual está representado en los datos como una variable multiclase con los valores: leve (1-25%), moderado (25-50%), severo (50-75%) y grave (75-100%). Posteriormente a una serie de análisis estadísticos se generó un nuevo target reagrupando las clases: Sin Daño (0% de daño), Menos50 (1% a 49% de daño) y Mas50 (50% a 100% de daño).

Para utilizar los mismos métodos de DM que en el caso de la estimación de ocurrencia de granizo, Gene Expression Programming con Regresión Logística (Ferreira, C. 2001, Hosmer y Lemeshow 2000), el primer target se trabajó con cuatro problemas binarios de acuerdo a los porcentajes de destrucción: leve/no leve, moderado/no moderado, severo/no severo y grave/no grave y la segunda opción a tres problemas binarios: Sin daño/ Con daño, <50%/Resto (Sin daño y >50%), >50%/resto (<50% y Sin Daño)

Además, las 82 variables se agruparon en diferentes conjuntos de datos de acuerdo al objetivo de análisis y a los dos targets buscados.

Los primeros conjuntos de datos se generaron para analizar el poder de clasificación y el comportamiento de las variables de doble polarización del radar para clasificar el daño realizado por el granizo en los cultivos. Como importa conocer el aporte de algunas variables relacionadas con los cultivos mencionadas en los antecedentes y que se pudieron relevar se generaron 12 data sets. En las figuras 3 y 4 y las tablas 1 y 2 se ve como queda configurada la muestra de casos con respecto del target en todos los data set.

Tabla 1. Características de los datos contenidos en los dataset para modelar el target Daño (tres clases).

				Target: Daño		
				Mas50	Menos50	Sin Daño
12 Elevac.	Sin Cultivo	Volúmenes Completos	H _{DR}	21 (1,43%)	245 (16,73%)	1.198 (81.8%)
12 Elevac.	Sin Cultivo	Volúmenes Completos + Incompletos	H _{DR}	97 (4.6%)	512 (24.27%)	1.501 (71.1%)

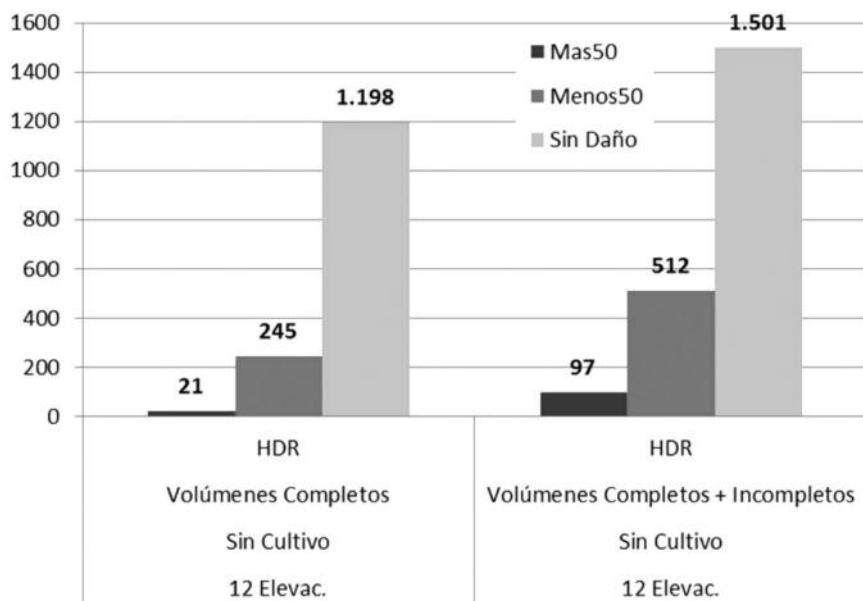


Figura 3. Distribución de los casos en los dos dataset para modelar el target Daño de tres clases.

Tabla 2. Características de los datos contenidos en los dataset para modelar el target Daño (cinco clases).

				Target: Daño				
				Leve	Moderado	Severo	Grave	Sin Daño
1 Elevac.	Con Cultivo	Volúmenes Completos		209 (74.1%)	25 (8.9%)	13 (4.6%)	8 (2.8%)	27 (9.6%)
1 Elevac.	Sin Cultivo	Volúmenes Completos		214 (15.9%)	31 (2.3%)	13 (1%)	8 (0.6%)	1.077 (80.2%)
1 Elevac.	Sin Cultivo	Volúmenes Completos	H _{DR}	214 (15.9%)	31 (2.3%)	13 (1%)	8 (0.6%)	1.077 (80.2%)
12 Elevac.	Con Cultivo	Volúmenes Completos		209 (74.1%)	25 (8.9%)	13 (4.6%)	8 (2.8%)	27 (9.6%)
12 Elevac.	Sin Cultivo	Volúmenes Completos		214 (14.6%)	31 (2.1%)	13 (0.9%)	8 (0.5%)	1.198 (81.8%)
12 Elevac.	Sin Cultivo	Volúmenes Completos	H _{DR}	214 (14.6%)	31 (2.1%)	13 (0.9%)	8 (0.5%)	1.198 (81.8%)
1 Elevac.	Con Cultivo	Volúmenes Completos + Incompletos		425 (63.2%)	60 (8.9%)	45 (6.7%)	49 (7.3%)	94 (14.0%)
1 Elevac.	Sin Cultivo	Volúmenes Completos + Incompletos		431 (21.9%)	67 (3.4%)	45 (2.3%)	51 (2.6%)	1.377 (69.9%)
12 Elevac.	Con Cultivo	Volúmenes Completos + Incompletos		436 (63.1%)	63 (9.1%)	46 (6.7%)	49 (7.1%)	97 (14.0%)
12 Elevac.	Sin Cultivo	Volúmenes Completos + Incompletos		442 (20.9%)	70 (3.3%)	46 (2.2%)	51 (2.4%)	1.502 (71.2%)

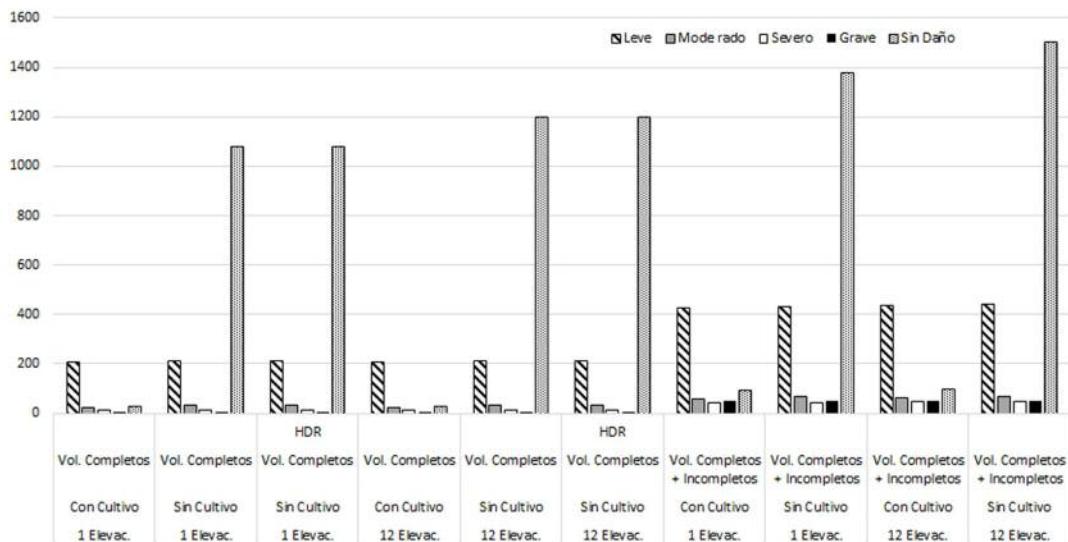


Figura 4. Distribución de los casos en los diez dataset para modelar el target Daño de cinco clases.

Análisis estadísticos

Se realizaron análisis ANOVA de un factor para determinar la importancia de cada variable en la separación de las clases del target Daño y determinar si las diferencias de las variables del radar entre las fechas completas de las incompletas es significativa. Para ejecutar estas tareas se usó el software R, su entorno RStudio y el paquete RCommander.

Ejecución de las herramientas de DataMining

Para ambas opciones de clases del target Daño, se utilizó como condición de parada llegar al valor 1000 en el “Max Fitness” o llegar a la generación 30.000, lo que ocurra primero.

Medidas de rendimiento de los modelos

Para presentar el rendimiento de los modelos obtenidos y compararlos se utilizan una serie de indicadores presentes en la literatura, basados en la matriz de confusión (tabla 4) de los datos de entrenamiento y testing utilizados en diferentes técnicas de DM (Alexiuk *et al.*, 1999; Alexiuk *et al.*, 1999b; Li P. C. *et al.* 2000; Marzban *et al.* 2001; Ramirez *et al.* 2001; Gagne *et al.* 2009; Peters *et al.* 2003; Zbigniew *et. al.* 2002; Tsaglidis *et al.* 2008; Zbigniew *et. al.*

2001) y en el ámbito de la predicción de fenómenos climáticos (Billet *et al.* 1997; Marzban *et al.* 2001; Aydin *et. al.* 1986; Ryzhkov A. V. *et al.* 2005; Holleman *et al.* 2000; Tabary P. *et al.* 2010; Lakshmanan V. *et. al.* 2009; Collino *et. al.* 2009; Xiang Li *et al.* 2004; Center NOAA, 2013). Algunos de los indicadores se comparten entre ambas temáticas; los más utilizados en clima, se concentran en el comportamiento del modelo con respecto de los verdaderos positivos, mientras que en DM también importan los verdaderos negativos.

RESULTADOS Y DISCUSIÓN

Características de los conjuntos de datos

Reflectividad

La tabla 5 presenta las estadísticas básicas por clase de daño de las variables de resumen de Z. Las cuatro variables presentan los menores valores registrados para la clase “Sin Daño”, como los casos sin daño pertenecen en su gran mayoría a casos sin granizo o de granizo muy leve, es lógico que los registros de Z para esta clase son los de menor valor (Rinehart 1997, Boettcher, 2013).

La clase “Grave” se diferencia de las demás porque sus valores mínimos son los más elevados,

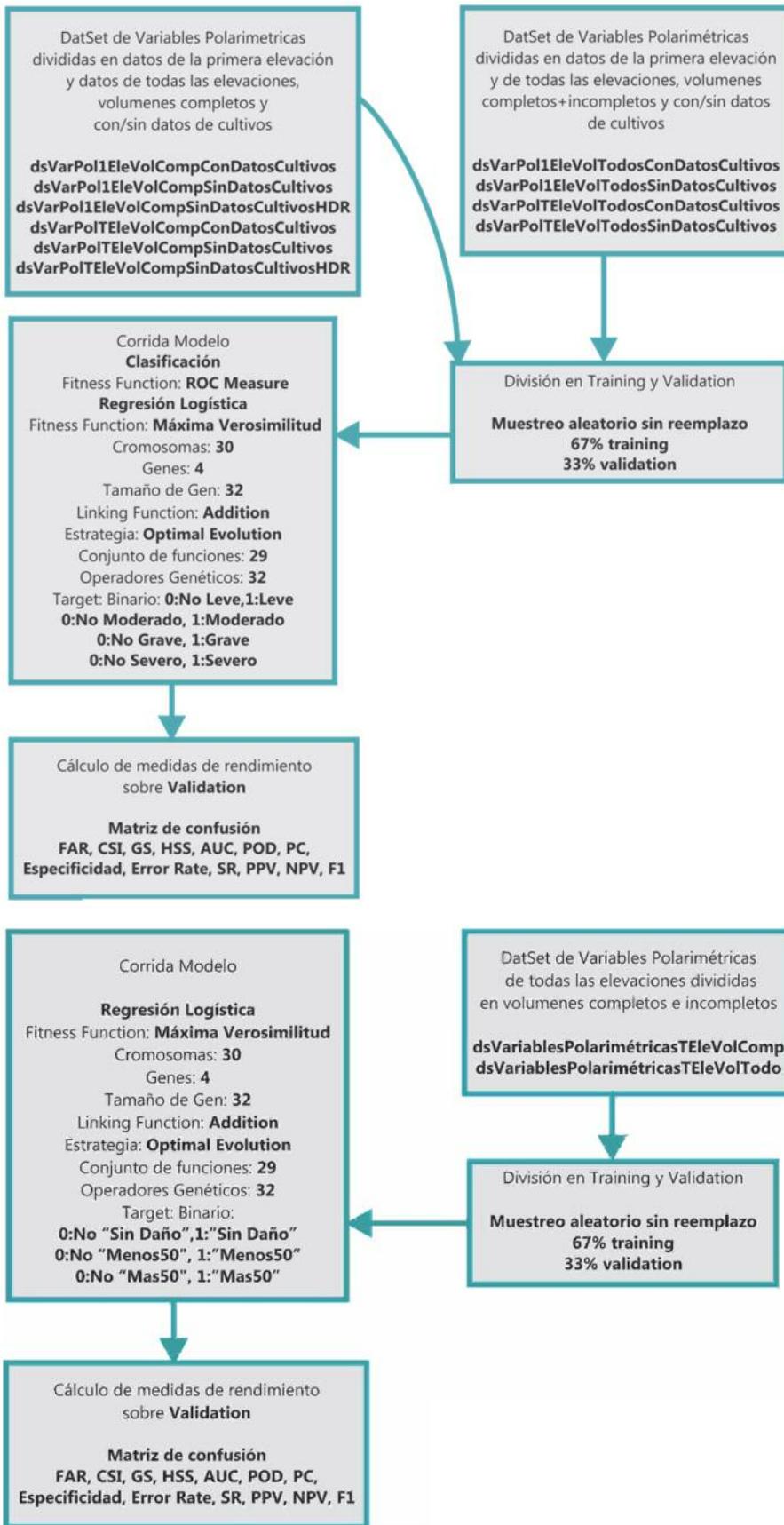


Tabla 4. Matriz de confusión y medidas de rendimiento disponibles (Adaptado de: Hosmer y Lemeshow 2000, Center NOAA 2013, León Aristizábal 2005).

		Matriz de Confusión (Contingencia)	
		Valores Observados	
		Si	No
Valores Predichos	Si	VP (Verdaderos Positivos) Identificados Correctamente	FP (Falso Positivo) Identificados incorrectamente
	No	FN (Falso Negativo) Incorrectamente rechazados	VN (Verdadero Negativo) Correctamente rechazados
Medidas de Rendimiento			
Nombre	Cálculo	Descripción	
FAR (False Alarm Ratio)	FP/(VP+FP)	(5)	Número de falsas alarmas positivas dividido por el número total de positivos predichos. Para un modelo perfecto FAR=0.
CSI (Critical Success Index), TS (Threat Score)	VP/(VP+FP+FN)	(6)	Total de verdaderos positivos dividido por el número total de las valores predichos SIN los VN por lo que no se ve afectado por el número de rechazos correctos.
GS (Gilbert Skill Score)	$(VP-CH)/(VP+FP+FN-CH)$ donde CH= $(VP+FP)(VP+FN)/n$	(7)	Se comporta similar al CSI, pero tiene en cuenta los verdaderos positivos obtenidos debido a la casualidad (CH).
HSS (Heidke Skill Score)	$(VP+VN-E)/(VP+FP+FN+VN-E)$ donde E= $[(VP+FP)(VP+FN)+(FP+VN)(FN+VN)]/n$	(8)	Rango (-1,1). Pronóstico al azar, HSS=0; pronóstico perfecto, HSS=1; pronóstico imperfecto HSS = -1. Valores negativos del HSS indican que los desaciertos dominan el análisis. Para un modelo aceptable debe ser positivo.
POD (Probability Of Detection), Sensibilidad, Recall, Exhaustividad	VP/(VP+FN)	(9)	Capacidad para identificar resultados positivos correctamente. Modelo perfecto POD=1.
PC (Percent Correct), Accuracy	(VP+VN)/(VP+VN+FP+FN)	(10)	Proporción de los resultados reales (VP y VN) en la población.
Especificidad	VN/(VN+FP)	(11)	Capacidad para identificar los resultados negativos.
Error rate	(FP+FN)/(VP+FP+FN+VN)	(12)	Proporción de casos identificados incorrectamente contra todos los casos.
Precision, PPV(Positive Predictive Value), SR (Success Ratio)	VP/(VP+FP) o 1-FAR	(13)	Proporción de VP contra todos los resultados positivos (VP Y FP). Lo contrario de FAR.
NPV (Negative Predictive Value)	VN/(VN+FN)	(14)	Proporción de VN contra todos los resultados negativos (VN Y FN).
AUC (Area Under the Curve)	Sensibilidad vs 1- especificidad	(15)	Representación gráfica de la sensibilidad frente a (1 – especificidad) para un clasificador binario según se varía el punto de corte.
Medida F1	$2 \cdot [(precision \cdot recall) / (precision + recall)]$	(16)	Promedio ponderado de precision y recall. El mejor valor de F1 es 1 y el peor es 0.

mientras que las clases “Leve”, “Moderado” y “Severo” se comportan de manera muy similar en las variables MxdBZ, MndBZ y AvdBZ; el análisis de ANOVA indica que no hay diferencias significativas entre estas tres clases con ninguna de estas tres variables (MxdBZ: $Pr(>F) = 0,121$, MndBZ: $Pr(>F) = 0,726$ y AvdBZ: $Pr(>F) = 0,57$), mientras que TotdBZ presenta diferencias significativas ($Pr(>F) = 0,00273$) para la clase “Severo”.

RhoHV

La variable MxRho no presenta diferencias entre las clases. MnRho presenta los mayores valores para la clase “Grave” (tabla 6). Un análisis ANOVA para cada variable entre las clases “Leve”, “Moderado” y “Severo” confirma que no hay diferencias significativas (MxRho: $Pr(>F) = 0,877$, MnRho: $Pr(>F) = 0,274$, AvRho: $Pr(>F) = 0,339$ y TotRho: $Pr(>F) = 0,941$).

ZDR

La clase “Grave” (mayor porcentaje de daño) presenta los valores de ZDR más cercanos a cero, sumado a valores de $Z > 50$ dBZ (tabla 4) concuerda con (Anderson *et al.* 2012) que sugiere la posibilidad

de observar esta anti correlación de Z y ZDR en banda C, ante la presencia de granizos grandes y en mucha cantidad, que dominan Z y aparecen como esferas perfectas. El resto de las clases presenta un promedio por encima de 4 dB, acorde al comportamiento de la banda C en presencia de granizo (Pierre Tabary *et al.* 2009, Anderson *et al.* 2012, Zrnic 2012, May *et. al.* 2003, Keenan 2003) (tabla 7). Estas tres variables tampoco presentan diferencias significativas entre las clases “Leve”, “Moderado” y “Severo” (MxZDR: $Pr(>F) = 0,381$, MnZDR: $Pr(>F) = 0,118$ y AvZDR: $Pr(>F) = 0,689$). La variable TotZDR presenta diferencia significativa ($Pr(>F) = 0,00624$) para la clase “Severo”.

HDR

El promedio de los valores máximos de HDR muestran una marcada diferencia de la clase “Sin Daño” con respecto del resto, coincidiendo con la bibliografía donde se marca que a mayores valores de HDR, mayor probabilidad de granizo y mayor tamaño de granizo (Aydin *et. al.* 1986; Ryzhkov A. *et al.* 2010). Los máximos valores de MxHDR están por debajo de los 50 dB que encontró (Mezher, R. N., *et. al.* 2011) asociado al daño “Grave” en el radar de

Tabla 5. Estadísticas básicas de las variables de resumen de Z en el dataset de todos los volúmenes, todas las elevaciones y sin datos de cultivos.

Variable	MxdBZ				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	53,9	51,6	54,3	29,2	49,4
Máximo	66,0	72,0	72,0	70,0	69,0
Mínimo	30,0	7,0	17,0	-1,5	23,5
Variable	MndBZ				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	2,9	0,1	0,4	-0,2	1,1
Máximo	12,5	14,0	12,5	17,0	13,0
Mínimo	-13,5	-19,5	-15,0	-27,5	-18,0
Variable	AvdBZ				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	17,1	13,8	14,6	9,2	14,9
Máximo	24,1	32,1	25,9	30,3	24,8
Mínimo	4,1	-6,0	0,8	-14,7	-5,0
Variable	TotDBZ				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	2.224,3	2.901,7	3.120,0	889,9	1.780,1
Máximo	8.085,5	13.443,5	13.052,5	15.053,5	5.705,5
Mínimo	572,5	-7.285,0	88,0	-25.415,0	-6.389,5

Tabla 6. Estadísticas básicas de las variables de resumen de RhoHV en el dataset de todos los volúmenes, todas las elevaciones y sin datos de cultivos.

Variable	MxRho				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	1,00	1,00	1,00	0,99	1,00
Máximo	1,00	1,00	1,00	1,00	1,00
Mínimo	1,00	1,00	1,00	0,43	1,00
Variable	MnRho				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	0,655	0,361	0,381	0,552	0,428
Máximo	0,925	0,957	0,976	1,000	0,886
Mínimo	0,024	0,008	0,020	0,000	0,024
Variable	AvRho				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	0,9481	0,9099	0,9128	0,9214	0,8909
Máximo	0,9920	0,9973	0,9955	1,0000	0,9907
Mínimo	0,6608	0,6222	0,6109	0,1820	0,6584
Variable	TotRho				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	174,17	298,60	290,56	185,57	286,38
Máximo	944,34	1.267,78	1.097,10	1.501,16	1.056,75
Mínimo	35,31	28,24	31,85	0,918	61,15

Tabla 7. Estadísticas básicas de las variables de resumen de ZDR en el dataset de todos los volúmenes, todas las elevaciones y sin datos de cultivos.

Variable	MxZDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	3,082	5,332	4,957	4,002	4,483
Máximo	12,000	12,000	12,000	12,000	11,921
Mínimo	0,819	-0,126	-0,126	-4,772	0,504
Variable	MnZDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	-3,856	-5,069	-4,800	-3,118	-4,478
Máximo	-1,543	-0,835	-0,992	5,071	-1,150
Mínimo	-8,000	-8,000	-8,000	-8,000	-7,921
Variable	AvZDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	-0,2023	0,2521	0,1677	0,4728	0,4153
Máximo	2,6227	4,9732	3,2044	9,9004	7,1355
Mínimo	-1,2052	-1,6749	-1,6579	-4,7720	-1,1770
Variable	TotZDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	110,52	301,02	317,97	282,94	786,70
Máximo	1.845,36	4.682,02	2.842,33	17.107,89	7.706,37
Mínimo	-198,68	-247,28	-209,06	-672,50	-181,14

Tabla 8. Estadísticas básicas de las variables de resumen de HDR en el dataset de todos los volúmenes, todas las elevaciones y sin datos de cultivos.

Variable	MxHDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	33,3	24,1	28,1	3,1	16,5
Máximo	42,0	41,5	42,5	38,5	42,0
Mínimo	9,0	-19,0	-11,0	-58,5	-22,5
Variable	MnHDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	-61,3	-55,9	-56,1	-52,4	-65,0
Máximo	-18,0	-12,5	-15,0	3,0	-45,5
Mínimo	-73,0	-78,0	-73,0	-86,0	-77,5
Variable	AvHDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	-15,601	-16,951	-12,637	-20,711	-27,461
Máximo	-2,822	3,720	0,059	3,656	0,160
Mínimo	-27,126	-60,995	-36,483	-66,078	-62,488
Variable	TotHDR				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	-11,3	-7,3	-6,4	-5,2	-26,8
Máximo	-175,0	479,9	8,2	474,9	19,2
Mínimo	-24.929,2	-57.078,7	-28.072,0	-111.540,8	-67.487,7

Tabla 9. Estadísticas básicas de las variables de resumen de E en el dataset de todos los volúmenes, todas las elevaciones y con datos de cultivos.

Variable	MxEWt				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	442,26	772,46	987,55	646,67	437,37
Máximo	2.931,2	10.613,1	8.905,6	9.565,9	4.891,8
Variable	AvEWt				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	13,43	19,92	20,43	15,98	9,96
Máximo	75,13	550,28	109,62	446,75	119,84
Variable	TotEWt				
Medida	Grave	Leve	Moderado	Sin Daño	Severo
Promedio	774,77	1.130,47	1.605,00	1.139,39	633,17
Máximo	6.635,33	12.659,61	11.400,67	12.955,88	5.675,55

INTA Paraná (tabla 8). El análisis ANOVA presenta a las variables AvHDR ($\text{Pr}(>F)=0.0364$), MxHDR ($\text{Pr}(>F)=0.0246$) y TotHDR ($\text{Pr}(>F)=1.32e-06$) con diferencias significativas entre las clases “Grave”, “Leve”, “Moderado” y “Severo”.

E

Esta variable es calculada con la función (4),

donde se aplica un peso que asigna el valor cero cuando la Z es menor a un umbral especificado; por esta razón la variable MnEWt y los mínimos de las otras tres variables valen cero en todos los casos y no se presentan en la tabla 9. El análisis de ANOVA de las tres variables, muestra diferencias significativas ($\text{Pr}(>F)=0,0961$) entre las clases solo para TotEWt.

Fenología, Cultivo y Tipo Cultivo

Las variables Fenología, Cultivos y TipoCultivo son categóricas, el estado fenológico se agrupo en tres categorías (tabla 10), el estado “Reproductivo” es el que mayor cantidad de casos presenta en total y se debe a las fechas incluidas en el análisis. En Cultivos se detectaron ocho categorías (tabla 11), de las cuales seis corresponden a cultivos de cosecha, una a forrajeras y una a otros usos del suelo. Para los cultivos de cosecha, “Girasol” y “Soja” son las dos categorías con mayor presencia, nuevamente debido a las fechas incluidas en el estudio. La variable TipoCultivo agrupa los cultivos en cosecha Fina o Gruesa, siendo esta última la de mayor cantidad de casos. El detalle de la distribución según el target se presenta en las tablas 10 a 12.

La cantidad de casos presentes por clase y categorías pueden ser insuficientes para analizar el aporte que este tipo de variable puede hacer al modelo.

Modelos Obtenidos

Target con cinco clases

Las medidas de rendimiento de los modelos obtenidos en los dataset de testing, en general fallan en diferenciar las clases de daño. En la tabla 13 se muestran los cuatro mejores modelos y se aprecia que presentan una importante cantidad de falsos positivos ($\text{FAR} > 40\%$) para todas las clases. El agregado de las variables de cultivos parece importante para clasificar las categorías “Moderado” y “Severo”; debido a la reducida cantidad de casos con los que se construyen y validan estos modelos, y a la subespecificación en el cálculo de las variables es necesario generar un conjunto de datos con mayor cantidad de casos e información más detallada para profundizar el análisis del aporte de estas variables a la clasificación del target.

Los cuatro modelos presentan una buena AUC ($> 74\%$), y excelentes valores de Especificidad

Tabla 10. Estadísticas básicas de las variables de resumen de Fenología en el dataset de todos los volúmenes, todas las elevaciones y con datos de cultivos.

Fenología	Sin Daño	Leve	Moderado	Severo	Grave
Madurez	0 (0%)	14 (74%)	2 (11%)	3 (16%)	0 (0%)
Reproductivo	73 (16%)	299 (64%)	50 (11%)	27 (6%)	21 (4%)
Vegetativo	16 (10%)	97 (60%)	10 (6%)	13 (8%)	26 (16%)
Sin Dato	8 (20%)	26 (65%)	1 (3%)	3 (8%)	2 (5%)

Tabla 11. Estadísticas básicas de las variables de resumen de Cultivos en el dataset de todos los volúmenes, todas las elevaciones y con datos de cultivos.

Cultivos	Sin Daño	Leve	Moderado	Severo	Grave
Avena	0 (0%)	25 (81%)	1 (3%)	310%	2 (6%)
Cebada	1 (2%)	32 (67%)	11 (23%)	12%	3 (6%)
Girasol	28 (17%)	108 (67%)	12 (7%)	85%	5 (3%)
Monte	8 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Maíz	13 (18%)	38 (51%)	6 (8%)	7 (9%)	10 (14%)
Sorgo	6 (38%)	10 (63%)	0 (0%)	0 (0%)	0 (0%)
Soja	30 (16%)	114 (60%)	17 (9%)	11 (6%)	17 (9%)
Trigo	11 (7%)	108 (66%)	16 (10%)	16 (10%)	12 (7%)

Tabla 12. Estadísticas básicas de las variables de resumen de Tipo Cultivos en el dataset de todos los volúmenes, todas las elevaciones y con datos de cultivos.

Tipo Cultivo	Sin Daño	Leve	Moderado	Severo	Grave
Fina	12 (5%)	166 (68%)	28 (12%)	20 (8%)	17 (7%)
Gruesa	77 (18%)	270 (61%)	35 (8%)	26 (6%)	32 (7%)
Sin Dato	8 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Tabla 12. Mejores modelos seleccionados para cada clase (Testing).

	Leve	Moderado	Severo	Grave	
	Volúmenes Completos				
	Todas las Elevaciones		Primera Elevación		TE
	Sin Cultivos	Con Cultivos			Sin Cultivos
	Sin H _{DR}		Con H _{DR}		Sin H _{DR}
Variables	10	12	12	13	10
Variables seleccionadas	MxdBZ, MndBZ, TotdBZ, AvdBZ, MxRho, MnRho, TotRho, MxZDR MnZDR TotZDR	TotdBZ, AvdBZ, MxRho, MxZDR, MnZDR, AvZDR, Fenología, tipoCultivo, Cultivos, TotEWt, AvEWt, MxEWt	TotdBZ1, AvdBZ1, MxRho1, TotRho1, AvRho1, MnZDR1, TotZDR1, AvZDR1, TotEWt, AvEWt, MxEWt, Cultivos	MndBZ1, TotdBZ1, AvdBZ1, MnRho1, TotRho1, AvRho1, MxZDR1, MnZDR1, AvZDR1, AvHDR, MxHDR, MnHDR, TotHDR	MxdBZ, TotdBZ, AvdBZ, MxRho, TotRho, AvRho, MxZDR, MnZDR, TotZDR, AvZDR
FAR	0.4300	0.5263	0.6000	0.7143	0.6000
CSI	0.4790	0.4091	0.3333	0.2857	0.2857
GS	0.4005	0.3359	0.3151	0.2825	0.2815
HSS	0.5720	0.5028	0.4792	0.4406	0.4393
AUC	0.8636	0.7947	0.9084	0.9932	0.7438
POD	0.7500	0.7500	0.6667	1.0000	0.5000
PC	0.8730	0.8617	0.9574	0.9888	0.9898
Especi	0.8956	0.8780	0.9670	0.9888	0.9938
Error Rate	0.1270	0.1383	0.0426	0.0112	0.0102
SR,PPV	0.5700	0.4737	0.4000	0.2857	0.4000
NPV	0.9510	0.9600	0.9888	1.0000	0.9959
F1	0.6477	0.5806	0.5000	0.4444	0.4444

(>87%), Error Rate (<13%), PC (>86%) y NPV (>95%); que evidencian un buen rendimiento en la identificación de los casos negativos; comportamiento esperable dada la mayor proporción de este tipo de casos en los conjuntos de datos. Todos los modelos utilizaron los dataset con los datos de los volúmenes completos. La clase “Grave” es la única que presenta dos modelos que “empatan” en su performance con mejores valores en seis medidas de rendimiento cada uno.

Uno de los modelos utilizó variables referidas a la primera elevación y el otro a variables calculadas con todas las elevaciones. Las clases “Leve” y “Moderado” seleccionaron variables construidas con datos de todas las elevaciones, mientras que

“Severo” eligió variables que resumían datos de la primera elevación. Las clases “Moderado” y “Severo” son las únicas que utilizaron datos relacionados con los cultivos. Solamente la clase “Grave” utilizó variables relacionadas a HDR, coincidiendo con la observación realizado por (Mezher, R. N. et. al. 2011) donde HDR es un buen indicador del daño más alto. Se destaca que en todos los modelos generados con variables de cultivos se seleccionó la variable E, apoyando la idea de (Schiesser 1990) que esta variable es un buen indicador del daño y que mayores valores de Z están relacionados con mayor daño en terreno, concordando en este aspecto con (Mezher, R. N. et. al. 2011). No se presentaron diferencias entre los modelos generados con la función

“ROC Measure” de aquellos generados con la función “Máxima Verosimilitud”.

Target con tres clases

La tabla 13 presenta los resultados de los modelos generados para el target daño con tres clases en el dataset de testing. La clase “Más de 50%”, tampoco presentó buenos rendimientos (FAR=65%, POD=58%), mientras que las clases “Sin Daño” y “Menos del 50%” mostraron muy buenos valores de FAR <13% y POD >94%, reforzando la idea que

estas herramientas son útiles para generar modelos sobre el daño en cultivos en dataset con más casos. Se destaca que la variable HDR fue seleccionada en todos los modelos, por lo que se debe incluir en futuros estudios.

En las figuras 7 a 9 se presentan los árboles de expresión (Ferreyra C. 2001) del mejor modelo de cada clase y las tablas 14 a 16 las funciones logísticas correspondientes (Ferreyra C. 2001, Hosmer y Lemeshow 2000).

Tabla 13. Resultados modelos Daño con tres clases (Testing).

Clase	Variables Polarimétricas			Variables Polarimétricas		
	Todas las elevaciones			Todas las elevaciones		
	Sin Cultivos			Sin Cultivos		
	Con HDR	Menos de 50%	Más de 50%	Con HDR	Menos de 50%	Más de 50%
Variables	16	16	16	16	16	16
Training	976	976	976	1.407	1.407	1.407
Testing	488	488	488	703	703	703
Variables seleccionadas	AvdBZ, MxRho, MnRho, TotRho, MnZDR, TotZDR, MxHDR	MxdBZ, MndBZ, AvdBZ, TotZDR, AvHDR, MnHDR	MxdBZ, TotdBZ, AvRho, MnZDR, AvHDR, TotZDR, AvHDR, MxHDR, MnHDR, TotHDR	MxdBZ, AvRho, MnZDR, AvHDR, MxHDR, MnHDR, TotHDR	MxdBZ, AvdBZ, MxRho, MnRho, AvZDR, AvHDR, MxHDR	MxdBZ, MndBZ, AvdBZ, MnRho, AvRho, MxZDR, TotZDR, AvZDR, MxHDR, MnHDR, TotdBZ
FAR	0.0957	0.1342	0.9333	0.1359	0.4078	0.6557
CSI	0.8750	0.8562	0.0625	0.8056	0.4473	0.2763
GS	0.4388	0.3006	0.0552	0.4160	0.3290	0.2453
HSS	0.6100	0.4622	0.1047	0.5875	0.4952	0.3940
AUC	0.8342	0.8400	0.8171	0.8632	0.8337	0.8322
POD	0.9643	0.9872	0.5000	0.9225	0.6463	0.5833
PC	0.8893	0.8668	0.9385	0.8407	0.8137	0.9218
Especi	0.5833	0.3750	0.9421	0.6350	0.8646	0.9400
Error Rate	0.1107	0.1332	0.0615	0.1593	0.1863	0.0782
SR,PPV	0.9043	0.8658	0.0667	0.8641	0.5922	0.3443
NPV	0.8000	0.8780	0.9956	0.7651	0.8893	0.9766
F1	0.9333	0.9225	0.1176	0.8923	0.6180	0.4329

Sub-ET 1

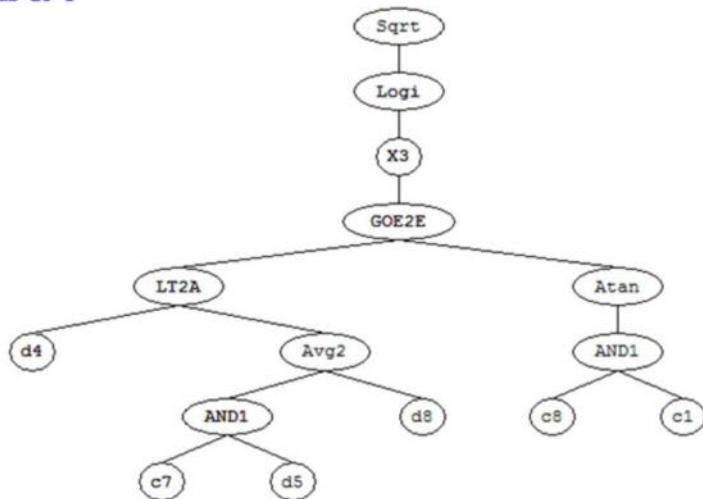
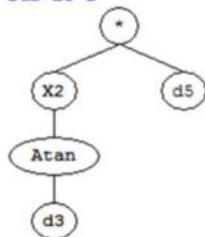
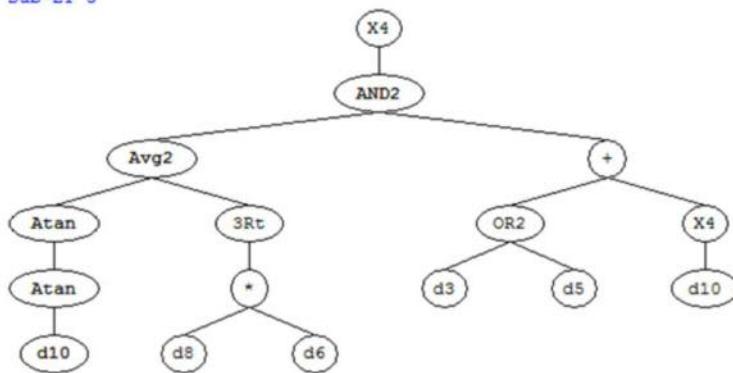


Figura 7. Árbol de expresión para la clase “Sin Daño”. Variables: AvdBZ = d₃, MxRho = d₄, MnRho = d₅, TotRho = d₆, MnZDR = d₈, TotZDR = d₁₀, MxHDR = d₁₃, Constantes: G₁C₈ = -5.04562517166662, G₁C₁ = 0.33051545762505, G₁C₇ = -5.06332590716269, G₄C₉ = 4.93575853755303, G₄C₁ = -9.5501571703238, G₄C₄ = -7.68059327982421.

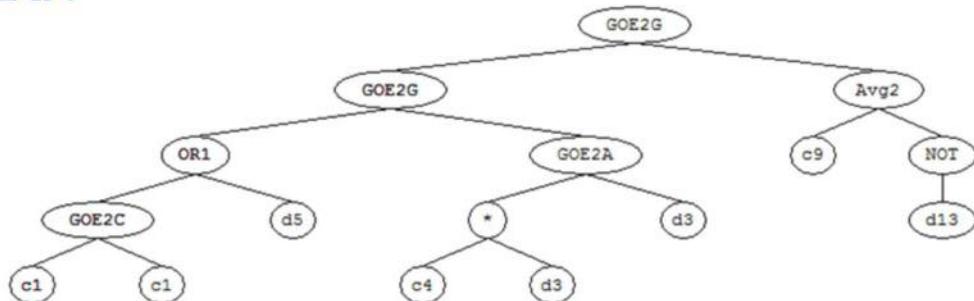
Sub-ET 2



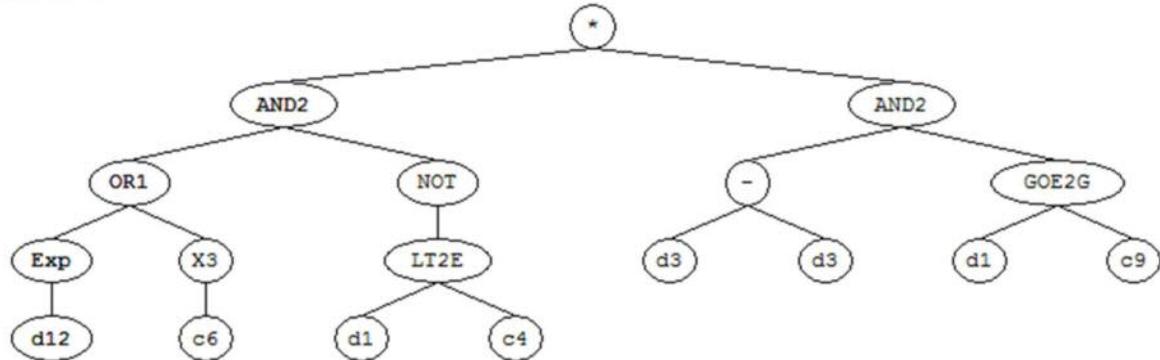
Sub-ET 3



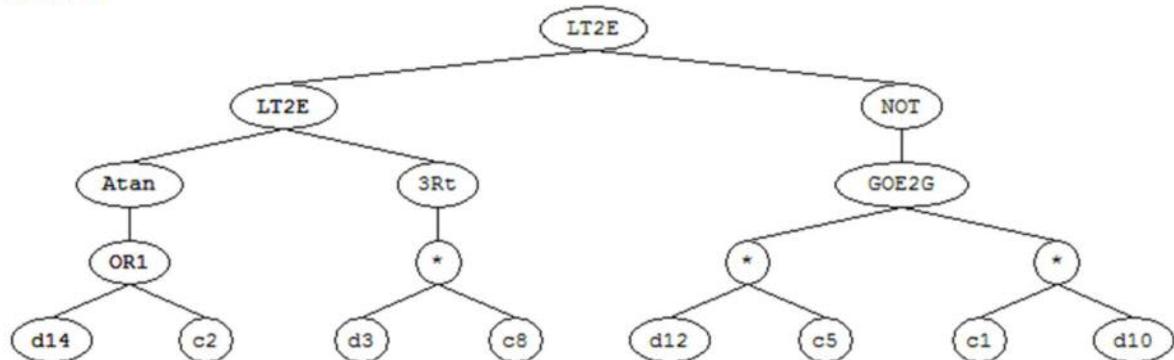
Sub-ET 4



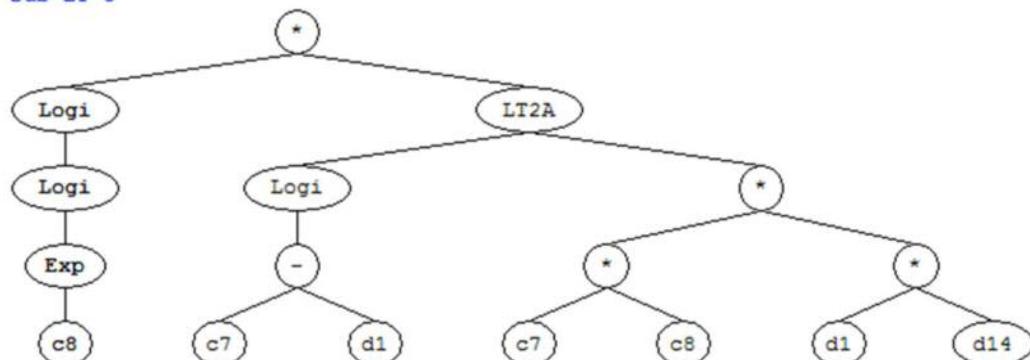
Sub-ET 1



Sub-ET 2



Sub-ET 3



Sub-ET 4

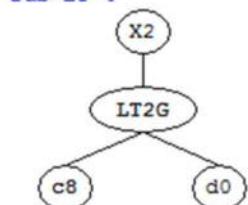
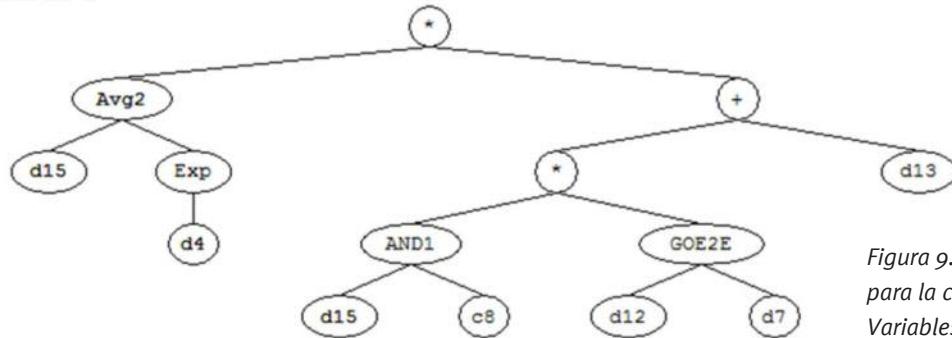
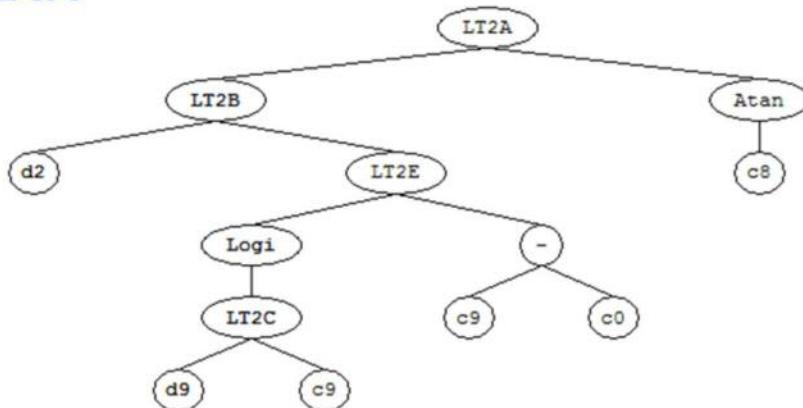


Figura 8. Árbol de expresión para la clase “Menos 50%”. Variables: $MxdBZ = do$, $MndBZ = d1$, $AvdBZ = d3$, $TotZDR = d10$, $AvHDR = d12$, $MnHDR = d14$. Constantes: $G1C9 = -2.05908383434553$, $G1C6 = -3.34879604480117$, $G1C4 = -3.62407300027467$, $G2C2 = -0.135196996978668$, $G2C8 = -0.356761375774407$, $G2C5 = 7.6281014435255$, $G2C1 = 5.92883083590197$, $G3C8 = 0.924405652027955$, $G3C7 = 2.61452070680868$, $G4C8 = -6.3365581225013$.

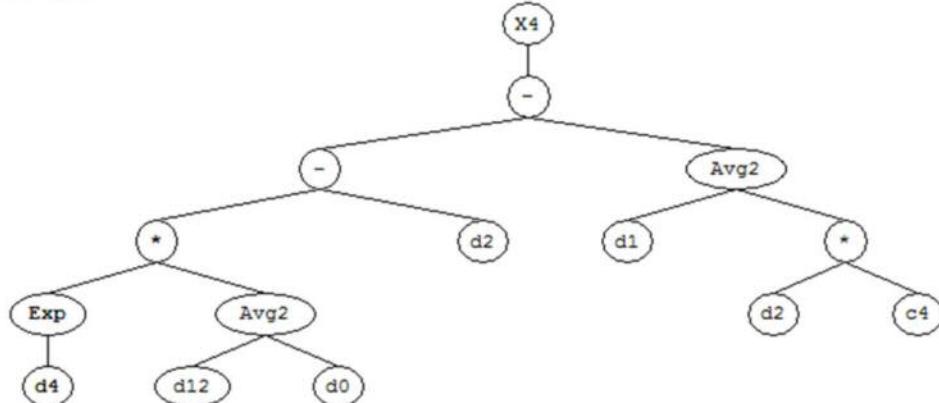
Sub-ET 1



Sub-ET 2



Sub-ET 3



Sub-ET 4

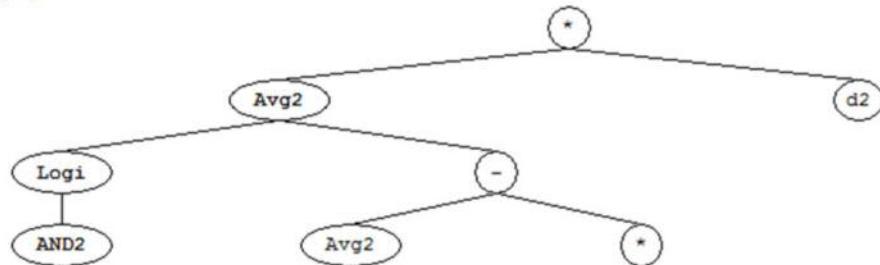


Figura 9. Árbol de expresión para la clase “Menos 50%”.
 Variables: $MxdBZ = d_0, MndBZ = 1, AvdBZ = d_2, MnRho = d_4, AvRho = d_6, MxZDR = d_7, TotZDR = d_9, AvZDR = d_{10}, MxHDR = d_{12}, MnHDR = d_{13}, TotdBZ = d_{15}$. Constantes:
 $G_1c_8 = -6.88711203344829, G_2c_8 = -1.01255836664937, G_2c_9 = 5.52171391949217, G_2c_{10} = 5.27695547349467, G_3c_4 = -5.55345316934721, G_4c_6 = -4.31440168462172$

Tabla 14. Función logística para la probabilidad positivo de la clase “Sin Daño” y detalles de su construcción.

Función logística						
$P(Y=1) = \frac{1}{1 + \exp(0.682238413415047 * y + 4.30616235143977)}$						
Armado de y						
$y =$ $\text{sqrt}(\text{gepLogi}(\text{pow}(\text{gepGOE2E}(\text{gepLT2A}(d[\text{MxRho}], ((\text{gepAND1}(\text{G1C7}, d[\text{MnRho}]) + d[\text{MnZDR}]) / 2.0)), \text{atan}(\text{gepAND1}(\text{G1C8}, \text{G1C1})))), 3.0)))$						
$y = y + (\text{pow}(\text{atan}(d[\text{AvdBZ}]), 2.0) * d[\text{MnRho}])$						
$y = y +$ $\text{pow}(\text{gepAND2}(((\text{atan}(\text{atan}(d[\text{TotZDR}])) + \text{gep3Rt}((d[\text{MnZDR}] * d[\text{TotRho}])) / 2.0), (\text{gepOR2}(d[\text{AvdBZ}], d[\text{MnRho}]) + \text{pow}(d[\text{TotZDR}], 4.0))), 4.0)$						
$y = y +$ $\text{gepGOE2G}(\text{gepGOE2G}(\text{gepOR1}(\text{gepGOE2C}(\text{G4C1}, \text{G4C1}), d[\text{MnRho}]), \text{gepGOE2A}((\text{G4C4} * d[\text{AvdBZ}]), d[\text{AvdBZ}]), ((\text{G4C9} + (1.0 - d[\text{MxHDR}])) / 2.0)))$						
Funciones lógicas para el armado de y						
gep3Rt	gepOR1	gepOR2	gepAND1			
Recibe x: if ($x < 0.0$): return $-\text{pow}(-x, (1.0 / 3.0))$, else: return $\text{pow}(x, (1.0 / 3.0))$	Recibe(x, y): if ($(x < 0.0)$ or ($y < 0.0$)): return 1.0, else: return 0.0	Recibe(x, y): if ($(x \geq 0.0)$ or ($y \geq 0.0$)): return 1.0, else: return 0.0	Recibe (x, y): if ($(x < 0.0)$ and ($y < 0.0$)): return 1.0 else: return 0.0			
gepAND2	gepLT2A	gepGOE2A	gepGOE2C			
Recibe (x, y): if ($(x \geq 0.0)$ and ($y \geq 0.0$)): return 1.0 else: return 0.0	Recibe (x, y): if ($x < y$): return x else: return y	Recibe (x, y): if ($x \geq y$): return x else: return y	Recibe (x, y): if ($x \geq y$): return $(x+y)$ else: return $(x-y)$			
gepGOE2E	gepGOE2G	gepLogi				
Recibe (x, y): if ($x \geq y$): return $(x+y)$ else: return $(x*y)$	Recibe (x, y): if ($x \geq y$): return $(x+y)$ else: return $\text{atan}(x*y)$	Recibe (x): if ($\text{abs}(x) > 709.0$): return $1.0 / (1.0 + \text{exp}(\text{abs}(x) / x * 709.0))$ else: return $1.0 / (1.0 + \text{exp}(-x))$				

Tabla 15. Función logística para la probabilidad positivo de la clase “Menos de 50%” y detalles de su construcción.

Función logística						
P(Y=1) =	$\frac{1}{1 + \exp(-(2.69480869158104E-03*y + -4.93561195211844))}$					
Armado de y						
$y = (\text{gepAND2}(\text{gepOR1}(\exp(d[\text{AvHDR}]), \text{pow}(G1C6, 3.0)), (1.0 - \text{gepLT2E}(d[\text{MndBZ}], G1C4))) * \text{gepAND2}((d[\text{AvdBZ}] - d[\text{AvdBZ}]), \text{gepGOE2G}(d[\text{MndBZ}], G1C9)))$						
$y = y + \text{gepLT2E}(\text{gepLT2E}(\text{atan}(\text{gepOR1}(d[\text{MnHDR}], G2C2)), \text{gep3Rt}((d[\text{AvdBZ}] * G2C8))), (1.0 - \text{gepGOE2G}((d[\text{AvHDR}] * G2C5), (G2C1 * d[\text{TotZDR}])))))$						
$y = y + (\text{gepLogi}(\text{gepLogi}(\exp(G3C8))) * \text{gepLT2A}(\text{gepLogi}((G3C7 - d[\text{MndBZ}]), ((G3C7 * G3C8) * (d[\text{MndBZ}] * d[\text{MnHDR}])))))$						
Funciones lógicas para el armado de y						
gep3Rt	gepOR1	gepAND2	gepLT2A			
Recibe (x): if (x < 0.0): return - pow(-x,(1.0/3.0)), else: return pow(x,(1.0/3.0))	Recibe (x, y): if ((x < 0.0) or (y < 0.0)): return 1.0, else: return 0.0	Recibe (x, y): if ((x >= 0.0) and (y >= 0.0)): return 1.0, else: return 0.0	Recibe (x, y): if (x < y): return x, else: return y			
gepLT2E	gepLT2G	gepGOE2G	gepLogi			
Recibe (x, y): if (x < y): return (x+y), else: return (x*y)	Recibe (x, y): if (x < y): return (x+y), else: return atan(x*y)	Recibe (x, y): if (x >= y): return (x+y), else: return atan(x*y)	Recibe (x): if (abs(x) > 709.0): return 1.0 / (1.0 + exp(abs(x) / x * 709.0)), else: return 1.0 / (1.0 + exp(-x))			

CONCLUSIONES

El daño fue tratado como problemas binarios con cinco y tres clases. Además de las variables polarimétricas del radar se incorporaron variables del cultivo. Los modelos obtenidos para las cinco clases fallan en diferenciarlas. Muestran una importante cantidad de falsos positivos (FAR > 40%) para todas las clases. Estos resultados eran esperados por la subespecificación de los datos del radar intradía, la poca cantidad de casos para algunas de las clases y la falta de variables relacionadas al cultivo. Sin embargo, se obtuvieron buenos resultados en otras medidas de rendimiento (AUC > 74%, PC > 86%, NPV > 95%, Error Rate < 13% y Especificidad > 87%) lo que sugiere que se pueden usar estas herramientas para análisis futuros en un conjunto de datos más grande y completo. En el segundo análisis del daño con tres clases, la clase correspondiente al mayor porcentaje de daño, no presentó buenos rendimientos (FAR=65%, POD=58%), este resultado también

se esperaba, por las mismas razones mencionadas en el párrafo anterior. Las otras dos clases se pueden diferenciar muy satisfactoriamente (FAR < 13% y POD > 94%), reforzando la idea que estas herramientas son útiles para generar modelos que clasifican el daño en cultivos.

Se pudieron utilizar como variables de entrada para los modelos, resúmenes (máximos, mínimos, promedios, totales y cantidades) de las variables Z, ZDR y RhoHV. También se pudieron calcular las variables E y HDR derivadas de Z y ZDR.

En el caso del Daño a cultivos, HDR y E se presentan como un buen indicador del mismo. Mayores valores de Z están relacionados con mayor daño en terreno. Para el daño más alto ZDR presenta valores más cercanos a cero, que puede asociarse a la presencia de granizos grandes y en mucha cantidad, que dominan Z y aparecen como esferas perfectas; para el resto de las clases ZDR presenta un promedio por encima de 4 dB, acorde al comportamiento de la banda C en presencia de granizo.

Tabla 16. Función logística para la probabilidad positivo de la clase “Más de 50%” y detalles de su construcción.

Función logística						
$P(Y=1) = \frac{1}{1 + \exp(-(3.84883671107244E-08*y + -7.00046478207421))}$						
Armado de y						
$y = (((d[TotdBZ]+exp(d[MnRho]))/2.0)*((gepAND1(d[TotdBZ],G1C8)*gepGOE2E(d[MxHDR],d[MxZDR]))+d[MnHDR]))$						
$y = y + \text{gepLT2A}(\text{gepLT2B}(d[AvdBZ],\text{gepLT2E}(\text{gepLogi}(\text{gepLT2C}(d[TotZDR],G2C9),(G2C9-G2C0))),\text{atan}(G2C8)))$						
$y = y + \text{pow}(((\exp(d[MnRho])*(d[MxHDR]+d[MxdBZ])/2.0))-d[AvdBZ])-((d[MndBZ]+(d[AvdBZ]*G3C4))/2.0)),4.0)$						
$y = y + (((\text{gepLogi}(\text{gepAND2}(\exp(G4C6),\text{pow}(d[AvRho],4.0))))+((d[MndBZ]+d[AvZDR])/2.0)-(d[MnHDR]*d[MxdBZ])))/2.0)*d[AvdBZ]))$						
Funciones lógicas para el armado de y						
gepAND1	gepAND2	gepLT2A	gepLT2B			
Recibe (x, y): if ((x < 0.0) and (y < 0.0)): return 1.0, else: return 0.0	Recibe (x, y): if ((x >= 0.0) and (y >= 0.0)): return 1.0, else: return 0.0	Recibe (x, y): if (x < y): return x, else: return y	Recibe (x, y): if (x < y): return 1.0, else: return 0.0			
gepLT2C	gepLT2E	gepGOE2E	gepLogi			
Recibe (x, y): if (x < y): return (x+y) else: return (x-y)	Recibe (x, y): if (x < y): return (x+y) else: return (x*y)	Recibe (x, y): if (x >= y): return (x+y) else: return (x*y)	Recibe (x): if (abs(x) > 709.0): return 1.0 / (1.0 + $\exp(\text{abs}(x) / x * 709.0))$ else: return 1.0 / (1.0 + $\exp(-x))$			

El método GEP tanto la regresión logística como la clasificación arrojaron resultados similares. Se debe completar un mayor set de datos para poder evaluar el rendimiento de la herramienta.

A partir de los resultados de este primer análisis exploratorio, sería conveniente en futuros trabajos, poder aislar el momento de ocurrencia de la tormenta y trabajar solo con los datos correspondientes a ese momento, evitando de esta manera suavizar los valores de las variables del radar. Otra característica a determinar, para evitar la subespecificación, son las elevaciones que realmente aportan a la identificación de la caída en superficie y a la diferenciación del daño. También sería muy importante darle continuidad a la actualización y ampliación de la información de verdad de campo utilizando el sistema de información desarrollado para tal fin.

Como el daño ocasionado está estrechamente

relacionado con el cultivo y su estado fenológico, para lograr mejores modelos sería conveniente analizar los datos por cultivo o tipos de cultivos similares y por momento de ocurrencia asociándolo a la fenología del mismo. La información podría ser provista por compañías de seguro, que cuentan con el detalle del cultivo afectado y su estado fenológico.

Las variables PhiDP y KDP se deberían incluir en futuros modelos para analizar su aporte a los target buscados, una vez que se pueda programar su lectura y procesamiento.

Se destaca que la variable HDR fue seleccionada en todos los modelos en los cuales se utilizó, por lo que también se debe incluir en futuros estudios; si se quiere generar un modelo que solo dependa de Z, se debe incluir también la variable E con diferentes configuraciones y métodos de cálculo.

REFERENCIAS BIBLIOGRÁFICAS

- Alexiuk, M., Li, P. C., Pizzi, N., & Pedrycz, W. 1999. "Classification of Hail and Tornado Storm Cells Using Neural Networks". In 1999 IEEE Western Canada Conference and Exhibition (pp. 15-21).
- _____, N. Pizzi, y W. Pedrycz. 1999b. «Classification of volumetric storm cell patterns». En Electrical and Computer Engineering, 1999 IEEE Canadian Conference on, 2:1081–1085. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=808201.
- Almada, Natalia, Florencia Díaz, Leticia Osorio, Juan Blatter, Héctor Rodriguez, Juan José De Battista, Norma Arias, y Carlos Bocchio. 2005. «Evaluación del daño por granizo en soja». Boletín Técnico EEA Concepción del Uruguay, Producción Vegetal, 46: 3.
- Anderson, M. E., L. D. Carey, W. A. Petersen, y K. R. Knupp. 2012. «C-band Dual-polarimetric Radar Signatures of Hail», julio. <http://www.nwas.org/ej/2011-EJ2/>.
- Aydin, K., T. A. Seliga, y V. Balaji. 1986. «Remote sensing of hail with a dual linear polarization radar». Journal of Climate and Applied Meteorology 25 (10): 1475–1484.
- Bellini Saibene, Y. 2015. "Estimación de ocurrencia de granizo en superficie y daño en cultivos utilizando datos de radar meteorológico y técnicas de minería de datos". Tesis de Maestría. Octubre.
- _____, Caldera, J. M., & Volpacchio, M. 2015. «HAR [Hail-ARchive]. desarrollo de un sistema de información y base de datos sobre granizo en la región semiárida pampeana central». In Congreso Argentino de Meteorología. 12. CONGREMET. 26-29 de mayo de 2015. Mar del Plata, Buenos Aires. AR.
- _____, Guillermo Casagrande. 2010. «Radar Meteorológico en la EEA Anguil». Horizonte Agropecuario, abril, 85 edición.
- ——, S. Banchero, R.N. Mezher. 2014. "Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA". Anales del 6º Congreso Argentino de Agroinformática: 74-86. Septiembre. Argentina.
- ——, S. Banchero. 2016. "RadINTA: Facilitando el uso de la red de radares de INTA mediante el desarrollo de herramientas libres." 3º Jornadas de Tecnología de Información Geográfica del Sur Argentino. Bahía Blanca, Argentina. 24 al 26 de Agosto.
- Billet, John, Mark DeLisi, Brian G. Smith, y Cory Gates. 1997. «Use of Regression Techniques to Predict Hail Size and the Probability of Large Hail». Weather and Forecasting 12 (1): 154-64. doi:10.1175/1520-0434(1997)012<0154:UORTTP>2.0.CO;2.
- Boettcher, Jami. 2013. «Dual Polarization RADAR Operation Course». Dual Polarization RADAR Operation Course. <http://www.wdtb.noaa.gov/courses/dualpol/RDA/Lesson1/player.html>. Accedido noviembre 2013
- Center, NOAA / NWS Space Weather Prediction. 2013. «Space Weather Prediction Center - Forecast Verification Glossary». Accedido noviembre 16. http://www.swpc.noaa.gov/forecast_verification/Glossary.html#G.
- Ceperuelo Mallafré, Manuel. 2008. «Identificación y caracterización del granizo mediante el radar meteorológico. Modelos de predicción del ciclo de vida de las células convectivas.» Tesis Doctoral.
- Changnon, David, y Stanley A. Changnon. 1997. «Surrogate data to estimate crop-hail loss». Journal of applied meteorology 36 (9): 1202–1210.
- Changnon Jr, Stanley A. 1971. «Hailfall characteristics related to crop damage». Journal of Applied Meteorology 10 (2): 270–274.
- Collino, Elena, Paolo Bonelli, y Luciano Gilli. 2009. «ST-AR (STorm-ARchive): A project developed to assess the ground effects of severe convective storms in the Po Valley». Atmospheric Research 93 (1-3): 483-89. doi:10.1016/j.atmosres.2008.10.021.
- Ferreira, Cândida. 2001. «Gene expression programming: a new adaptive algorithm for solving problems». arXiv preprint cs/0102027. <http://arxiv.org/abs/cs/0102027>.
- Foote, G. Brant, T. W. Krauss, y V. Makitov. 2005. «Hail metrics using conventional radar». En Proc., 16th Conference on Planned and Inadvertent Weather Modification. https://ams.confex.com/ams/Annual2005/techprogram/aper_86773.htm.
- Gagne, David John, Amy McGovern, y Jerry Brotzge. 2009. «Classification of convective areas using decision trees». Journal of Atmospheric and Oceanic Technology 26 (7): 1341–1353.
- Hohl, Roman, y Martin Beniston. 2001. «Relationship between hailfall intensity and hail damage on ground, determined by radar and lightning observations». Doctorial thesis. University Freiburg, Switzerland. <http://ethesis.unifr.ch/theses/HohlRM.pdf>.
- Hohl, Roman, Hans-Heinrich Schiesser, y Dörte Aller. 2002. «Hailfall: the relationship between radar-derived hail kinetic energy and hail damage to buildings». Atmospheric Research 63 (3): 177–207.
- _____, _____ y Ingeborg Knepper. 2002b. «The use of weather radars to estimate hail damage to automobiles: an exploratory study in Switzerland». Atmospheric research 61 (3): 215–238.
- Holleman, I., H. R. Wessels, J. R. Onvlee, y S. J. Barlag.

2000. «Development of a hail-detection-product». Physics and Chemistry of the Earth Part B Hydrology Oceans and Atmosphere 25: 1293–1297.
- Hosmer, D. W., y S. Lemeshow. 2000. Applied logistic regression. Second. Vol. 354. Wiley-Interscience.
 - Katz, Richard W., y Rolando R. Garcia. 1981. «Statistical relationships between hailfall and damage to wheat». Agricultural Meteorology 24: 29–43.
 - Keenan, T. 2003. «Hydrometeor classification with a C-band polarimetric radar». Australian Meteorological Magazine 52 (1): 23–31.
 - Lakshmanan, V., & Smith, T. 2009. Data mining storm attributes from spatial grids. Journal of Atmospheric and Oceanic Technology, 26(11), 2353-2365.
 - León Aristizábal, Gloria. 2005. «Verificación de los modelos Meterológicos». Instituto de Hidrología, Meteorología y Estudios Ambientales - IDEAM.
 - Li, Pak C., Nicolino Pizzi, Witold Pedrycz, Dave Westmore, y Rodrigo Vivanco. 2000. «Severe storm cell classification using derived products optimized by genetic algorithms». En Electrical and Computer Engineering, 2000 Canadian Conference on, 1:445–448. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=849748.
 - Makitov, Viktor. 2007. «Radar measurements of integral parameters of hailstorms used on hail suppression projects». Atmospheric Research, European Conference on Severe Storms 2004 ECSS 2004 European Conference on Severe Storms 2004, 83 (2–4): 380-88. doi:10.1016/j.atmosres.2005.09.016.
 - Marzban, C., y A. Witt. 2001. «A Bayesian neural network for severe-hail size prediction». Weather and Forecasting 16 (5): 600–610.
 - May, P. T., y T. D. Keenan. 2003. «Four-dimensional microphysical data from Darwin». En 13th ARM Science Team Meeting Proceedings. https://education.arm.gov/publications/proceedings/conf13/extended_abs/may2-pt.pdf.
 - Mezher, R. N, y P. A Mercuri. 2010. «Uso de la red de radares de INTA para la detección de granizo». XIII Reunión Argentina y VI Latinoamericana de Agrometeorología., octubre.
 - ———, Luciano Vidal, y Paola Salio. 2011. «Hailstorms Analysis using Polarimetric Weather Radars and Microwave Sensors in Argentina». 6th European Conference on Severe Storms (ECSS 2011).
 - ———, Santiago Banchero, y Yanina Noemí Bellini Saibene. 2012. «Identificación de granizo con la utilización de variables polarimétricas de los radares de Paraná y Anguil, el radar de Pergamino y daño en cultivos.» En Congreso Argentino de Meteorología. 11., 28 de mayo al 1 de junio de 2012. Mendoza. AR. <http://orton.catie.ac.cr/cgi-bin/wxis.exe/?IsisScript=INTA2.xis&method=post&formato=2&cantidad=1&expresion=mfn=067676>.
 - Morgan Jr, Griffith M., y Neil G. Towery. 1976. «Crop Damage-Hailpad Parameter Study In Illinois». Illinois State Water Survey, Urbana 101: 55.
 - Peters, J. F., Suraj, Z., Shan, S., Ramanna, S., Pedrycz, W. y Pizzi, N. 2003. “Classification of meteorological volumetric radar data using rough set methods,” Pattern Recognition Letters, vol. 24, no. 6, pp. 911–920, Mar.
 - Ramirez, L., W. Pedrycz, y N. Pizzi. 2001. «Severe storm cell classification using support vector machines and radial basis function approaches». En Electrical and Computer Engineering, 2001. Canadian Conference on, 1:87–91. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=933663.
 - Rinehart, R. E. 1997. “Radar for meteorologists.” Dept. of Atmospheric Sciences, Center for Aerospace Sciences, University of North Dakota.
 - Rose, R., Hoople, C., McGill, M., Dehmer, K., Raebler, C., Byham, R., Skinner, L., Cleary, W. 2000. «Programming a Microsoft SQL Server 2000 DataBase». Microsoft Training and Certification. 2329A course.
 - Ryzhkov, A. V., T. J. Schuur, D. W. Burgess, P. L. Heinselman, S. E. Giangrande, y D. S. Zrnic. 2005. «The joint polarization experiment. polarimetric Rainfall Measurement and Hidrometeor Classification». Bull. Amer. Meteor. Soc 86: 809–824.
 - Ryzhkov, A., D. Zrnic, J. Krause, M. Kumjian, y S. Ganson. 2010. «Discrimination Between Large And Small Hail. Final Report». NSSL.
 - Schiesser, H. H. 1990. «Hailfall: the relationship between radar measurements and crop damage». Atmospheric Research 25 (6): 559–582.
 - Sánchez, J. L., L. López, E. García-Ortega, y B. Gil. 2013. «Nowcasting of kinetic energy of hail precipitation using radar». Atmospheric Research 123: 48–60.
 - Sánchez, J. L., R. Fraile, J. L. De La Madrid, M. T. De La Fuente, P. Rodríguez, y A. Castro. 1996. «Crop damage: The hail size factor». Journal of Applied Meteorology 35 (9): 1535–1541.
 - Seino, H. 1980. «On the characteristics of hail size distribution related to crop damage». Journal of Agricultural Meteorology, 36(2), 81-88.
 - Skripnikova, Katerina, Daniela Rezacova, y Katerina Skripnikova. 2014. «Testing radar-based hail detection criteria». Accedido agosto 2. http://www.meteo.fr/cic/meetings/2012/ERAD/extended_abs/NOW_367_ext_abs.pdf.
 - Tabary, P., C. Berthet, P. Dupuy, J. Figueras i Ventura, B. Fradon, J. F. Georgis, R. Hogan, F. Kabeche, y J. P. Wasselin. 2010. «Hail detection and quantification with C-band polarimetric radar». Atmospheric Research 94 (1–4): 1–12.

- metric radars: results from a two-year objective comparison against hailpads in the south of France». En Proceedings of 6th European Conference on Radar in Meteorology and Hydrology, Sibiu, Romania, 6-10 Sept.
- Tsagalidis, E., Tsitouridis, K., Evangelidis, G., & Dervos, D. 2008. Hail Size Estimation and Prediction using Data Mining Techniques. In ERAD 2008—The 5th European Conference on Radar in Meteorology and Hydrology.
 - Visser, Petrus JM, y J. van Heerden. 2000. «Comparisons of hail kinetic energy derived from radar reflectivity with crop damage reports over the eastern Free State». Water SA-Pretoria- 26 (1): 91–96.
 - Waldvogel, A., W. Schmid, y B. Federer. 1978. «The Kinetic Energy of Hailfalls. Part I: Hailstone Spectra». Journal of Applied Meteorology 17 (4): 515-20. doi:10.1175/1520-0450(1978)017<0515:TKEOHP>2.0.CO;2.
 - Williamson, Samuel P. 2006. Doppler Radar meteorological observations. Part c wsr-88d products and algorithms. Federal meteorological handbook 11. Washington, DC: Office Of The Federal Coordinator For Meteorological Services And Supporting Research. Http://www.ofcm.gov.
 - Witt, A., M. D Eilts, G. J Stumpf, J. T. Johnson, E. D.W Mitchell, y K. W Thomas. 1998. «An enhanced hail detection algorithm for the WSR-88D». Weather and Forecasting 13 (2): 286–303.
 - Xiang, Li, Rahul Ramachandran, John Rushing, Sara Graves, Kevin Kelleher, S. Lakshminarahan, Douglas Kennedy, y Jason Levit. 2004. «Mining nexrad radar data: An investigative study». En American Meteorology Society annual meeting. <https://ams.confex.com/ams/pdfpapers/69342.pdf>.
 - Zbigniew, Suraj, J.F. Peters, y W. Rzasa. 2002. «A Comparison of Different Decision Algorithms Used in Volumetric Storm Cells Classification». Fundamenta Informaticae 51 (1): 201-14.
 - _____, y W. Rzasa. 2001. «Volumetric storm cell classification with the use of rough set methods». Zeszyty Naukowe WSI Iz 1. <http://logic.mimuw.edu.pl/Grant2003/prace/GSCDPSuraj1.pdf>.
 - Zrnic, D. 2012. «Dual Polarization for Weather Observations, Clasification of Hydrometeors and Measurement of Rain.» presentado en Simposio de Radarización y Sistema de Alertas Hidrometeorológicos del Cono Sur, Buenos Aires, Argentina, septiembre 11. <http://pampero.cima.fcen.uba.ar/SimposioAlertas/>.
-

HAR (HAIL ARCHIVE): sistema de información para registrar eventos de granizo, sus efectos en superficie y los datos de radares meteorológicos en la región pampeana central

8.

Bellini Saibene, Yanina¹; Caldera, Juan Marcelo¹

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA
bellini.yanina@inta.gob.ar

RESUMEN

Para analizar el granizo se necesitan datos con alta resolución espacial y temporal debido a la pequeña escala y corta duración de este fenómeno. Estos datos son obtenidos desde diversas fuentes de información que necesitaban ser integradas y organizadas en una base de datos. Fue necesario desarrollar un sistema de información para gestionar y analizar la diversidad de los datos. La base de datos contempla dos tipos de datos: 1) los registrados por el radar y 2) los datos de campo que detallan la ocurrencia de granizo y daño en cultivos. La base de datos contiene aproximadamente 3.700 registros geo referenciados para los años 2009 a 2013, en las provincias de La Pampa, Buenos Aires, Córdoba, San Luis y Santa Fé.

La base de datos relacional se realizó en SQL Server 2008 Express con extensión para el tratamiento de datos geográficos. Se administra por medio de un sistema de información desarrollado en la plataforma web ASP.NET, utilizando Visual Basic .NET como lenguaje de programación. La herramienta se encuentra disponible en <http://rian.inta.gob.ar/DanioGranizo>

Palabras clave: integración de información, sensores remotos, agrometeorología, datos abiertos

ABSTRACT

Data with high spatial and temporal resolution are necessary to analyze the characteristics of the

hail due to the small scale and short duration of this phenomena. This type of data is provided by various sources of information that needed to be integrated and organized in a database. It was also necessary to develop an information system that allowed handling different kind of data and different analysis. The database includes two types of data: 1) those recorded by a weather radar and 2) hail ground effect reports. The database contains approximately 3,700 geo-referenced points of occurrence of hail in the period March 2009 to March 2013 in the provinces of La Pampa, Buenos Aires, Córdoba, San Luis and Santa Fe.

The database was developed using SQL Server 2008 Express with extension for geographic data processing. The information system for managing the database was developed in ASP.NET using Visual Basic .NET programming language. This information system is available in <http://rian.inta.gob.ar/DanioGranizo>.

Keywords: data integration, remote sensing, agrometeorology, open data

INTRODUCCIÓN

El análisis de eventos climáticos severos, como el granizo, resulta importante para las personas, el sector público y la industria de los seguros, quienes deben afrontar el costo y sufrir las consecuencias del daño que ocasionan dichos eventos. Las tormentas severas suelen ser localizadas y de corta dura-

ción, aunque el daño que ocasionan puede ser muy importante, tanto en cultivos, infraestructura, medios de transporte e incluso personas (Ponce de Leon, 1985, Kats *et. al.* 1981, Collino, *et. al.* 2009). Ante estas necesidades de información, además del nowcasting (pronóstico a corto plazo), los análisis post evento también son importantes (Collino, *et. al.* 2009).

Para el caso específico del granizo, se necesitan datos con alta resolución espacial y temporal debido a la pequeña escala y corta duración de este fenómeno meteorológico extremo (Ponce de Leon, 1985). Cuando no se cuenta con una red densa de observación o una red de granzómetros, como la presente en la provincia de Mendoza (Bustos *et. al.* 1982), se utilizan sensores remotos como los satélites y radares meteorológicos (Hohl, R. *et. al.*, 2002, Collino, *et. al.* 2009) que permiten contar con información de alta resolución temporal y espacial.

Para poder analizar los efectos en terreno del granizo, validar los productos que genera el software del RADAR y desarrollar nuevos productos, los registros realizados por el instrumento deben ser contrastados con datos de terreno.

De acuerdo a Collino, *et. al.* 2009, piezas de información descriptiva sobre el daño ocasionado pueden suplir la falta de información de observaciones desde redes de estaciones, observadores o granzómetros.

Estas piezas de información se pueden encontrar en diferentes fuentes como la prensa (Collino, *et. al.* 2009, Bellini Saibene, Y. 2015), los reportes de compañías de seguro (Bellini Saibene, Y. 2015, Visser, P., *et. al.* 2000, Mezher *et. al.* 2012), entre otras. Por esta razón en diversos países existen sistemas nacionales de registro de este tipo de información: la European Severe Weather Database (ESWD) se creó para organizar información que permitiera validar modelos para estudios climáticos (Collino *et. al.* 2009), el Storm Archive (Collino *et. al.* 2009) se generó para determinar los efectos en terreno de tormentas convectivas en una región de Italia, organizado información de radar, periódicos y rayos caídos; Schiesser et. al 1995 recolecta información de sensores remotos, noticias de diarios y compañías de seguro para evaluar el daño de tormentas severas en Suiza. Muchos autores resaltan la importancia de contar con conjuntos de datos con verdad de campo para poder validar y entrenar modelos de clima severo (Bellini Saibene, Y. 2015, Schiesser et.

al 1995, Collino *et. al.* 2009). En nuestro país, el emprendimiento más reciente del Servicio Meteorológico Nacional y el INTA llamado AlertAR, permite a un usuario de SmartPhone registrar un evento climático e informarlo a un sistema central, con la idea que estos datos conformen un repositorio de información de campo para mejorar y generar modelos. El mismo es muy reciente y no integra datos con otras fuentes.

No existe a nivel nacional un sistema que centre y registre de forma exhaustiva los eventos de granizo y sus efectos en superficie; por lo que estos datos son provistos por diversas fuentes de información que necesitan ser integradas y organizadas.

La herramienta adecuada para organizar, almacenar y procesar datos de fuentes diversas es una base de datos relacional, administrada por un sistema de información, con el objetivo que la mayor cantidad de usuarios puedan gestionar la diversidad de: los orígenes de los datos, los tipos de datos, la frecuencia de los datos y la necesidad de los análisis de estos datos.

Este trabajo presenta un modelo de datos, el sistema de información desarrollado y la información registrada sobre la ocurrencia de granizo en la región semiárida pampeana central desde nueve orígenes de datos diferentes.

METODOLOGÍA

La base de datos contempla datos que registra el radar meteorológico y datos de campo que detallan la ocurrencia de granizo y daño en cultivos.

Datos del radar

Los datos del sensor remoto incorporados al sistema, corresponden al Radar Doppler de banda C situado en la Estación Experimental Agropecuaria Anguil, del Instituto Nacional de Tecnología Agropecuaria (INTA) en La Pampa. Tienen una resolución temporal de 10 minutos y espacial de 1 km² en un rango de 240 km y 12 elevaciones (inicia a 0,5° y aumenta el ángulo de elevación 11 veces más: 0,9°; 1,3°; 1,9°; 2,3°; 3,0°; 3,5°; 5,0°; 6,9°; 9,1°; 11,8° y 15,1°).

Se registran las variables:

- factor de reflectividad (Z) (Rinehart, 1997),
- reflectividad diferencial (ZDR) (Rinehart, 1997) y

- coeficiente de correlación polarimétrica (RhoHV) (Rinehart, 1997).

Estos datos se transformaron en imágenes raster georeferenciadas utilizando el software RadINTA (Bellini, Saibene, Y. et. al., 2015b). Para un día completo se generan 144 tomas x 3 variables x 12 elevaciones = 5.184 imágenes.

Datos de campo

Los datos de campo se recolectaron, organizaron e integraron desde ocho fuentes distintas, las cuales se utilizan ampliamente para la validación de métodos de estimación de ocurrencia y tamaño de granizo (Bellini Saibene, Y. et. al, 2015b, Bellini Saibene, Y., 2015d):

- a)** *Reportes de compañías de seguros agrícolas:* se contactaron 29 empresas, sólo 3 facilitaron información (SanCor, La Dulce y La Segunda).
- b)** *Reportes en medios de comunicación y redes sociales:* se recolectaron 226 reportes en 63 medios de comunicación y redes sociales.
- c)** *Redes de informantes como:*

- c.1)** *Red de voluntarios del Servicio*

Meteorológico Nacional: los cuales reportan eventos extremos a este organismo que se publican en un sitio web (Smn.gov.ar, 2013)

c.2) *Red de pluviómetros de la Policía de La Pampa* (APA-PLP, 2009-2013),

c.3) *Red termo pluviométrica de la Red de Información Agropecuaria Nacional (RIAN):* que releva información diaria de precipitaciones en aproximadamente 100 puntos en el área de estudio (Bellini Saibene, y. et. al. 2011, Rian.inta.gov.ar, 2013).

c.4) *Sistema Integrado de Información Agropecuaria:* que genera informes semanales de estimaciones de superficie sembrada por delegación provincial, indicando para cada cultivo el estado general, avances de labores, estado fenológico y principales adversidades (SIIA, 2013)

d) *Informantes calificados como técnicos de INTA, profesionales y asesores agropecuarios y productores*

e) *Recorridas a campo posteriores a una tormenta:* se aprovecharon datos recolectados en las recorridas mensuales de la RIAN donde se registran datos referentes al estado y evolución de los

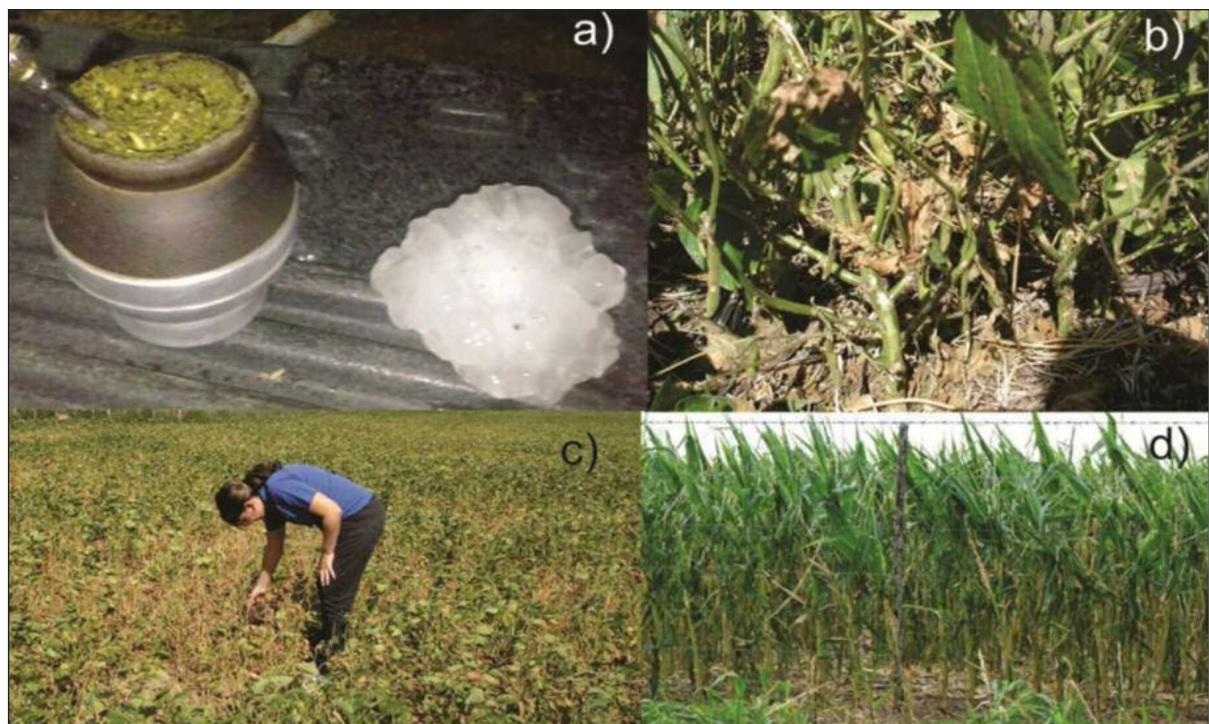


Figura 1. a) Granizo caído el 24-12-2012 en zona rural deslinde entre La Pampa y Córdoba. b) Daño en lote de soja tormenta del 10-12-2012. c) Lote de soja con daño de granizo tormenta del 01-03-2013, d) Lote de maíz con daño tormenta del 10-12-2012 (Bellini Saibene Y., 2015).

cultivos y se incluyen las principales adversidades de origen ambiental como granizo, heladas y sequías entre otras (Bellini Saibene, Y. et. al., 2011). También se registran salidas posteriores a tormentas realizadas especialmente para analizar el daño en terreno de dichas tormentas. Estas recorridas se realizaron por rutas y caminos vecinales, acompañados por productores y contratistas locales, los cuales guían en el recorrido y permiten el acceso a los lotes con daño en sus establecimientos agropecuarios. Durante el recorrido se georreferencian los lotes y se toma información de cultivo dañado, porcentaje de daño y tareas a realizar en los lotes con daño por granizo (ej.: resiembra, picado, etc.). En la figura 1 se presentan fotos de lotes dañados por granizo, recolectados en diversas salidas a campo (Bellini Saibene Y., 2015).

Desde estas fuentes se obtuvo información de: fecha del evento, hora del evento (inicio y fin), ocurrencia de granizo (si, no), posición geográfica (latitud y longitud), cultivo (uso del suelo en caso que no tenga un cultivo), estado fenológico (mínimo: Vegetativo, Reproductivo, Madurez, de ser posible nombre correspondientes de acuerdo a escalas fenológicas definidas para cada cultivo por la RIAN), daño y área afectada. Toda esta información está georeferenciada con coordenadas geográficas Datum WGS84.

A partir del dato de latitud y longitud presente en los datos de campo, se busca en las imágenes del radar el pixel que corresponde a cada punto y se extraen los valores de cada una de las tres variables. De esta manera a cada punto de campo se pueden aparear los datos registrados por el radar en esa localización.

Estos valores se almacenan en la base de datos

diseñada y se realizaron cálculos de valores máximos, mínimos, promedios, totales y cantidades para cada elevación e integrando las 12 elevaciones. También se integraron los datos a nivel diario. Estos cálculos se realizaron por medio de consultas Transact Structure Query Languaje (T-SQL) y quedaron disponibles en la base de datos por medio de Vistas.

La base de datos se implementó en SQL Server 2008 Express con extensión para el tratamiento de datos geográficos. El sistema de información se desarrolló en la plataforma web ASP.NET, utilizando Visual Basic .NET como lenguaje de programación (Bellini Saibene, Y. et. al., 2015b).

RESULTADOS Y DISCUSIÓN

La figura 3 muestra el Diagrama de Entidad Relación (DER) del modelo de datos diseñado (Bellini Saibene, Y. et. al 2015c). La figura 4 presenta la implementación de dicho modelo con el detalle de los atributos de cada tabla y relación. El modelo es muy sencillo con 9 tablas (nombre en negrita y color blanco, los atributos subrayados indican clave principal).

Las tablas de datos del radar (inician su nombre con la palabra *datos*) tienen una estructura similar:

- *IdLocXEvento*: contiene el identificador único de la relación que une los lotes identificados a campo con un evento de tormenta severa.
- *Elevación*: un número del 1 al 12 que indica en qué ángulo de elevación se registró el dato del radar.
- *Horario*: un dato tipo hora que almacena el horario en que se realizó el registro del dato del radar.
- *VolCompleto*: campo binario que vale uno en

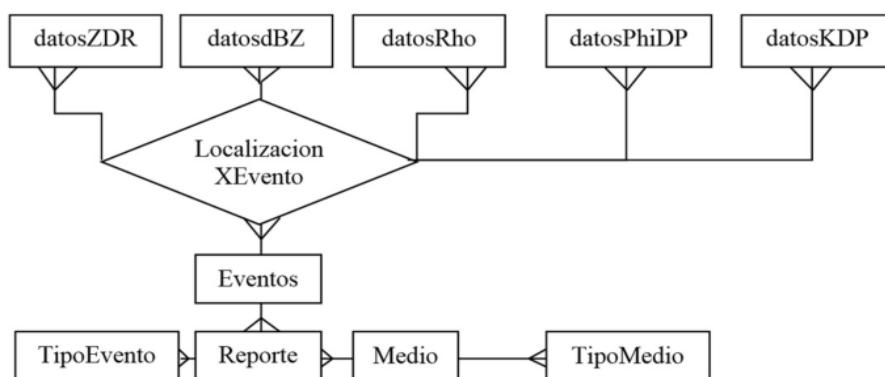


Figura 3. Diagrama de entidad relación de la base de datos (Bellini Saibene, Y. et. al. 2015c)

caso que el volumen esté completo (sin falta de datos o elevaciones) y un cero en el caso contrario.

- *ZDR, dBZ, RhoHV PhiDP, y KDP*: hacen referencia a la variable registrada. dBZ se refiere a Reflectividad (Z).

La tabla *LocalizacionXEvento* representa la relación y sus atributos son:

- *IdLocXEvento*: contiene el identificador único de la relación que une los lotes identificados a campo con un evento de tormenta severa.
- *IdEvento*: identificador único del evento severo.
- *Latitud, Longitud*: coordenadas geográficas del lote identificado.
- *Fuente*: el medio, compañía de seguro, informante calificado, red de observación, etc. desde donde se obtuvo este dato.
- *Granizo*: atributo binario que vale 1 en caso de caída de granizo y 0 en caso contrario.
- *Danio*: contiene los valores Sin Daño, Leve, Moderado, Severo y Grave.
- *PorcentajeDanio*: en caso de contar con el valor del porcentaje del daño ocasionado en el cultivo se almacena en este campo.
- *Cultivo, TipoCultivo, Fenologia*: información del uso del suelo que tiene ese lote.

La tabla *Evento* tiene los atributos:

- *IdEvento*: identificador único del evento severo.
- *fechaEvento, HoraInicio, HoraFin*: fecha y horarios de ocurrencia del evento.

La tabla *Reporte* tiene los campos

- *IdReporte*: identificador único de las noticias sobre eventos severos encontradas en los medios de comunicación
- *fechaReporte*: fecha en que se publicó la noticia
- *linkReporte*: url donde está publicada la noticia.
- *resumenReporte*: breve resumen del texto de la noticia rescatando los datos mas importantes y de interés para el sistema.
- *TipoEvento*: granizo, lluvia, viento fuerte, tormenta eléctrica o sus combinaciones,
- *IdMedio*: identificador único del medio de comunicación desde donde se sacó la noticia
- *Titulo*: título de la noticia
- *FechaEvento y HoraEvento*: fecha y hora que la noticia le asigna al evento.
- *IdEvento*: identificador único del evento severo.

Las entidades *TipoEvento* y *TipoMedio* son atributos en las tablas correspondientes.

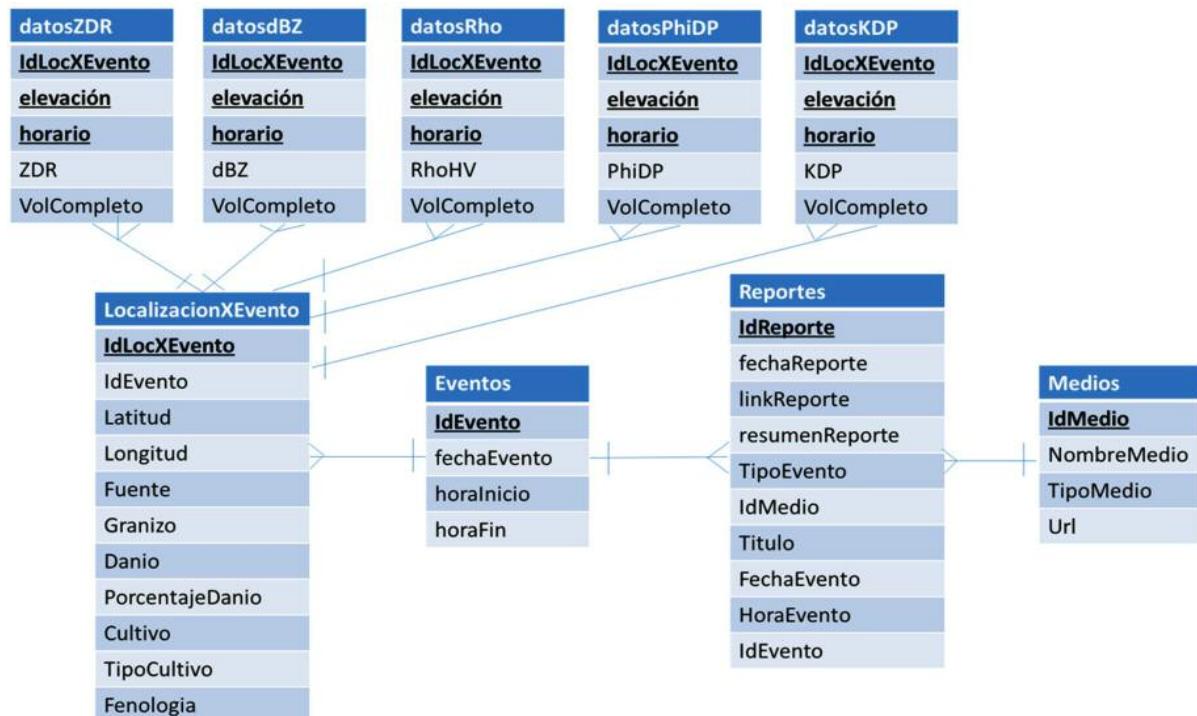


Figura 4. Implementación del DER

Tabla 1. Cantidad de registros en cada tabla de la base de datos

Tabla	Registros
datosdBZ	4.869.678
datosRhoHV	4.852.218
datosZDR	4.819.182
datosPhiDP	0
datosKDP	0
Eventos	111
Localizaciones	3.969
Medio	65
Reportes	226

La base de datos contiene 3.969 puntos geo referenciados con el detalle de ocurrencia (Si o No) de granizo en el período Marzo de 2009 a Marzo 2013, en las provincias de La Pampa, Buenos Aires, Córdoba, San Luis y Santa Fe.

El 80% de los lotes tiene información sobre intensidad y porcentaje de daño en cultivos. La tabla 1 detalla la información contenida.

Esta base se administra por medio de un sistema de información y se encuentra disponible en <http://rian.inta.gob.ar/DanioGranizo>. Las figuras 5 a 7 presentan diferentes pantallas de las consultas web disponibles para los usuarios.

Estimación de ocurrencia de Granizo en superficie

Inicio Acerca de Contacto

Estimación de ocurrencia de granizo en superficie y daño en cultivos, mediante datos del radar meteorológico utilizando técnicas de Data Mining.

UNIVERSIDAD AUSTRAL

INTA

Selección de fecha a consultar

Selecciona la fecha por la que quiere consultar:

02/03/2009

Mapa con localizaciones del evento

El mapa muestra la ubicación de cada uno de los punto donde el evento se manifestó, mostrando una nube en casa de lluvia y una nube con granizo en caso que registrar caída del mismo

Consultar >

Figura 5. Pantalla de inicio del sistema web para consulta de los datos

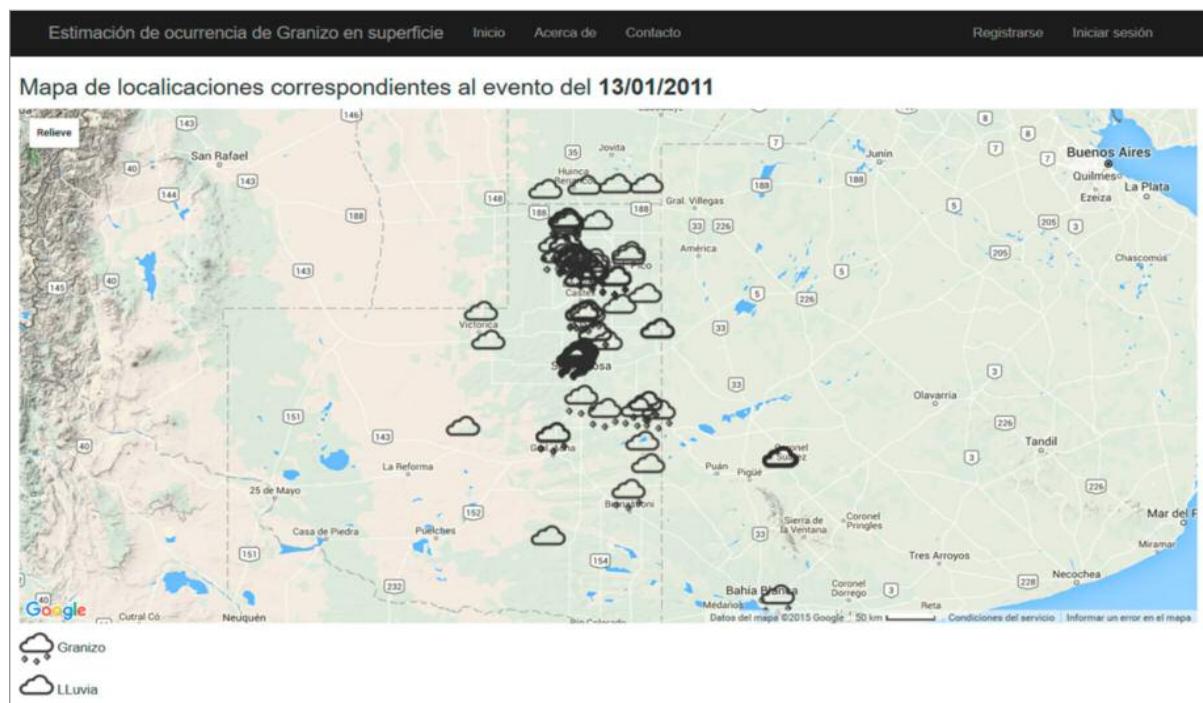


Figura 6. Ejemplo de visualización en un mapa de las localizaciones relevadas a campo (Bellini Saibene Y., 2015).

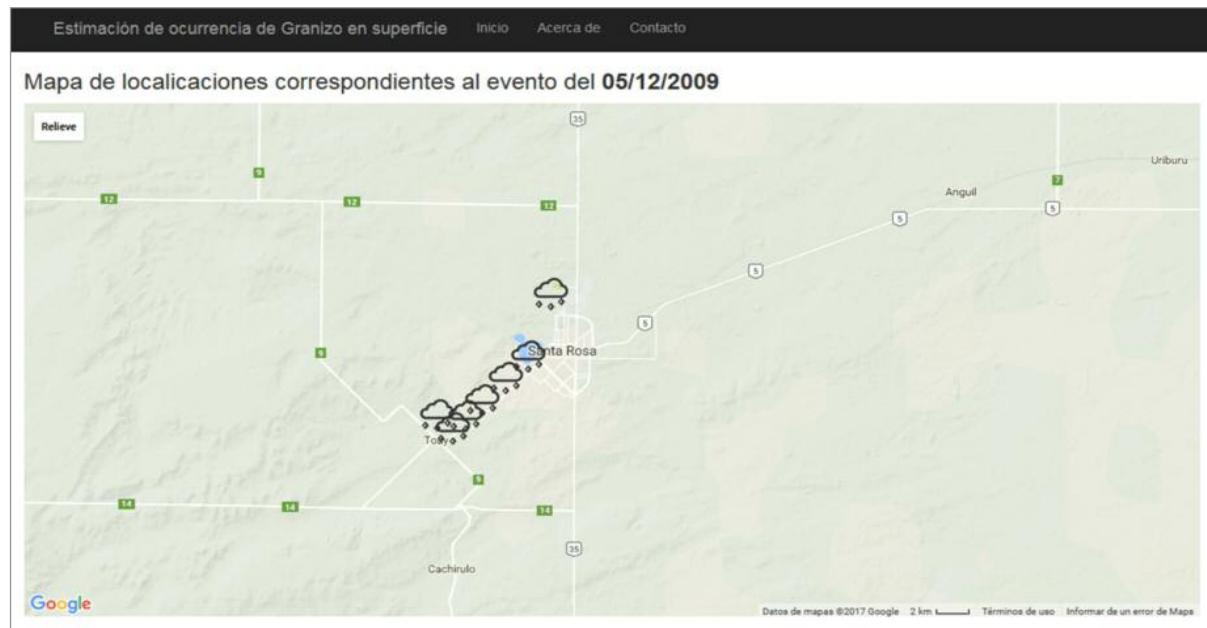


Figura 7. Zoom realizado sobre datos de campo contiguos.

CONCLUSIONES Y TRABAJOS FUTUROS

Se sistematizó y puso a disposición información de ocurrencia de granizo, daño a cultivos y variables registradas por un radar meteorológico. También se desarrolló un software que permite la actualización continua y la consulta on line de los datos registrados.

La base de datos fue utilizada para realizar diversos trabajos:

- a) análisis y caracterización de variables polarimétricas ante la presencia de granizo (Mezher et. al, 2012),
- b) generación de modelos de clasificación/estimación de ocurrencia de granizo y daño (Bellini, Saibene, Y. et. al., 2015c, Bellini, Saibene, Y. et. al., 2015d Bellini Saibene Y. 2015, Banchero, S. et. al. 2016)
- c) entrenamientos de modelos de procesamiento de lenguaje natural para clasificar noticias en medios de comunicación digitales de eventos severos de interés agropecuarios (Proyecto Ivituy, 2014)

Estos usos refuerzan la idea que la información ordenada y disponible de forma on-line es una herramienta importante para la realización de estudios sobre eventos severos en el área.

Como todos los sistemas de información la tarea

más compleja es mantenerlo actualizado, en ese sentido a partir del trabajo realizado en los proyectos regionales con enfoque territorial de INTA Anguil se actualizará esta información en la medida de las posibilidades de realizar campañas de recolección de datos a campo.

Como trabajos futuros se prevé completar los datos de las variables *Fase de propagación diferencial (PhiDP)* y *Fase diferencial específica (KDP)* del radar. También se prevé guardar una versión PDF de los reportes periodísticos como lo hace Collino et. al. 2009, ya que algunos links han cambiado o han sido dados de baja con el tiempo, perdiendo el acceso a la noticia original.

Finalmente se agregarán información de lluvias al modelo de datos y se ampliará el sistema de información y consultas web para incorporar las nuevas variables, permitir mayor interacción de los usuarios y más formatos de exportación de datos.

REFERENCIAS BIBLIOGRÁFICAS

- APA-PLP, (2009-2013). Administración Provincial del Agua - Policia de La Pampa. [online] Disponible en: <http://www.policia.lapampa.gov.ar/lluvias.php> [Accedido 2 Dic. 2013]
- Banchero, Santiago, Marcelo Soria, Romina Mezher. 2016."Big Data para el análisis de tormentas severas"

AGRANDA. ISSN: 2451-7569.pp:(119-127).

- Bellini Saibene, Yanina. 2015. "Clasificación de ocurrencia de granizo en superficie y daño en cultivos utilizando datos de radar meteorológico y técnicas de Data Mining". Tesis de Maestría. Octubre.
 - Bellini Saibene, Yanina; Schaab Lucas, Ramos Lucas, Belmonte M. L., y Fuentes M. E. 2011. "Anuario RIAN-RIAP 2009-2010," Boletín de Divulgación Técnica EEA Anguil, vol. 105, p. 54, Oct.
 - Bellini Saibene,Y.; Banchero,S.; Mezher,R.; Volpacchio,M. 2015b. "Desarrollo de software para el procesamiento y análisis de datos de la red de radares meteorológicos del INTA. CongreMet XII. 27 de mayo de 2015. Mar del Plata, Argentina. Poster.
 - Bellini Saibene,Y.; Caldera,J.M.; Volpacchio,M. 2015c. "HAR (Hail-ARchive): desarrollo de un sistema de información y base de datos sobre granizo en la región semiárida pampeana central" CongreMet XII. 27 de mayo de 2015. Mar del Plata, Argentina.
 - Bellini Saibene. Y.; Volpacchio, M. 2015d. "Clasificación de granizo en superficie usando técnicas de minería de datos y datos de radar meteorológico". 16º Argentine Symposium on Artificial Intelligence. ISSN: 1850-2784. 31/08/2015. Rosario, Santa Fe, Argentina.
 - Bustos C. y H. Videla, 1982. "Modelo estadístico de predicción de tormentas a corto plazo para la provincia de Mendoza," in Anales del XI Congreso Argentino de Meteorología. Catuogno, GA
 - Collino, Elena, Paolo Bonelli, y Luciano Gilli. 2009. «ST-AR (STorm-ARchive): A project developed to assess the ground effects of severe convective storms in the Po Valley». Atmospheric Research 93 (1-3): 483-89. doi:10.1016/j.atmosres.2008.10.021.
 - Hohl, R., H.-H. Schiesser, y Ingeborg Knepper. 2002. «The use of weather radars to estimate hail damage to automobiles: an exploratory study in Switzerland». Atmospheric research 61 (3): 215-238.
 - Katz, Richard W., y Rolando R. Garcia. 1981. «Statistical relationships between hailfall and damage to wheat». Agricultural Meteorology 24: 29-43.
 - Mezher R., Banchero S. y Bellini S.Y. Identificación de granizo con la utilización de variables polarimétricas de los radares de Paraná y Anguil, el radar de Pergamino y daño en cultivos. CongreMet XI. 28-05 al 1-06 de 2012. Mendoza, Argentina.
 - Ponce de Leon, E. 1985. "Granizo." Servicio Meteorológico Nacional.
 - Rian.inta.gov.ar. (2013). INTA - Red de Información Agropecuaria Nacional. [online] Disponible en: <http://rian.inta.gov.ar/agua/informes.aspx> [Accedido 2 Dic. 2013].
 - Rinehart, R. E. 1997. "Radar for meteorologists." Dept. of Atmospheric Sciences, Center for Aerospace Sciences, University of North Dakota.
 - Schiesser, H. H., Houze Jr, R. A., & Huntrieser, H. 1995. The mesoscale structure of severe precipitation systems in Switzerland. Monthly weather review, 123(7), 2070-2097.
 - SIIA, 2013. Sistema Integrado de Información Agropecuaria. [online] Disponible en: <http://www.siiia.gob.ar/informes> [Accedido 2 Dic. 2013].
 - Smn.gov.ar. (2013). Servicio Meteorológico Nacional. [online] Disponible en: <http://www.smn.gov.ar/?mod=voluntarios&id=1> [Accedido 2 Dic. 2013].
 - Tabary, P. , B. Fradon, A. J. Illingworth, and G. Vulpiani, 2009. "Hail detection and quantification with a C-band polarimetric radar: Challenges and promises," in 34th Conference on Radar Meteorology, pp. 5-9.
 - Visser, Petrus JM, y J. van Heerden. 2000. «Comparisons of hail kinetic energy derived from radar reflectivity with crop damage reports over the eastern Free State». Water Sa-Pretoria- 26 (1): 91-96.
-

RadINTA: un conjunto de herramientas de software libre para facilitar la explotación de datos de los radares meteorológicos del INTA

9.

*Bellini Saibene, Yanina¹; Banchero, Santiago²;
Mezher, Romina²; Caldera, Juan Marcelo¹*

¹ Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA

² Instituto de Clima y Agua del INTA

bellini.yanina@inta.gob.ar

RESUMEN

Los datos generados por la red de radares de INTA son de gran utilidad para la identificación, análisis, monitoreo y evaluación de fenómenos hidrometeorológicos, la previsión de contingencias y los pronósticos a corto plazo, entre otras aplicaciones. Dentro de los propósitos de la red se encuentra poner a disposición de todos los usuarios los datos de los radares, con el objetivo que puedan usarlos y realizar sus propios análisis y productos, incrementando el uso y aprovechamiento de la red; este acceso y procesamiento se veía dificultado por la complejidad de los datos, el costo de los software comerciales y la falta de sistemas libres que atacaran el problema de forma integral. Por estas razones, se desarrollaron herramientas que permitan minimizar el uso de software propietario para el manejo de los datos y faciliten el acceso y procesamiento de los mismos. Las herramientas desarrolladas se dividen en tres grupos de tareas: la transformación a formatos estándares y georeferenciados, la obtención de información puntual y la generación de productos. El desarrollo del software se hizo con software libre y de código abierto siendo Python 2.7 el lenguaje de programación. El uso de este tipo de software asegura la transparencia en el tratamiento de la información y la posibilidad de extender y mejorar el desarrollo realizado. Las herramientas se puede ejecutar en plataforma Linux o Windows y permite el uso de las base de datos relacionales. Se utilizó exitosamente en tesis de maestría y diversos trabajos académicos. En el repositorio <https://github.com/INTA-Radar> se encuentran todas las herramientas desarrolladas y las instrucciones de instalación y uso.

Palabras clave: SIG, sector agropecuario y agroindustrial, eventos severos, procesamiento de datos

ABSTRACT

The weather radar network of the National Institute of Agricultural Technology (INTA) is a highly useful tool for applications such as the study and evaluation of hydro-meteorological phenomena, monitoring and forecasting contingencies and short-term forecasts, among others. One of the network aims is make available to all users the radar data, with the objectives that they can use them and make their own analysis and products, increasing the use of the network. This access and processing was hampered by the complexity of the data, the cost of commercial software and the lack of free systems that attacked the problem in an integral way. For these reasons, tools were developed to minimize the use of proprietary software for data management. and facilitate access and processing data for future studies.

The tools developed are divided into three groups of tasks: the transformation geo-referenced standards and formats, obtaining timely information and product generation. Software development was done with Python 2.7

The use of open source software ensures transparency in the treatment of information and the possibility of extending and improving the development carried out. The tools can be run on Linux or Windows platform and allows the use of relational databases. It was successfully used in several master's theses and academic papers. The developed tools can be downlo-

Key words: GIS, agroindustrial and agricultural sector, data processing, severe weather

INTRODUCCIÓN

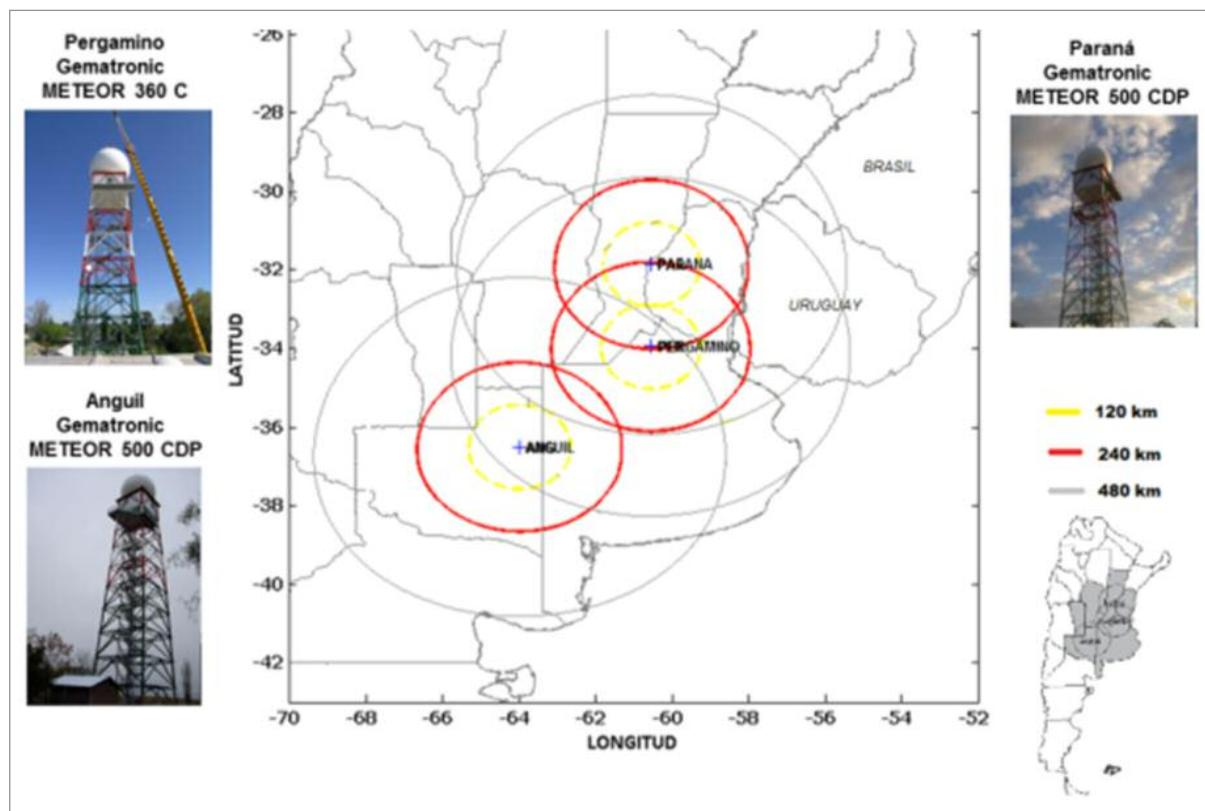
La red de INTA cuenta con tres radares meteorológicos ubicados en las Estaciones Experimentales Agropecuarias Paraná (Entre Ríos), Pergamino (Buenos Aires) y Anguil (La Pampa) cubriendo aproximadamente 46 millones de hectáreas (figura 1) y que está operativa desde el año 2007.

Los radares meteorológicos son sensores remotos y funcionan girando una antena 360 grados en forma horizontal y en una serie de ángulos de elevación vertical mientras emiten a la atmósfera pulsos de energía electromagnética en la frecuencia de las microondas. Cuando dichos pulsos encuentran una partícula (gotas de agua, cristales de hielo, polvo, insectos y otros objetos), parte de esa energía es absorbida por la partícula y el resto se dispersa en todas direcciones, devolviendo una fracción hacia el radar (Reinhart 1992); este eco se registra y a partir

del mismo se generan las variables del radar.

Los datos generados por esta red de radares son de gran utilidad para diversos aplicaciones, entre las que se pueden mencionar: a) Descripción del estado del tiempo, generación de pronósticos a mediano y corto plazo (Nowcasting), b) Previsión y monitoreo de contingencias ambientales (granizo, lluvias torrenciales, tormentas severas, entre otros.), c) Seguridad en la navegación y aeronavegación, d) Estudios de física de la atmósfera, e) Estudios de riesgos agroclimáticos, f) Suministro de datos básicos para la investigación científica y tecnológica, y g) Suministro de datos de entrada para modelos hidrológicos (ej: inundaciones) (Divan *et al.* 2015).

Dentro de los propósitos de la red se encuentra poner a disposición de todos los usuarios los datos de los radares, con el objetivo que puedan usar dichos datos y realizar sus propios análisis y productos, incrementando el uso y aprovechamiento de la red (Bellini Saibene *et al.* 2014); sin embargo, varios obstáculos técnicos dificultan su acceso y uso, por un lado los datos de los radares de INTA se presenta en un formato complejo, propietario y binario, que requiere una gran cantidad de conocimientos en su manejo, incluso en el uso de productos de soft-



ware comercial (Heistermann, Jacobi, y Pfaff 2013; Bellini Saibene *et al.* 2014), el cual suele ser además de un alto costo y sin la posibilidad de modificación de su código fuente (Bellini Saibene *et al.* 2014); por otro lado la falta de experiencia en el trabajo con coordenadas polares, en la georreferenciación de los datos de exploración de tres dimensiones y la falta de herramientas de visualización espacial con excepción de los entornos de trabajo de Sistemas de Información Geográfico (SIG) (Heistermann, Jacobi, y Pfaff 2013).

Además, dado el amplio alcance de la información que los radares meteorológicos pueden proporcionar, sus usuarios tienen una amplia gama de necesidades, desde la visualización de los datos básicos hasta el desarrollo de complejos procesamientos y generación de productos (Heistermann, Jacobi, y Pfaff 2013; Helmus y Collis 2016).

Por otra parte, algunas herramientas de procesamiento de datos disponibles libremente, como TITAN (Dixon y Wiener 1993) o la iniciativa Hydro-NEXRAD (Krajewski *et al.* 2011), se circunscriben a resolver solo una parte del problema, en el primer caso solo procesan una de las variables “crudas” que registran los radares y en el segundo solo maneja datos de la red de radares NEXRAD de Estados Unidos.

Por lo antes descripto, a partir del 2010, se inicia el desarrollo de herramientas que permitan minimizar el uso de software propietario para el manejo de los datos de los radares de la red de INTA y faciliten el acceso y procesamiento de los mismos. En un principio se inició con la lectura y extracción de información de una serie de productos generados por el software de los radares (Hartmann,

Tamburrino, y Bareilles 2010) y posteriormente se avanzó en el desarrollo de la lectura de datos crudos y la generación de productos propios (Bellini Saibene *et al.* 2014).

METODOLOGÍA

Las herramientas desarrolladas tienen como objetivo realizar tareas de procesamiento y análisis de datos planos, que no son realizadas por el radar (figura 2); se dividen en tres grandes grupos: la transformación de los datos a formatos estándar (ASCII y GeoTIFF), la obtención de información puntual y la generación de productos, todos con su componente espacial.

Tecnología utilizada

El desarrollo de las herramientas se basa enteramente en software libre y de código abierto, lo que asegura la transparencia en el tratamiento de la información y la posibilidad de extender y mejorar el desarrollo realizado. El lenguaje principal de implementación utilizado fue Python 2.7 (<http://python.org>) que es un lenguaje de alto nivel, código abierto, con librerías científicas utilizadas mundialmente, bien documentado, con sintaxis sencilla y por ende fácil de aprender. Las librerías externas utilizadas fueron: PyQt4, Numpy, lxml, Gdal y Pyodbc. También permite utilizar librerías en otros lenguajes como C y es independiente de la plataforma, por lo que el software puede funcionar tanto en entornos Linux como Windows.



Figura 2. Tareas de procesamiento, corrección y análisis de datos de información de radares meteorológicos.

Los radares y sus datos

Los radares de INTA son de industria alemana, de marca Gematronik, operan en banda C a una frecuencia de 5,64 Ghz y longitud de onda de 5,4 cm y poseen sistema doppler. Los radares de Anguil y Paraná son de doble polarización mientras que el de Pergamino es de simple polarización (figura 1). Sus antenas permiten un giro en el sentido horizontal (azimut) de 360º y puede elevarse en ángulo vertical hasta 45º. Estos radares están configurados para 12 ángulos de elevación, entre 0,5º de base y 15,1º de tope, en rangos de 120 km, 240 km y 480 km con una resolución espacial de 1 km² (Bellini Saibene *et al.* 2014). La frecuencia de este escaneo completo está programada cada 10 minutos, totalizando, bajo funcionamiento normal, 144 adquisiciones diarias (el día del radar se extienden de las 00:00 h a las 23:50 h.) registrando los datos planos o “variables” (tabla 1) y generando datos procesados o “productos”. Cada escaneo, variable y rango de alcance genera un archivo individual llamado volumen.

Una de las formas más utilizadas de mostrar los datos del radar es por medio de imágenes (Reinhart 1992), una de las más comunes es la Plan Position Indicator (PPI- Indicador de posición en un plano), la cual se arma cuando durante el escaneo el radar tiene su elevación fija y varía su ángulo de acimut (Reinhart 1992).

Lectura y transformación

Los volúmenes y se caracterizan por una sección XML (Extensible Markup Language) que almacena datos del contexto de la toma de datos y una sección de datos en formato binario comprimido que contiene el dato plano y también presenta una estructura al estilo XML, situación impone una restricción al uso de herramientas estándares para manipular XML (Hartmann, Tamburrino, y Bareilles 2010).

Diversos autores (Heistermann, Jacobi, y Pfaff 2013; Helmus y Collis 2016; Bellini Saibene *et al.* 2014) definen el formato propietario y binario de los datos de radar como una de las mayores barreras para el aprovechamiento de los mismos. A su vez la georreferenciación, que incluye la conversión desde coordenadas polares a coordenadas geográficas, es una tarea particularmente importante, porque la única referencia geográfica que se provee con los datos planos es la ubicación del radar (latitud, longi-

Tabla 1. Variables planas registradas por los radares según distancia.

Variable	120 km	240 km	480 km
Factor de reflectividad (Z-dBZ)	X	X	X
Coeficiente de correlación polarimétrica (RhoHV)	X	X	
Reflectividad diferencial (ZDR)	X	X	
Desplazamiento de fase diferencial específica (KDP)	X	X	
Desplazamiento de fase diferencial (PhiDP)	X	X	
Desplazamiento de fase diferencial sin correcciones (uPhiDP)	X	X	
Velocidad radial (V)	X		
Anchura del espectro (W)	X		
Elevaciones	12	12	1
Total archivos (1 día)	1152	864	288

tud y altitud).

Para realizar la tarea de transformación al estándar ASCII (American Standard Code for Information Interchange) el proceso por lotes *Proc-vol* recorre un directorio con archivos de datos crudos del radar y los pasa como parámetro al proceso *Vol.py*. Este script lee los volúmenes, los descomprime, transforma las coordenadas polares a geográficas, realiza el cálculo del dato “crudo” de las variables *dBZ*, *PhiDP*, *KDP*, *ZDR* y *RhoHV* de acuerdo a lo indicado en y genera un archivo de texto por cada elevación con el formato: latitud, longitud y valor de la variable (figura 3).

Estos archivos ASCII son la base para generar los archivos GeoTIFF, el segundo estándar. GeoTIFF es un formato de intercambio basado en Tagged Image File Format (TIFF) para imágenes raster georeferenciada. Es un estándar de metadatos y es ampliamente soportado por aplicaciones SIG y de manipulación de imágenes. Este formato raster representa los datos del radar en imágenes de 1 km² de resolución, en coordenadas geográficas (latitud y longitud) y Datum WGS84.

Para generar las imágenes se utiliza *Proc-todo* que recorre el directorio de los archivos ASCII y los pasa como parámetro al proceso por lotes *Procesar*. Este algoritmo recibe cada archivo de texto y genera una imagen de 487x505 píxeles por medio de la utilidad *gdal_rasterize* de la librería GDAL (versión

Figura 3. a) Ejemplo de datos presentes en un volumen en formato binario original y b) los mismos datos transformados a ASCII (Bellini Saibene et al. 2014).

1.11.0), que convierte un archivo vectorial, en este caso de puntos, a una imagen raster (figura 4). Se generó una plantilla GeoTIFF con el sistema de coordenadas que coincide con el de los archivos ASCII, para realizar una correcta proyección de los datos (Bellini Saibene, Yanina 2015). Como el haz emitido por el radar se va ensanchando a medida que se aleja, la toma de los datos no es completamente uniforme (las muestras más alejadas abarcan un volumen

mayor a 1 km³ y las más cercanas un volumen menor), por lo que en la transformación de puntos a raster pueden quedar pixeles sin datos. Para completar toda la matriz correctamente, se ejecuta el script *Completablancos.py* el cual rellena todos los pixeles del área del radar interpolando el dato máximo (o mínimo) correspondiente (Bellini Saibene *et al.* 2014). Se genera un archivo por cada elevación, variable y momento de la toma de datos (Figuras 5 y 6).

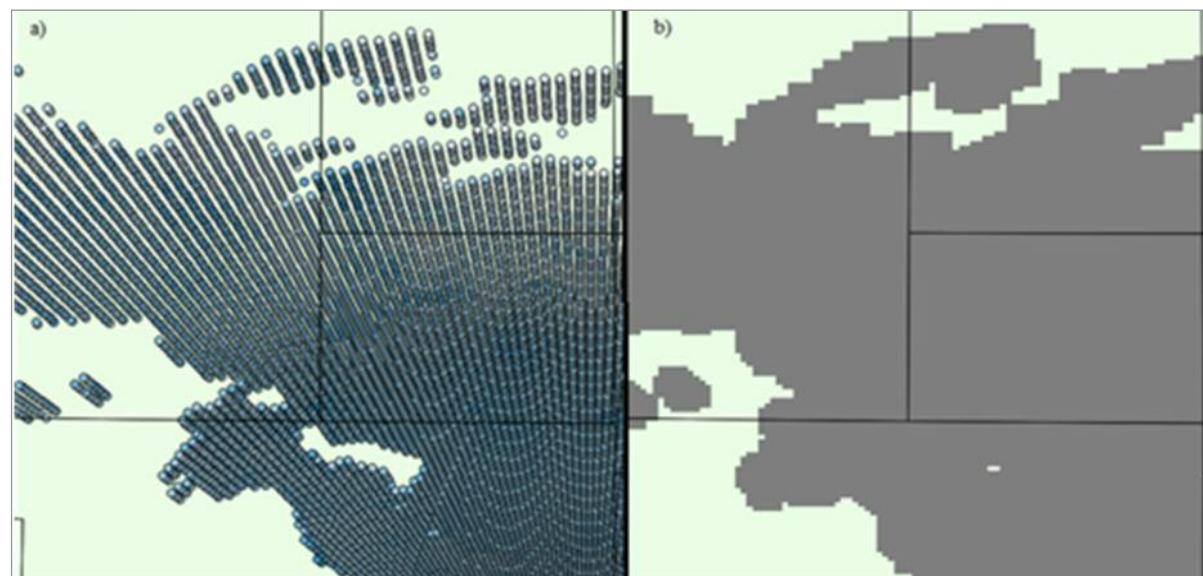


Figura 4. a) Datos del radar en formato punto como están contenido en el archivo ASCII. b) Los mismos datos convertidos a formato raster donde cada dato ocupa un pixel de 1km² (Bellini Saibene et al., 2014).

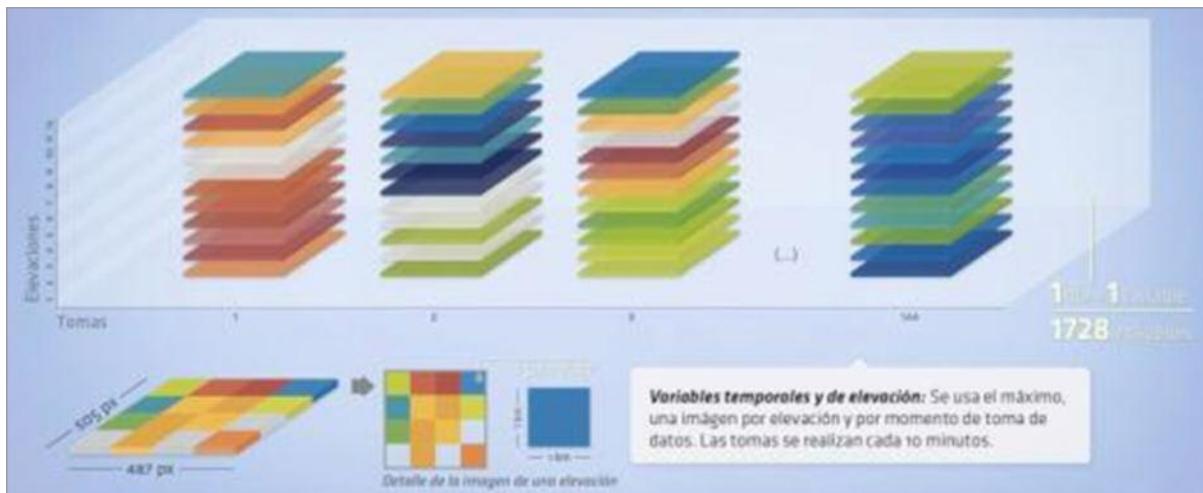


Figura 5. Esquema detalle de la generación de las imágenes correspondientes a cada elevación, cantidad de archivos y resolución espacial de las imágenes (Bellini Saibene, Yanina 2015).

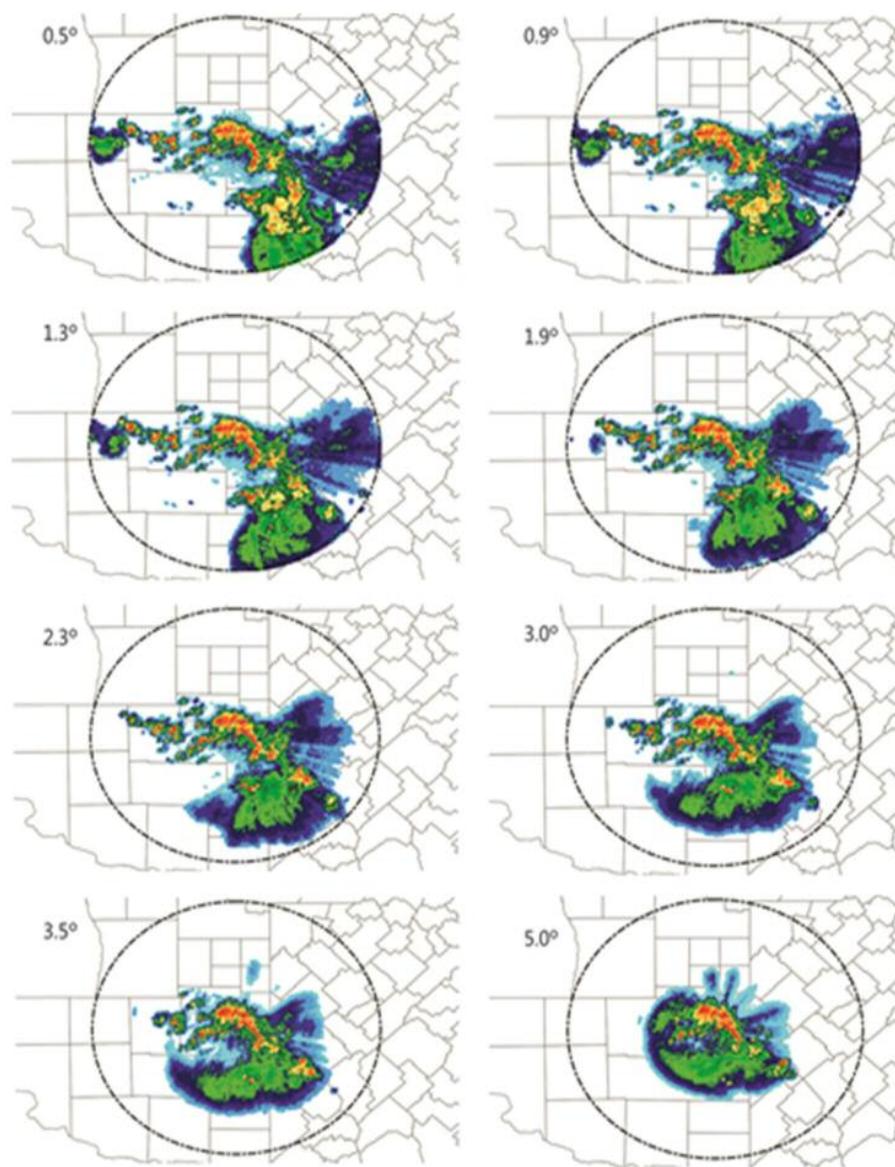


Figura 6. Indicador de Posición en el Plano (PPI) de las 12 elevaciones del 15-01-2011 a las 23.40 hs (Bellini Saibene et al. 2014).

Generación de productos de datos

Los productos del radar implican el cálculo de resúmenes con diferentes frecuencias temporales, por ejemplo diario, mensual, etc., de una o más variables. Los cálculos pueden ser máximos, mínimos, promedios, totales, cantidades, entre otros. Estos cálculos se pueden realizar utilizando todas las elevaciones disponibles, en este caso entre diez y doce, lo que se denomina perfil vertical de la tormenta o bien utilizando solo algunas. El script *GIC.py* calcula imágenes compuestas (figura 7) a partir de los argumentos indicados por el usuario (tabla 2). En la tabla 3 se presenta un ejemplo de uso.

En ocasiones los productos también utilizan variables calculadas a partir de una o varias variables planas del radar, las cuales pueden ser de interés para el cálculo de precipitaciones o la diferenciación del tipo de precipitación.

Tabla 2. Argumentos del Script *GIC.py*

Argumentos posicionales:	
Path_img	Ubicación de los archivos raster a procesar para generar la imagen compuesta
Fecha	Fecha a procesar, formato: aaaammdd
Extensión	Valor por defecto: .tif
Variable	Posibles valores: dBZ, ZDR, RhoHV, KDP, PhiDP, E, EW, EWs. Valor por defecto: dBZ
Argumentosopcionales:	
-h, --help	Muestra la ayuda
-mto	Calcula imágenes compuestas por cada paso de toma de datos (default: 10 minutos). Para todas las elevaciones.
-ele	Número de elevación a procesar. Posibles valores: 1 a 12. Si no se indica se procesan todas
-d	Indica que hace falta generar las imágenes horarias para la imagen compuesta de 24 horas y todas las elevaciones.
-maxi	Genera la imagen compuesta con el valor máximo.
-mini	Genera la imagen compuesta con el valor mínimo.
-prom	Genera la imagen compuesta con el valor promedio.
-tot	Genera la imagen compuesta con el valor total (suma).

Tabla 3. Ejemplos de código de procesamiento

Se calcula el valor promedio de todas las elevaciones de la variable dBZ del 08 de Noviembre de 2012:

```
Python GIC.py c:\volumenes tif dBZ  
20121108 -prom
```

Se calcula el valor máximo y mínimo de la primera elevación de la variable ZDR del 08 de Noviembre de 2012:

```
Python GIC.py c:\volumenes tif ZDR  
20121108 -ele 1 -mini -maxi
```

En este set de herramientas se programaron una serie de medidas que se utilizan en la identificación de granizo y en el cálculo del daño que ocasiona. Se buscó en la bibliografía disponible las ecuaciones correspondientes y su uso en Argentina para contar con datos de implementación específicos para nuestro país.

Las medidas programadas son la energía cinética que se calcula a partir de Z de acuerdo a la ecuación 1 (Waldvogel, Schmid, y Federer 1978) y HDR, que se obtiene a partir de Z y ZDR de acuerdo a la ecuación 2 (Aydin, Seliga, y Balaji 1986; Mezher, Vidal, y Salio 2011). El script *EK.py* y *HDR.py* realizan los cálculos de cada variable derivada, respectivamente.

$$E = \left(\sum_{i=\text{base}}^{i=\text{topo}} 5 \times 10^{-6} \times 10^{0.084 \times Z_i} \times W(Z_i) \right) \cdot t \quad (1)$$

Donde:

$i = 1 \text{ a } 12 \text{ elevaciones}$

$t = 600 \text{ segundos}$

$$W(Z) = \begin{cases} 0 & \text{para } Z \leq 40 \\ \frac{Z - 40}{40 - 55} & \text{para } 40 < Z < 55 \\ 1 & \text{para } Z \geq 55 \end{cases}$$

$$H_{DR} = Z - f(Z_{DR}) \quad (2)$$

Donde:

$$\begin{aligned} f(Z_{DR}) &= 21 \text{ dB} & \text{si } Z_{DR} \leq 0 \text{ dB} \\ &= 19 Z_{DR} + 27 \text{ dB} & \text{si } 0 < Z_{DR} \leq 1.74 \text{ dB} \\ &= 60 \text{ dB} & \text{si } Z_{DR} > 1.74 \text{ dB} \end{aligned}$$

También se ha desarrollado el script *ModeloGranizo.py* que implementa el modelo de estimación de caída de granizo en superficie ex pos

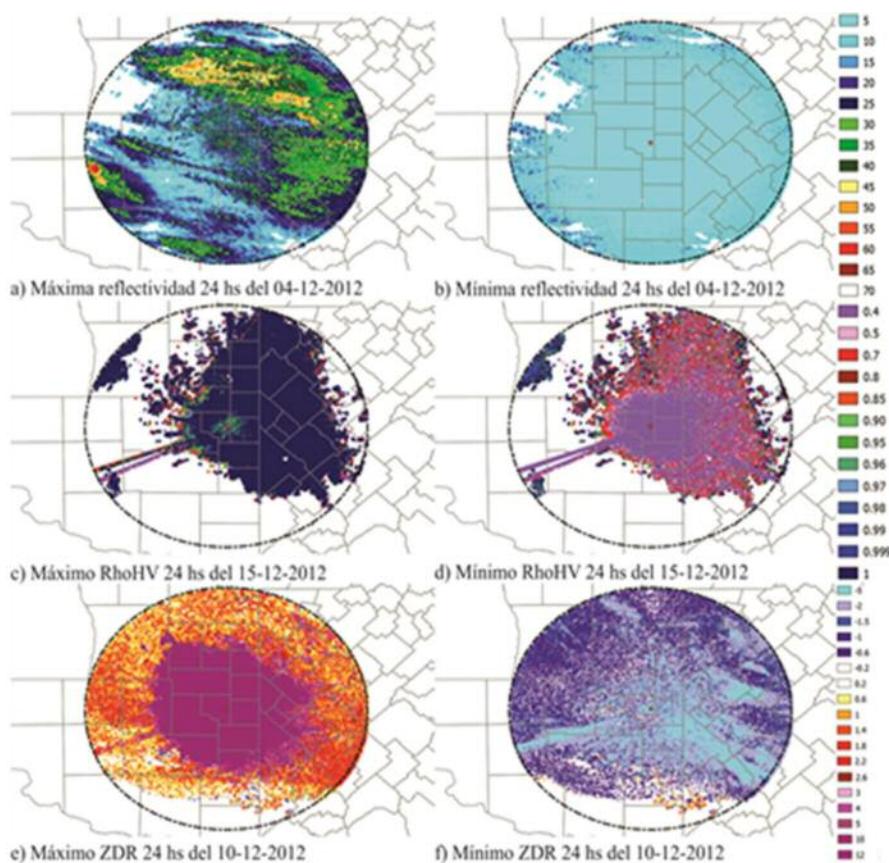


Figura 7. Ejemplos de las imágenes compuestas de la primera elevación (0.5°) y 24 horas generadas con el script GIC.py para tres fechas con tormenta del 2012 (Bellini Saibene, Yanina 2015)

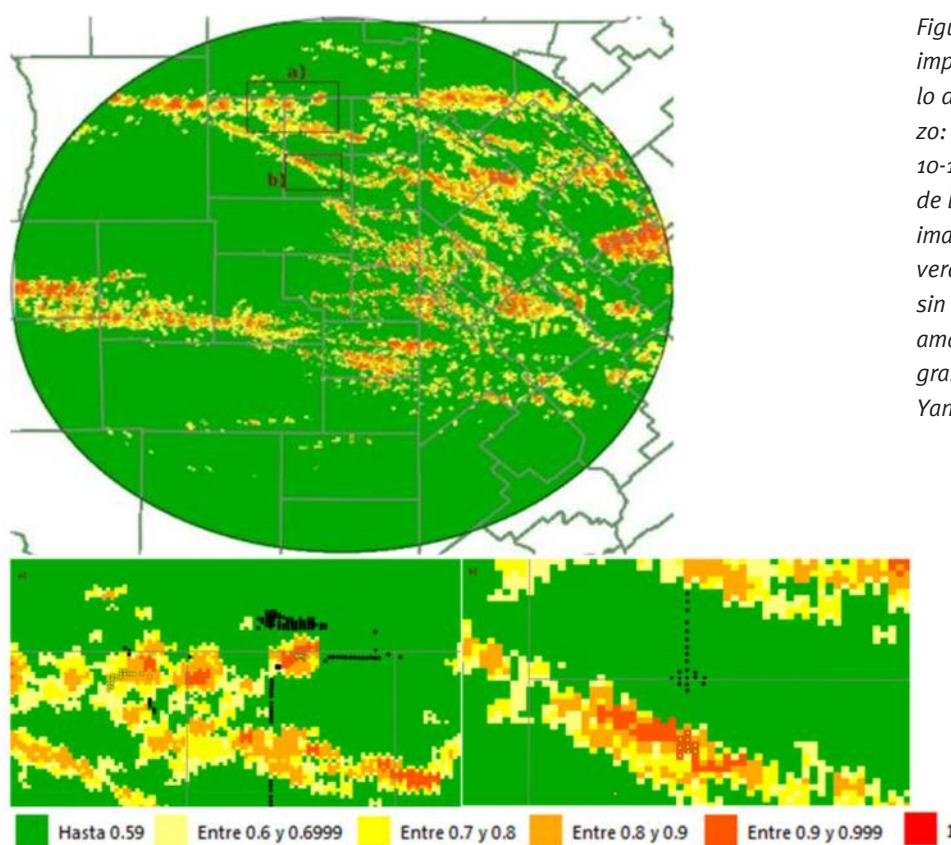


Figura 8. Resultados de la implementación del modelo de clasificación de granizo: imagen completa del 10-12-2012. a) y b) Detalle de la clasificación de la imagen y los puntos de verdad de campo: negro sin registro de granizo, amarillo con registro de granizo (Bellini Saibene, Yanina 2015).

a la ocurrencia de una tormenta, desarrollado por (Bellini Saibene, Yanina 2015). Este modelo genera como resultado una imagen clasificada indicando la probabilidad de caída de granizo en superficie posterior a una tormenta (figura 8).

Obtención de información puntual

En algunos análisis, se necesita conocer el valor de las variables del radar o de los productos y resúmenes realizados, en una serie de localizaciones específicas en vez de toda la imagen completa. Para obtener este tipo de dato puntual se generó el script *PuntosfromRaster1ominutos.py* que obtiene los valores del pixel correspondiente a cada par de coordenadas geográficas (latitud y longitud) pasadas como parámetros, desde las imágenes de las tomas cada 10 minutos y *Datos1ominutosResumen.py* que los obtiene desde las imágenes resúmenes o compuestas generadas como productos (Ej: imagen de resumen diario, o imagen de una toma de 10 minutos, con el resumen de las 12 elevaciones). La información recuperada se puede grabar en una base de datos relacional (soporta PostgreSQL y SQL Server Express) o en un archivo ASCII. Utiliza el script *Identify.py* el cual convierte las coordenadas geográficas a coordenadas en pixel de la imagen raster a procesar y devuelve el valor correspondiente al pixel determinado. Esta funcionalidad permite analizar la información de casos de campo, apareando los valores registrados por el radar con eventos ocurridos en superficie.

RESULTADOS Y DISCUSIÓN

Se desarrollaron un set de herramientas que permiten leer, transformar y procesar los datos de los radares de INTA (figura 9). Estos programas son software libre y puede ser redistribuido y/o modificado bajo los términos de la licencia GNU General Public License, como está publicada en la Free Software Foundation. En el repositorio <https://github.com/INTA-Radar> se encuentran todas las herramientas desarrolladas y las instrucciones de instalación y uso, tanto en plataforma Linux como Windows.

En las primeras versiones del software (Bellini Saibene *et al.* 2014), se contemplaba un script para descarga de los volúmenes directamente desde los

radares llamado *ftpRADAR.py* (figura 9), este script incluía información de acceso a la red de comunicaciones, lo que implicaba un riesgo de seguridad. Además, Esta descarga se realizaba del servidor donde el radar almacena la información en tiempo real, por lo que solo tiene disponible una serie de meses. Esto impide acceder de esta manera a toda la información histórica, la cual debe ser solicitada vía mail para poder obtenerla. Para resolver estas dos situaciones, se está trabajando en la generación de un repositorio de grandes datos de acceso libre para agilizar el acceso y uso de los volúmenes (Divan *et al.* 2015).

El uso de estándares para la representación de los datos, permite que los mismos puedan ser utilizados por una variedad de software también libre presente en el mercado, por ejemplo analizar las imágenes GeoTIFF con QGIS para realizar geoprocесamientos o importar los archivos ASCII en paquetes estadísticos como R o de minería de datos, como Weka o RapidMiner, para su análisis. El software ha sido utilizado en tesis de maestría (Bellini Saibene, Yanina 2015), en diversos trabajos académicos, en convenios de trabajos de INTA con Universidades (Ej: Universidad Nacional de La Pampa), y ha sido solicitado por compañías de seguros agrícolas nacionales (Ej: La Segunda y SanCor) y por universidades de otros países, como Ecuador que cuentan con el mismo tipo de radar.

Como detallan (Heistermann, Jacobi, y Pfaff 2013; Helmus y Collis 2016), la programación de funcionalidades de análisis y procesamiento de los datos, implica la búsqueda en la bibliografía de algoritmos probados, los cuales son abundantes en aquellos países que tienen una historia en el uso de radares meteorológicos, pero no es tan cuantiosa en nuestro país, por lo que hay mucho trabajo por delante en la adecuación de estos algoritmos a las zonas de influencia de la red de INTA.

También hay un interesante campo a desarrollar en la búsqueda de nuevos algoritmos generados con herramientas como la minería de datos y el procesamiento de flujos de datos (Bellini Saibene, Yanina 2015; Divan *et al.* 2015) y el aprovechamiento de los productos para aplicaciones en industrias como la agrícola y de los seguros.

Durante el desarrollo de este set de herramientas libres, durante el 2013 se liberaron dos softwares para el tratamiento de datos de radares meteorológicos, el primera es una librería para Python ll-

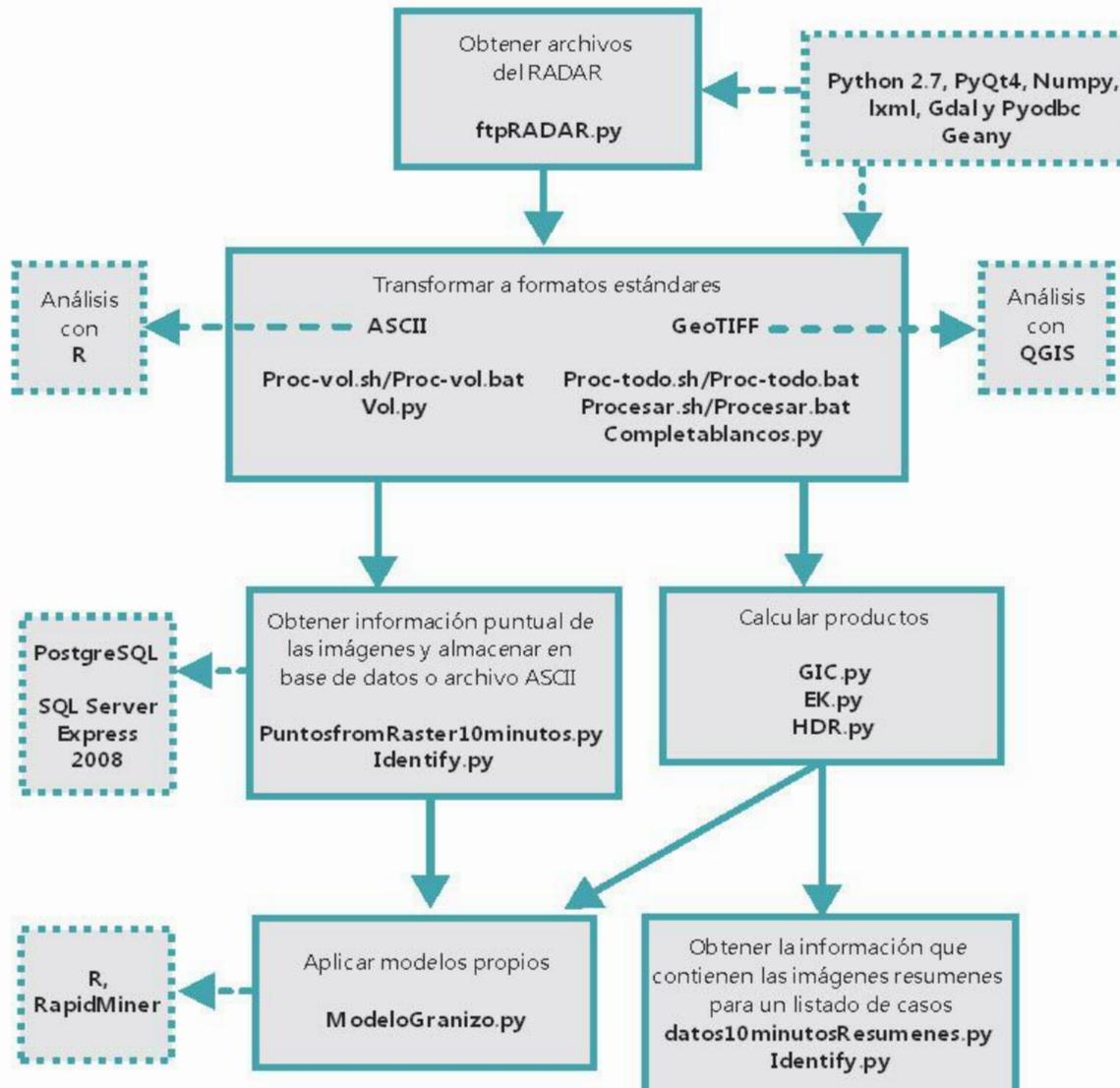


Figura 9. Esquema de las herramientas desarrolladas y software utilizado.

mada wradlib (Heistermann, Jacobi, y Pfaff 2013) que tiene como objetivo proveer un software que permite generar flujos de procesamiento complejos y la segunda es un paquete para la lectura, visualización y análisis de datos de radares, también realizada en Python, llamado Python ARM Radar Toolkit (Py-ART) (Helmus y Collis 2016). Ambas herramientas están muy bien documentadas y presentan un interesante número de funcionalidades ya desarrolladas, además que sus autores ven las mismas como complementarias (Heistermann, Jacobi, y Pfaff 2013). En el caso de wradlib, durante el 2015 incorporó la lectura del formato de los radares de INTA. Como estas librerías cuentan con una comunidad de desarrollo que las respalda y se nota un incre-

mento en su uso, se considera incorporar funcionalidades de RadINTA a las mismas, para contribuir a un marco de trabajo común en la comunidad científica. RadINTA cuenta con desarrollos de funcionalidades relacionadas con el uso de la doble polarización y la discriminación de hidrometeoros, que wradlib aún tiene que implementar (Heistermann, Jacobi, y Pfaff 2013) y en el caso de Py-ART, aún no cuenta con la facilidad de leer el formato de los radares de INTA (Helmus y Collis 2016), que se podría incorporar.

Igualmente, RadINTA presenta un set de herramientas listas para usar por un usuario sin necesidad de programar, mientras que las dos herramientas anteriores necesitan conocimientos de progra-

mación en Python para poder aprovechar su uso. Además, los algoritmos se encuentran configurados con implementaciones nacionales de los parámetros necesarios para su funcionamiento.

Finalmente, disponer de este tipo de programas con la documentación en castellano, aumenta las posibilidades de uso por usuarios no científicos/académicos, como las compañías de seguro.

CONCLUSIONES

El desarrollo de RadINTA permitió el uso de los datos de la red de INTA sin necesidad del uso de software propietario. Como el desarrollo se realizó con herramientas libres, se asegura la transparencia en el tratamiento de los datos, al contar con el código fuente de los diferentes algoritmos, adaptados a la realidad local en caso de implementaciones disponibles y dando la posibilidad de extender y mejorar el desarrollo realizado. Para los usuarios científicos brinda además una herramienta que propicia la reproducibilidad de los análisis y procesamientos realizados.

RadINTA brinda herramientas a los usuarios de los datos de la red permitiendo el procesamiento de los mismos, propiciando el uso de formatos estándares y abiertos de datos para facilitar la interacción tanto entre las herramientas de software elegidas como entre los usuarios que las utilizan, incrementando el uso y aprovechamiento de los radares meteorológicos del INTA.

REFERENCIAS BIBLIOGRÁFICAS

- Aydin, K., T. A. Seliga, y V. Balaji. 1986. «Remote sensing of hail with a dual linear polarization radar». *Journal of Climate and Applied Meteorology* 25 (10): 1475–1484.
- Bellini Saibene, Yanina. 2015. «Estimación de ocurrencia de granizo en superficie y daño en cultivos mediante datos del radar meteorológico utilizando técnicas de Data Mining». Magister, Buenos Aires: Austral.
- Bellini Saibene, Yanina, Martín Volpaccio, Santiago Banchero, y Romina Mezher. 2014. «Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA». En XLIII Jornadas Argentinas de Informática e Investigación Operativa (43JALIO)-VI Congreso Argentino de Agroinformática (CAI)(Buenos Aires, 2014). <http://sedici.unlp.edu.ar/handle/10915/41989>.
- Divan, Mario, Yanina Noemí Bellini Saibene, María de los Angeles Martín, María Laura Belmonte, Guillermo Lafuente, y Juan Marcelo Caldera. 2015. «Hacia una Arquitectura de Procesamiento de Datos del RADAR Meteorológico de INTA Anguil.» En Jornadas Argentinas de Informática. 44. Simposio de Informática en el Estado. 9. 2015 08-09 31-04, 31 de agosto al 4 de septiembre de 2015. Rosario, Santa Fé. AR. <http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=inta2.xis&method=post&formato=2&cantidad=1&expresion=mfn=078823>.
- Dixon, Michael, y Gerry Wiener. 1993. «TITAN: Thunderstorm identification, tracking, analysis, and nowcasting-A radar-based methodology». *Journal of Atmospheric and Oceanic Technology* 10 (6): 785–797.
- Hartmann, Tomás, M. Tamburino, y Federico Bareilles. 2010. «Análisis preliminar de datos obtenidos por la red de radares del INTA para el estudio de precipitaciones en la Región Pampeana». En 39º Jornadas Argentinas de Informáticas-2º Congreso Argentino de Agroinformática, Buenos Aires, 826. http://inta.gob.ar/sites/default/files/script-tmp-9_analisis_preliminar_de_datos_obtenidos_por_la_red_d.doc.
- Heistermann, M., S. Jacobi, y T. Pfaff. 2013. «Technical Note: An open source library for processing weather radar data (wradlib)». *Hydrology and Earth System Sciences* 17 (2): 863–871.
- Helmus, Jonathan, y Scott Collis. 2016. «The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language». *Journal of Open Research Software* 4 (1). http://openresearchsoftware.metajnl.com/articles/10.5334/jors.119/?toggle_hypothesis=on.
- Krajewski, Witold F., Anton Kruger, James A. Smith, Ramon Lawrence, Charles Gunyon, Radoslaw Goska, Bong-Chul Seo, et al. 2011. «Towards better utilization of NEXRAD data in hydrology: an overview of Hydro-NEXRAD». *Journal of hydroinformatics* 13 (2): 255–266.
- Mezher, R. N., L. Vidal, y P. Salio. 2011. «Hailstorms Analysis using Polarimetric Weather Radars and Microwave Sensors in Argentina». En 6th Eur. Conf. Sev. Storms ECSS 2011. Vol. 26082011.
- Reinhart, R. E. 1992. Radar for meteorologists. North Dakota, USA.
- Waldvogel, A., W. Schmid, y B. Federer. 1978. «The kinetic energy of hailfalls. Part I: Hailstone spectra». *Journal of Applied Meteorology* 17 (4): 515–520.

10.

Hacia una arquitectura de procesamiento de datos del RADAR Meteorológico de INTA Anguil

*Diván, Mario¹; Bellini Saibene, Yanina²;
Martín, María de los Angeles¹; Belmonte, María Laura²;
Lafuente, Guillermo¹ & Caldera, Juan Marcelo²*

¹ Facultad de Ingeniería, UNLPam, General Pico, La Pampa, Argentina

² Estación Experimental Agropecuaria Anguil “Ing. Agr. Guillermo Covas” del INTA

mjdivan@ing.unlpam.edu.ar

Este trabajo recibió el Premio 2015 a la Excelencia en Gobierno Electrónico en la categoría Proyecto otorgado por la Sociedad Argentina de Informática e Investigación Operativa en el marco del Simposio de Informática en el Estado.

RESUMEN

El Radar Meteorológico (RM) de la Estación Experimental Agropecuaria INTA Anguil produce en forma diaria un volumen de 17GB de datos, lo cual representa anualmente alrededor de 6,2 Tb. La utilización de dichos datos al momento en que son generados, como así también su posterior gestión, utilización y la posibilidad de brindar servicios al público representan un desafío por su volumen y complejidad. El enfoque integrado de procesamiento de flujos de datos centrado en metadatos de mediciones (EIPFDcMM), es un gestor de flujos de datos sustentado en un marco de medición y evaluación, el cual incorpora comportamiento detectivo y predictivo, mediante el empleo de las mediciones y metadatos asociados. Es-te trabajo propone una arquitectura de procesamiento que extiende el EIPFDcMM para incorporar el procesamiento de grandes datos almacenados. Ello permitiría dotar al RM de un comportamiento detectivo y predictivo sobre los datos en línea, como así también incluir una capa de servicios al público, que fomente el consumo de datos generados por el RM de INTA Anguil.

Palabras clave: Grandes Datos, Teledetección, Flujo de Datos, Memoria Organizacional

ABSTRACT

The Weather Radar (WR) of the Experimental Agricultural Station (EAS) INTA Anguil produces daily a volume of 17GB of data, which represents about 6.2 Tb annually. The use of such data when

they are generated, as well as its subsequent management, use and the possibility of providing services to the public represent a challenge in terms of volume and complexity. The Strategy for Data Stream Processing based on Measurement Metadata (SDSPbMM) is a data stream manager sustained in a measurement and evaluation framework, which incorporates detective and predictive behavior, through the use of measurements and associated metadata. This paper proposes a processing architecture that extends the SDSPbMM to incorporate the processing of big data. This would provide the WR of a detective and predictive behavior on online data, as well as include a layer of public services, which encourages the consumption of data generated by the WR of INTA Anguil.

Key words: Big Data, Remote sensing, Data Stream Processing, Organizational Memory

INTRODUCCIÓN

Actualmente, existen aplicaciones que procesan un conjunto de datos a medida, generados en forma continua, a los efectos de responder a consultas y/o adecuar su comportamiento en función del propio arribo de los datos (Gehrke, *et al.*, 2008), como es el caso de las aplicaciones para el monitoreo de signos vitales de pacientes; del comportamiento de los mercados financieros; entre otras. En dicho tipo de aplicaciones, se enmarca el Enfoque Integrado de Procesamiento de Datos centrado en Metadatos de Mediciones (EIPFDcMM) (Divan, *et al.*, 2011; Divan y Olsina, 2014), el cual sustentado en el marco de

medición y evaluación C-INCAMI (Context-Information Need, Concept Model, Attribute, Metric and Indicator) (Olsina y Molina, 2007, Papa *et al.*, 2007), incorpora metadatos al proceso de medición, promoviendo la repetitividad, comparabilidad y consistencia del mismo. Desde el punto de vista del sustento semántico y formal para la medición y evaluación (M&E), el marco conceptual C-INCAMI establece una ontología que incluye los conceptos y relaciones necesarias para especificar los datos y metadatos de cualquier proyecto de M&E. Por otra parte, y a diferencia de otras estrategias de procesamiento de flujos de datos (Aref *et al.*, 2005; Chandrasekaran *et al.*, 2003; Ahmad *et al.*, 2005), gracias a la incorporación de metadatos, el EIPFDcMM es capaz de guiar el procesamiento de las medidas provenientes de fuentes de datos heterogéneas, analizando cada una de ellas dentro de su contexto de procedencia, como así también su significado dentro del proyecto de M&E en el que se definió.

La Estación Experimental Agropecuaria (EEA) INTA Anguil tiene instalado un RM capaz de generar un flujo de datos de 17GB diarios aproximadamente, lo que representa un desafío para su almacenamiento, gestión y posterior servicio al público, principalmente considerando la importancia que los datos poseen para la región productiva de influencia. Por otro lado, y si bien el EIFPDcMM incorporó en su esquema de procesamiento la posibilidad de manejar grandes volúmenes de datos como memoria organizacional para el entrenamiento de sus clasificadores (Divan *et al.*, 2013) no estaba orientado al acopio de la totalidad de los datos procesados en línea y menos aún, al hecho de contar con la posibilidad de integrar la prestación de servicios de datos públicos a terceros.

Así, y como contribuciones específicas se plantea, (i) relacionado con la configuración de las fuentes de datos: la posibilidad de incorporar dentro de la arquitectura de procesamiento del EIPFDcMM como una fuente heterogénea al RM de INTA Anguil en forma transparente, lo que permite incorporar un comportamiento detectivo y predictivo, a partir del proyecto de M&E que los expertos del INTA definan sobre las medidas observadas; (ii) relacionado con la captación de medidas: la posibilidad de definir formalmente un proyecto de M&E que permite detectar eventuales desvíos del funcionamiento del RM y promover los ajustes meteorológicos pertinentes, basando los cursos de acción probables a partir

de la información disponible en la memoria organizacional; (iii) relacionado con la arquitectura de procesamiento: ahora se tiene una arquitectura actualizada para el EIPFDcMM que permite incorporar la gestión y utilización de grandes volúmenes de datos persistentes junto con el procesamiento de flujos de datos; (iv) relacionado con los servicios de datos: la arquitectura incorpora en forma transparente la posibilidad de prestar servicios de datos a terceros interesados. De este modo, los datos generados por el RM de la EEA Anguil, podrían ser consumidos en forma directa y sin mediar solicitud o intervención humana, por Centros de Investigación, Servicios Meteorológicos u otros organismos o personas.

Estas contribuciones representan un importante avance con respecto al modelo de procesamiento presentado en (Divan y Olsina, 2014; Divan *et al.*, 2013) ya que ahora se incorpora la posibilidad de gestionar grandes volúmenes de datos persistentes en forma adicional a su procesamiento-to en línea, y contar con la posibilidad de establecer capas de servicios para el eventual consumo de los mismos.

El presente artículo se organiza en seis secciones. La sección 2 resume el funcionamiento del RM y la tipología de los datos generados. La sección 3 resume el marco conceptual C-INCAMI. La sección 4 esquematiza la nueva arquitectura de procesamiento para el EIPFDcMM junto con la nueva perspectiva de capas de servicios. La sección 5 discute los trabajos relacionados, y por último, se resumen las conclusiones y trabajos a futuro.

Los RM son sensores activos de teledetección que emiten pulsos de energía electromagnética hacia la atmósfera en el rango de frecuencias de las microondas. Sus mediciones se basan, por un lado, en que la radiación electromagnética a medida que se propaga en la atmósfera es dispersada por los objetos y partículas en ella existentes, y por otro lado, en la capacidad de las antenas para emitir radiación direccional y captar la radiación incidente procedente de una dirección determinada. Estos sensores representan herramientas que permiten monitorear en forma permanente variables medio ambientales, y específicamente, la identificación, análisis, monitoreo, pronóstico y evaluación de fenómenos hidrometeorológicos, como así también de los procesos físicos que estos involucran, considerando el análisis de los riesgos que pueden ocasionar los eventos severos. Entre sus principales aplicaciones pueden mencionarse: a) Descripción

del estado del tiempo, generación de pronósticos a mediano y corto plazo (Nowcasting), b) Previsión y monitoreo de contingencias ambientales (granizo, lluvias torrenciales, tormentas severas, entre otros.), c) Seguridad en la navegación y aeronavegación, d) Estudios de física de la atmósfera, e) Estudios de riesgos agroclimáticos, f) Suministro de datos básicos para la investigación científica y tecnológica, y g) Suministro de datos de entrada para modelos hidrológicos.

Los componentes básicos de un RM son: a) Un transmisor, que genera una señal de voltaje; b) Una antena, que direcciona esta señal concentrada en un haz y recibe la radiación retrodispersada por los blancos contenidos en un volumen de muestreo; y c) Un receptor que recoge la señal de voltaje inducida en la antena cuando ésta capta radiación incidente y la amplifica para facilitar su procesamiento. La potencia recibida por un RM está relacionada con las propiedades retrodisper-soras de los hidrometeoros

contenidos en cada volumen de muestreo. Esta medida de la capacidad de los blancos de interceptar y devolver la energía del RM, se denomina reflectividad y es una medida de la intensidad de los ecos recibidos por la antena radar. La reflectividad depende de los parámetros físicos del objeto a medir, como su tamaño, forma, orientación y composición (Rinehart, 2007).

METODOLOGÍA

Radar Meteorológico de la EEA Anguil

El RM ubicado en Anguil (La Pampa), es de industria alemana marca Gematronik modelo Meteor 600C (Fig 1.a). Posee sistema doppler y es de doble polarización (DP). Opera en banda C a una frecuencia de 5,64 Ghz y longitud de onda de 5,4 cm (Gematronik, 2005).

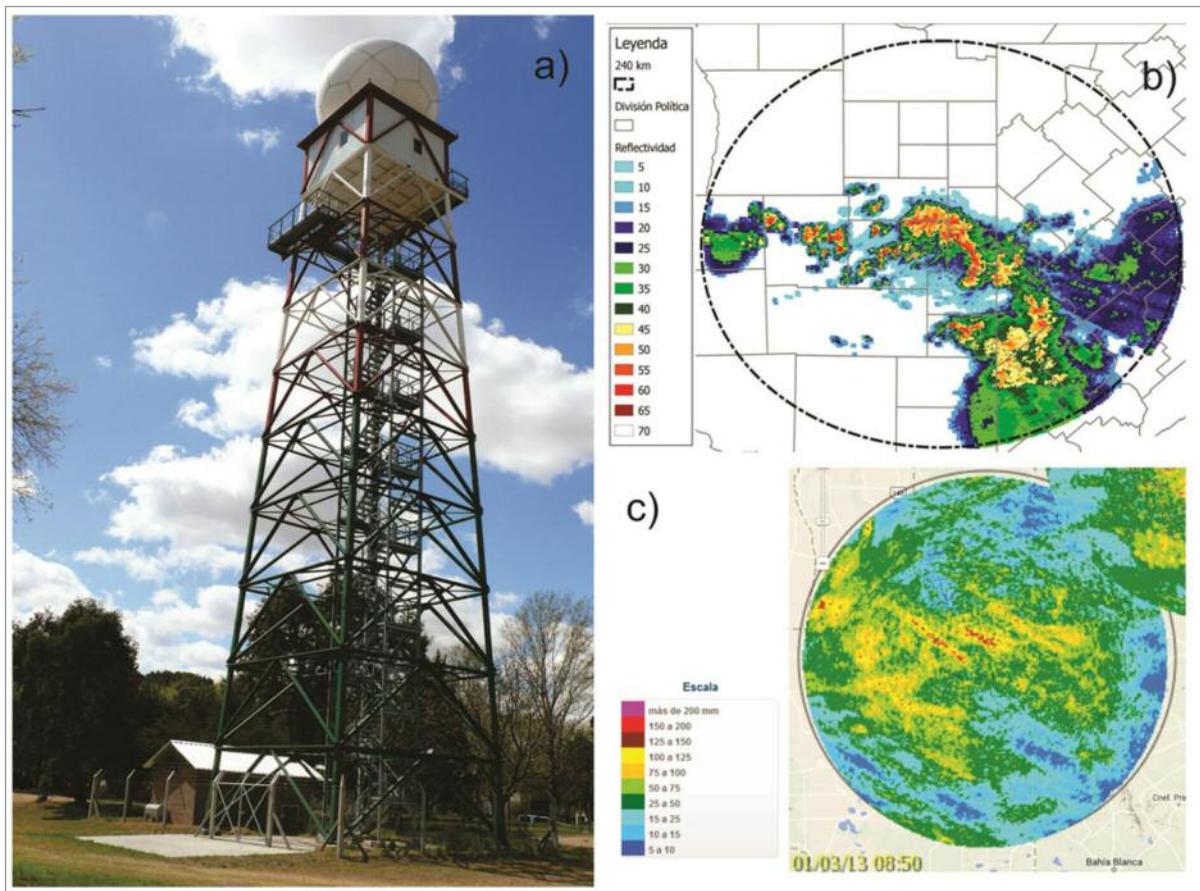


Figura 1. a) Infraestructura del RM instalado en la EEA Anguil, b) Imagen de reflectividad de la primera elevación ($0,5^{\circ}$) del 15-01-2011 a las 23:40hs generado con Software de INTA, c) PAC (Precipitation Accumulation) de Febrero de 2013 generado con software Rainbow 5 de Gematronic.

La antena permite un giro en el sentido horizontal (azimut) y puede elevarse en ángulo vertical hasta 45°. Este RM está configurado para completar una serie de giros a 360° que se repite para 12 ángulos de elevación, entre 0,5° de base y 15,1° de tope, en rangos de 120 km, 240 km y 480 km (Hartmann *et al.*, 2010) un ejemplo para la primera elevación es expuesto en la figura 1(b). La frecuencia de este escaneo completo está programada cada 10 minutos, totalizando, bajo funcionamiento normal, 144 adquisiciones diarias.

Desde el momento de su instalación en marzo de 2009, INTA cuenta con un convenio con el Servicio Meteorológico Nacional (SMN) para su mantenimiento, por lo que es factible su empleo continuado a lo largo de diferentes proyectos de investigación propios e interdisciplinarios con otras Instituciones tales como la Facultad de Ingeniería de la UNLPam, mediante comisiones de estudio y/o cooperaciones técnicas.

Recolección de Datos

La información registrada por el RM es recolectada a través de escaneos volumétricos y los datos son almacenados en archivos separados llamados volúmenes. Los datos contienen las distintas variables: factor de reflectividad (Z), reflectividad diferencial (ZDR), coeficiente de correlación polarimétrica (RhoHV), desplazamiento de fase diferencial (PhiDP), desplazamiento de fase diferencial específica (KDP), velocidad radial (V) y anchura del espectro (W). Se distinguen dos tipos de datos: a) Datos “crudos” y b) Datos con algún nivel de procesamiento o “productos”. En ambos casos, la unidad de muestreo es de 1 km² y 1°, mientras que cada variable o producto se almacena en archivos separados (Gematronik, 2005).

Las variables a registrar se configuran de acuerdo al rango, y posteriormente, se almacenan en volúmenes que contienen: el valor de la variable medida para cada elevación, y la posición de cada unidad muestreada con respecto del RM en coordenadas polares. Los volúmenes tienen un formato propietario, con una sección XML que almacena datos del contexto de la toma de datos, y una sección de datos en formato binario con compresión que contiene el dato plano, pero que también presenta una estructura al estilo XML ((Hartmann *et al.*, 2010; Gematronik, 2005) e impone una restricción al

uso de herramientas estándares para manipular este tipo de archivos. Ante esta situación, se desarrollaron un conjunto de programas que permiten descargar los volúmenes, transformarlos a dos formatos estándares (ASCII y GeoTIFF) y procesarlos (Ver figura 1.b y 1.c). De este modo, se genera un archivo ASCII por cada elevación con el formato: latitud, longitud y valor de la variable. Bajo funcionamiento normal, en un día completo (00:00 a 23:50h), se generan 8.640 archivos sólo para el rango de 240 kilómetros. Luego, a partir de los ASCII, se generan archivos raster en una matriz de 487x505 píxeles de 1 km² de resolución, en coordenadas geográficas (latitud y longitud) y Datum WGS84 (Bellini *et al.*, 2014).

A partir de datos “crudos” y mediante el software propietario Rainbow 5, se pueden obtener diferentes procesamientos, por ejemplo, algunos productos hidrológicos que estiman características de la precipitación como SRI (Surface Rain Intensity), que genera valores de intensidad o PAC (Precipitation Accumulation), que calcula un acumulado de lluvia en un intervalo de tiempo predefinido. Estos productos pueden tener formato al estilo XML o de imagen raster (Gematronik, 2005). Asimismo, INTA desarrolló un software que permite generar más productos a partir de datos crudos radar, por ejemplo aplicaciones de modelos para estimar ocurrencia de granizo y daño en cultivos (Bellini *et al.*, 2014). Así, pueden mencionarse como beneficiarios directos (o consumidores) de los datos y productos generados, con acceso libre y abierto a: i) Sistema Nacional de Radar Meteorológico (SiNaRaMe), ii) Servicio Meteorológico Nacional (SMN), iii) Subsecretaría de Recursos Hídricos de la Nación (SSRH), iv) Instituto Nacional del Agua (INA), v) Defensa Civil, vi) Fuerza Aérea Argentina, vii) Aviación comercial y general, viii) Dirección de Agricultura y Contingencias Climáticas (DACC, Mendoza), ix) Oficina de Riesgo Agroclimático (ORA), xi) Universidades, xii) Grupos de Investigación y Desarrollo de productos relacionados, xiii) Compañías de Seguros, xiv) Medios de comunicación, e xv) INTA.

Panorama de C-INCAMI

C-INCAMI es un marco conceptual [4,5] que define los módulos, conceptos y relaciones que intervienen en el área de M&E, para organizaciones de software.

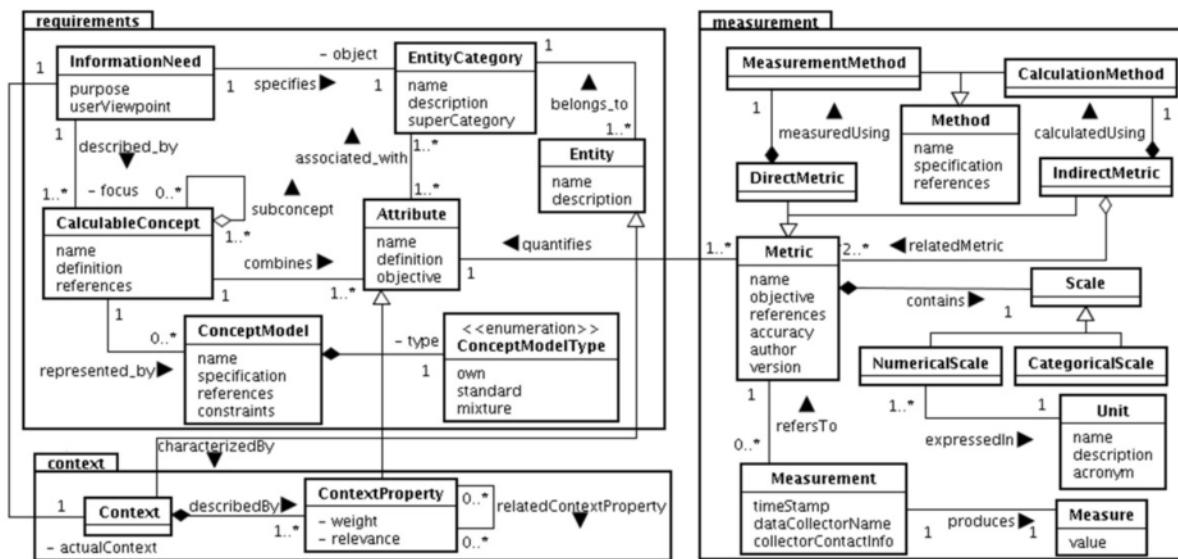


Figura 2. Principales conceptos y relaciones de los componentes Especificación de Requerimientos no Funcionales, Especificación del Contexto y de la Medición.

Se basa en un enfoque en el cual la especificación de requerimientos, la medición y evaluación de entidades y la posterior interpretación de los resultados están orientadas a satisfacer una necesidad de información particular. Está integrado por los siguientes componentes principales: 1) Gestión de Proyectos de M&E; 2) Especificación de Requerimientos no Funcionales; 3) Especificación del Contexto del Proyecto; 4) Diseño y Ejecución de la Medición; y 5) Diseño y Ejecución de la Evaluación. La mayoría de los componentes están soportados por los términos ontológicos definidos en [5]. En la figura 2, se muestra un diagrama con los principales conceptos y relaciones para los componentes de requerimientos, contexto y medición.

Los flujos de medidas que se informan desde las fuentes de datos al EIPFDcMM, se estructuran incorporando a las medidas, metadatos basados en C-INCAMI tales como la métrica a la que corresponde, el grupo de seguimiento asociado, el atributo de la entidad que se mide, entre otros. Dentro del flujo, se etiquetan conjuntamente con cada medida asociada al atributo, las medidas asociadas a cada propiedad de contexto. Gracias a la formalización del proyecto de M&E en base a C-INCAMI, el hecho de procesar el flujo etiquetado, permite la estructuración del contenido de un modo consistente y alineado con el objetivo del proyecto. Esta estructuración de las mediciones dentro del EIPFDcMM, mantiene el concepto con el que se asocia cada medida; por ejem-

ple, si es una medida de atributo o bien de propiedad contextual. De este modo, se enriquece el análisis estadístico dado que es posible en forma directa, verificar la consistencia formal y sintáctica de cada medida contra su definición formal, en forma previa a avanzar con técnicas estadísticas más complejas. En este caso, la entidad bajo monitoreo la constituye la situación meteorológica con un radio de hasta 480 km con centro en EEA Anguil, y dentro de los atributos que desean monitorearse pueden encontrarse, por ejemplo, la cantidad de agua presente en la atmósfera, para lo que el valor del factor de reflectividad (Z , ver sección 2.2) constituye una métrica plausible de análisis. Asociado a él, un indicador de precipitaciones puede definirse, junto con criterios de decisión y cursos de acción que favorezcan la emisión temprana de alertas.

Grandes Volúmenes de Datos Persistentes y Capas de Servicios en el EIPFDcMM

Arquitectura de Procesamiento desde la Perspectiva del Flujo de Datos

El EIPFDcMM es un gestor de flujos semi-estructurados de mediciones, enriquecidos con metadatos sustentados en C-INCAMI, especializado en proyectos de M&E, que incorpora comportamiento detectivo y predictivo en línea junto con la capacidad de gestionar y proveer grandes volúmenes de datos

sobre demanda. Como puede apreciarse en la Figura 3, la idea que subyace al modelo en términos de procesamiento de flujos (Divan, 2011) es la siguiente.

Las mediciones se generan en las fuentes de datos heterogéneas (por ejemplo, el RM del INTA), las cuales abastecen a un módulo denominado adaptador de mediciones (MA en la Figura 3) generalmente embebido en los dispositivos de mediciones. MA incorpora junto a los valores medidos, los metadatos del proyecto de medición y los informa a una función de reunión central (Gathering Function -GF). GF incorpora los flujos de mediciones paralelamente en: a) El repositorio de grandes datos en forma persistente, b) En un flujo C-INCAMI/MIS para los consumidores suscriptos que deseen procesar la información al momento en que se genera (por ejemplo, si fueren datos del RM del INTA, un consumidor podría ser el Servicio Meteorológico Nacional), y c) Dentro de un buffer organizado por grupos de seguimiento –modo dinámico de agrupar a las fuentes de datos definido por el director del proyecto de M&E- con el objeto de permitir análisis estadísticos consistentes a nivel de grupo de seguimiento o bien por región geográfica donde se localicen las fuentes de datos, sin que ello implique una

carga adicional de procesamiento. Adicionalmente, GF incorpora técnicas de load shedding (Rundensteiner *et al.*, 2008) que permiten gestionar la cola de servicios asociada a las mediciones, mitigando los riesgos de desborde independientemente el modo en que se agrupen. Así, el flujo C-INCAMI/MIS es incorporado al repositorio de grandes de datos con mediciones y metadatos, y permanece disponible para atender peticiones de servicios asociados con datos de medición históricos (Big Data Repository y The Historical Data Services en la figura 3). Por otro lado, y además de que el flujo de mediciones es remitido a los consumidores suscriptos, una copia de este continúa dentro del procesador de flujo de datos y se le aplica análisis descriptivo, de correlación y componentes principales (Analysis & Smoothing Function -ASF- Fig. 3) guiados por sus propios metadatos, a los efectos de detectar situaciones inconsistentes con respecto a su definición formal, tendencias, correlaciones y/o identificar las componentes del sistema que más aportan en términos de variabilidad. De detectarse alguna situación en ASF, se dispara una alerta estadística al tomador de decisiones (Decision Maker - DM) para que evalúe si corresponde o no disparar la alerta externa (vía e-mail, SMS, y otros) que informe

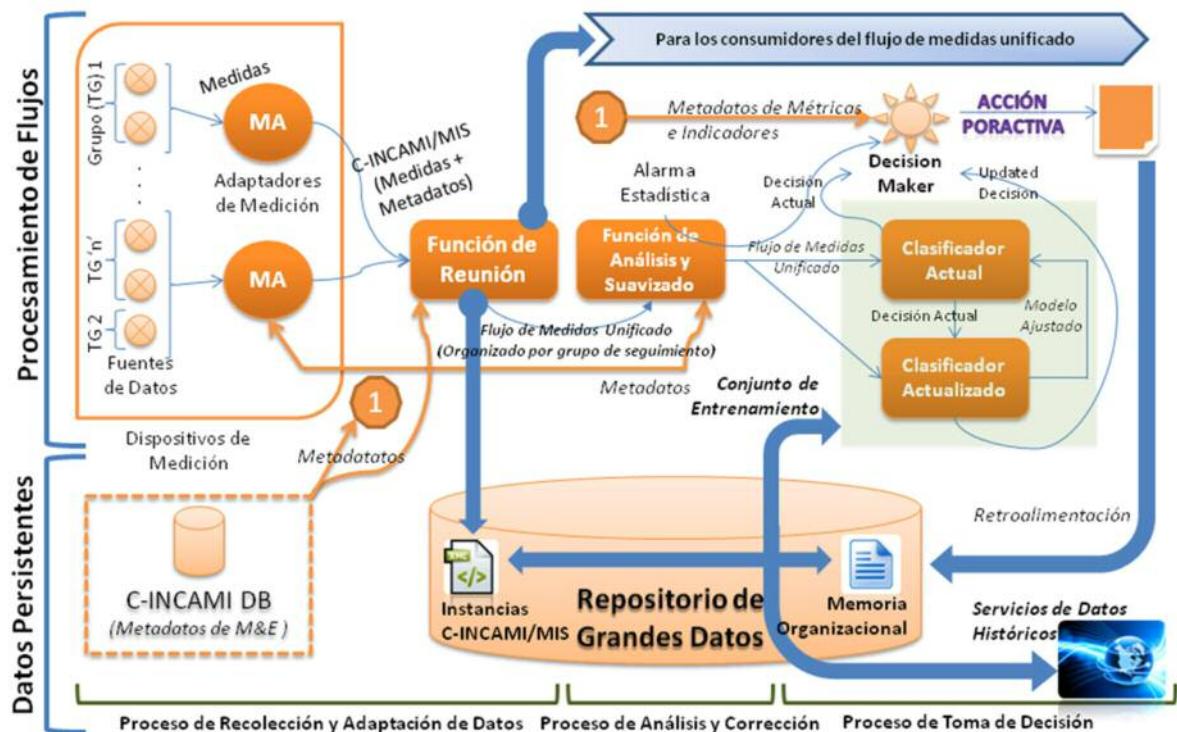


Figura 3. Esquema Conceptual del EIPFDcMM con Big Data.

al personal responsable de monitoreo sobre la situación (Por ejemplo, en el caso del RM del INTA, podría señalar un eventual error de dispositivo o descalibración del RM). En paralelo los nuevos flujos de mediciones son comunicados al clasificador vigente (Current Classifier -CC), quien deberá clasificar las nuevas mediciones si corresponden o no a una situación de riesgo e informar dicha decisión al DM. Simultáneamente, se reconstruye en línea el CC incorporando las nuevas mediciones al conjunto de entrenamiento y produciendo con ellas un nuevo modelo (Updated Classifier -UC). El UC clasificará las nuevas mediciones y producirá una decisión actualizada que también será comunicada al DM. El DM determinará si las decisiones indicadas por los clasificadores (CC y UC) corresponden a una situación de riesgo y en cuyo caso con qué probabilidad de ocurrencia, actuando en consecuencia según lo definido en el umbral mínimo de probabilidad de ocurrencia definido por el director del proyecto. Finalmente, independientemente de las decisiones adoptadas, el UC se torna en CC sustituyendo al anterior, en la medida que exista una mejora en su capacidad de clasificación según el modelo de ajuste basado en curvas ROC (Receiver Operating Characteristic) (Duin *et al.*, 2008).

Las decisiones a las que arriba el DM (sean a partir de una alerta estadística o bien de los clasificado-

res), permite retroalimentar la memoria organizacional basada en casos de la arquitectura (Organizational Memory –OM- en Fig. 3). De este modo, OM considera las variables bajo monitoreo de las distintas entidades bajo análisis, a través del flujo de medición C-INCAMI/MIS, incorporando las probabilidades de ocurrencia de los distintos eventos configurados, a partir de los criterios de decisión de los indicadores definidos para el proyecto de M&E.

Adicionalmente, el hecho de contar con un repositorio de grandes volúmenes de flujos C-INCAMI/MIS en OM, retroalimentado en forma permanente a partir del procesamiento de los distintos flujos de datos, hace posible disponer de una base de entrenamiento consistente para el start-up inicial de los clasificadores.

Arquitectura de Procesamiento desde la Perspectiva de Capas de Servicios

Desde la perspectiva de la capa de servicios, el sistema que dará soporte al procesamiento de datos del RM y la provisión de servicios a terceros desde la EEA Anguil, sigue un modelo arquitectónico distribuido en capas (Fig. 4), constituyendo una estructura modular, flexible, segura y extensible de la aplicación. Consiste en una arquitectura genérica, que está basada en Servicios Web y Streaming. Dicha arquitectura está pensada para ser usada como núcleo de aplicaciones de gestión datos del RM con distintas finalidades y fue desarrollada teniendo en cuenta tres aspectos principales:



Figura 4. Arquitectura de Procesamiento desde la Perspectiva de las Capas de Servicios.

I. Es necesario integrar distintos niveles de datos e información, a saber: 1) Los datos “crudos” (Ver sección 2.2), 2) Datos validados, clasificados y corregidos por el procesador de flujos de mediciones (GF y ASF en Fig. 3), 3) Los metadatos y reglas de validación necesarias para el procesamiento de flujos (C-INCAMI DB y MA en Fig. 3), y 4) Una Memoria Organizacional con conocimiento adquirido de casos con datos del RM previos (Fig. 3). Cualquier uno de estos cuatro niveles de datos puede ser accedido a través de servicios básicos de consultas que brinda el sistema (Fig. 4).

II. Facilitar la escalabilidad y extensibilidad. Las tres capas lógicas se distribuyen en al menos tres capas físicas basadas en nodos con capacidad de procesamiento paralelo e integración

dentro de uno o más clústers. De este modo, es posible regular, escalar y extender las prestaciones de las capas de: a) Servicios externos, la cual posee contacto directo con el RM, usuario, servicios y herramientas de terceros; b) Acceso a Datos, responsable de regular el acceso, actualización de datos y aplicación de las políticas de datos en general, constituyéndose en un clúster de servicios y c) Persistencia, responsable por el resguardo de los datos mediante un clúster de almacenamiento distribuido.

III. Incorporar Seguridad y Protección de los datos, sea para los accesos permitidos o bien para la actualización de los datos del RM efectuados mediante la capa de servicios externos.

En la figura 4 se sintetiza la arquitectura procesamiento de datos del RM desde la perspectiva de las capas de servicios (Ver sección 4.1 para una vista secuenciada de procesamiento). De este modo, para el diseño de la arquitectura, se ha elegido un estilo arquitectónico de tres capas, a saber:

- Capa de Persistencia: almacena de manera persistente tanto los datos del RM sin procesar (tal y como son obtenidos del RM), como los datos del RM ajustados y validados mediante el procesamiento en línea.
- Capa de Acceso a Datos: Brinda la funcionalidad de acceso a los datos del RM en forma uniforme y controlada para todos los servicios. Esto permite aplicar políticas de seguridad que resguarden los datos del RM y garanticen su consistencia e integridad. Se compone de los siguientes servicios a) Servicio de Acceso a Datos (Data Access Service), el cual permite consultas sobre los datos e información del RM almacenada en el Big Data, b) Servicio de Actualización de Datos (Update Service), permite actualizar los datos del RM de la capa de persistencia, brindando funcionalidad al módulo de integración de datos de la capa de servicios externos, y c) Servicios de Seguridad (Security Service), el cual además de verificar los usuarios (ideales o reales) que desean acceder a los datos del RM, se verifican sus permisos asociados y se aplican las políticas de seguridad en forma uniforme a todos los módulos de la capa actual, como así también a las que dependen de ésta.
- Capa de Servicios Externos: Por un lado y a través de los servicios básicos de consulta (Basic

Service of Query), implementa la funcionalidad de consultas ad-hoc o consumo de streaming de mediciones del RM para los usuarios, los agentes u otras herramientas que los requieran. Por otro lado, mediante el procesador de flujos de mediciones y las capacidades de transformación, se permite el procesamiento en línea (con comportamiento detectivo y predictivo) sobre el propio flujo de mediciones del RM, incorporando la posibilidad de transformar los datos (guiado por sus metadatos embebidos) mediante el esquema de integración de datos (Data Integration) para su posterior actualización.

DISCUSIÓN

Existen trabajos que enfocan el procesamiento de flujos de datos desde una óptica sintáctica, donde se permite el modelado del flujo de procesamiento y la consulta continua sobre el flujo, es realizada en términos de atributos y sus valores asociados mediante CQL (Continuous Query Language) Widom y Babu, 2001; Bockermann y Blom, 2012). Nuestra estrategia, incorpora la capacidad de introducir metadatos basados en un marco formal de M&E, que guían la organización de las medidas (datos) en el buffer, facilitando análisis consistentes y comparables desde el punto de vista estadístico, con la posibilidad de disparar alertas en forma proactiva a partir de los diferentes análisis estadísticos o bien de la decisión a la que arriben los clasificadores. Adicionalmente, cuenta con los procesos formalizados mediante SPEM, lo que promueve una especificación bien establecida, comunicable y extensible.

El Servicio Climático de Estados Unidos posee una aplicación web en los que es posible consumir datos provenientes desde diferentes RM distribuidos a lo largo del país. Nuestra propuesta, incorpora metadatos de medición al momento de la generación de datos del RM, lo que permite un estudio de la medición en base a su semántica y no solo en términos del valor cuantitativo arrojado, con las posibilidades de automatización que ello implica. Más aún, la base C-INCAMI prevé la posibilidad de gestionar distribuciones de probabilidad para una medida dada, si es que un valor determinístico no puede ser obtenido.

La Agencia Estatal de Meteorología (AEMET) del Gobierno de España, tiene diversos mecanismos

por el cual hace público los datos climáticos provenientes de los radares de su región. Si bien AEMET posee una política de datos abiertos, nuestra propuesta incorpora metadatos a los datos desde su propio origen para guiar su procesamiento sin alterarlo, y adicionalmente, permite al suscriptor y/o usuario final decidir qué tipos de datos necesita consultar (sea por streaming, o bien, por demanda a partir del repositorio Big Data), siempre dentro de un marco de seguridad asociado.

El Instituto Nacional de Investigación Espacial (Ministerio de Ciencia, Tecnología e Innovación) de Brasil dispone también de aplicaciones Web mediante las cuales hacen públicos los datos provenientes de sus RM, contando con la posibilidad incluso de seleccionar el RM de origen que se desee analizar en particular. Nuestra propuesta prevé el aprovisionamiento por suscripción de datos generados en tiempo real desde el RM, adicionalmente guiado por metadatos sustentado en una base ontológica de M&E, que permiten determinar si las mediciones arrojadas son consistentes con la definición del objeto bajo estudio.

En Argentina y adicionalmente a INTA, existen entidades que cuentan con RM en funcionamiento tales como el Servicio Meteorológico Nacional (SMN), la Dirección de Agricultura y Contingencias Climáticas (DACC, Mendoza), como así también proyectos tales como el Sistema Nacional de RM (SINA-RAME), con reciente instalación en Córdoba, de su primer RM con tecnología doppler de doble polarización de fabricación Argentina (Rocca, 2015). Dichas entidades ponen a disposición de los usuarios información en tiempo real a través de diferentes productos de sus RM en un formato de imagen (GIF, PNG, JPG) mediante diferentes sitios web como el SMN, la Dirección de Agricultura y Contingencias climáticas de Mendoza y el INTA. A diferencia de los servicios actuales mencionados, nuestra propuesta incorpora la posibilidad de acceder tanto a datos “crudos” en línea, como a datos “procesados” (o productos) vigentes y/o históricos, como así también a su detalle y a los mecanismos empleados para la generación de productos.

CONCLUSIONES

En el presente artículo hemos discutido la incorporación del RM de la EEA Anguil como fuente de

datos heterogénea del EIPFDcMM, a los efectos de incorporar comportamiento detectivo y predictivo sobre las mediciones a medida que son generadas. Adicionalmente, se planteó la actualización de la arquitectura de procesamiento, para poder no solo gestionar servicios por suscripción en tiempo real sobre los datos que genera el RM, sino también poder satisfacer a partir de repositorios Big Data la petición de datos del RM ad-hoc dentro de un marco de seguridad dado. Esto permite y promueve adicionalmente, la capacidad de extender el procesamiento a otros RM de INTA (o incluso otras Instituciones), ya que sus incorporaciones serían en forma transparente como una fuente de datos heterogénea más, posibilitándose un monito-reo detectivo y predictivo en línea de la red de RM. Por otro lado, la arquitectura de procesamiento también ha sido expuesta desde la perspectiva de la capa de servicios, lo que incorpora en forma transparente la prestación de datos a terceros interesados o beneficiarios en general. Esto permitiría que los datos fuesen consumidos sin mediar solicitud o intervención humana, por diferentes interesados. De este modo, el hecho de poder definir un proyecto de M&E sobre una base ontológica dada e incorporarla dentro de la arquitectura de procesamiento como se ha expuesto, permite que las mediciones provenientes del RM sean monitoreadas ante eventuales desvíos en su funcionamiento, promoviendo los ajustes meteorológicos pertinentes, junto con los cursos de acción pre-definidos dentro del proyecto de M&E.

Uno de los principales aspectos en cualquier proceso de medición, reside en la comparabilidad de sus medidas a lo largo del tiempo. En tal sentido, nuestra estrategia, permite incrementar la confiabilidad en el procesamiento con respecto al proyecto de M&E, haciendo consistente el cómputo sobre las medidas, y promoviendo así, la interoperabilidad con respecto a las diferentes fuentes de datos y/o los destinatarios que deseen emplear tal información, gracias a: a) encontrarse sustentada en un marco formal de M&E como C-INCAMI, b) contar con una ontología subyacente de M&E, c) guiar su procesamiento de datos en base a los metadatos asociados con la entidad bajo análisis, incluyendo también a su contexto, y finalmente, d) contar con los procesos formalizados que facilitan la comunicabilidad y extensibilidad de la estrategia.

Trabajo Futuro

Como trabajo a futuro, se avanzará sobre la definición formal del proyecto de M&E conjuntamente con los especialistas del INTA, con el objetivo de establecer los umbrales de funcionamiento normales del RM para las variables bajo estudio, como así también pre-definir los ajustes meteorológicos y cursos de acción ante eventuales desvíos, los cuales serán incorporados posteriormente a EIPFDcMM.

Reconocimientos. Esta investigación está sostenida por los proyectos PICTO 2011-0277 y CD 066/12 por la UNLPam y por el Proyecto PAMSL-1282101 de INTA Anguil, en el marco del convenio de cooperación técnica entre la Facultad de Ingeniería de la UNLPam y el INTA Anguil.

REFERENCIAS BIBLIOGRÁFICAS

- Agencia Estatal de Meteorología. Gobierno de España. (2015, Abril) El Tiempo. Radar - Composición Península y Baleares - Últimas 12 horas - Agencia Estatal de Meteorología - AEMET. Gobierno de España. [Online]. <http://www.aemet.es/es/eltiempo/observacion/radar>. Último acceso abril 4 de 2015.
- Ahmad, D.; Y. Balazinska, M. Cetintemel, U. Cherniack, M. Hwang, J. Lindner, W. Maskey, A. Rasin, A. Ryvkina, E. Tatbul, N. Xing, Y. & Zdonik, S. Abadi, 2005. "The Design of the Borealis Stream Processing Engine," in Conference on Innovative Data Systems Re-search (CIDR), Asilomar, CA, pp. 277-289.
- Aref, M.; W. Bose, R. Elmagarmid, A. Helal, A. Kamel, I. Mokbel, M. Ali, 2005. "NILE-PDT: A Phenomenon Detection and Tracking Framework for Data Stream Management Systems," in VLDB, Trondheim, Norway, pp. 1295-1298.
- Bellini Saibene, Y.; M. Volpaccio, S. Banchero, y R. Mezher, 2014. "Desarrollo y uso de herramientas libres para la explotación de datos de los radares meteorológicos del INTA," in 43º Jornadas Argentinas de Informática - 6º Congreso Argentino de Agroinformática, Buenos Aires, pp. 74-86.
- Bockermann, C. y H. Blom, 2012. "Processing Data Streams with The RapidMiner Streams Plugin," Technical University of Dortmund, Dortmund, Germany
- Chandrasekaran, S.; S. Cooper, O. Deshpande, A. Franklin, M. Hellerstein, J. Hong, W. Madden, S. Reiss, F. y Shah, M. Krishnamurthy, 2003. "TelegraphCQ: An Architectural Status Report," IEEE Data Engineering Bulletin, vol. 26.
- Dirección de Agricultura y Contingencias Climáticas. Mendoza, Argentina. (2015, Abril) Imágenes de Radar. [Online]. <http://www.contingencias.mendoza.gov.ar/web1/Radar/radar.html>. Último acceso abril 20 de 2015.
- Diván, M., 2011. "Enfoque Integrado de Procesamiento de Flujos de Datos centrado en Metadatos de Mediciones," UNLP, La Plata, PhD Thesis.
- Diván, M.; L Olsina, y S. Gordillo, 2011. "Strategy for Data Stream Processing Based on Measurement Metadata: An Outpatient Monitoring Scenario," Journal of Software Engineering and Applications, vol. 4, no. 12, pp. 653-665.
- Diván, M.; M Martín, y L. Olsina, 2013. "Towards the feedback of the Data Stream Processing based on Organizational Memory (in spanish)," Congreso Nacional de Ingeniería Informática/Sistemas de Información, Córdoba, Argentina.
- Diván, M. y L. Olsina, 2014. "Process View for a Data Stream Processing Strategy based on Measurement Metadata," Electronic Journal of SADIO, vol. 13, no. 1, pp. 16-31.
- Duin, C.; R. Tortorella, F. Marrocco, 2008. "Maximizing the area under the ROC curve by pairwise feature combination," ACM Pattern Recognition, pp. 1961-1974
- Gematronik, 2005. Rainbow® 5 Products & Algorithms. Neuss, Germany: Gematronik GmbH.
- Gehrke, J; J. Balakrishnan, H. Namit, 2008. "Towards a Streaming SQL Standard," Proceedings of the VLDB Endowment, vol. 1, no. 2, pp. 1379-1390.
- Hartmann, T.; M. Tamburrino, y S Bareilles, 2010. "Análisis preliminar de datos obtenidos por la red de radares del INTA para el estudio de precipitaciones en la Región Pampeana," in 39º Jornadas Argentinas de Informáticas - 2º Congreso Argentino de Agroinformática, Buenos Aires, p. 826.
- Instituto Nacional de Investigación Espacial (Ministerio de Ciencia, Tecnología e Innovación) de Brasil. (2015, Abril) Precipitación por Radar. [Online]. <http://sigma.cptec.inpe.br/radar/>. Último acceso abril 4 de 2015.
- Instituto Nacional de Tecnología Agropecuaria. (2015, Abril) Red de Radares Meteorológicos. [Online]. <http://radar.inta.gob.ar>. Último acceso abril 20 de 2015
- National Weather Service - USA. (2015, April) National Weather Service Doppler Radar Images. [Online]. <http://radar.weather.gov/>. Último acceso abril 6 de 2015.
- Olsina L. y H. Molina, 2007. "Towards the Support of Contextual Information to a Measurement and Evaluation Framework," in QUATIC, Lisboa, Portugal, pp. 154-163.
- Papa F, Molina H. y L. Olsina, 2007. "How to Measure and Evaluate Web Applications in a Con-sistent Way," in Ch. 13 in Web Engineering.: Springer, pp. 385-420.
- Rinehart, R., 2007. RADAR for Meteorologists, 5th ed. St. Joseph, USA, Misuri.
- Rodríguez and H. Ciappesoni, "Sistema NAcional de

Radares Meteorológicos. SINARAME," in Simposio de Radarización y Sistemas de Alerta Hidrometeorológicas del Cono Sur, Buenos Aires, Argentina, 2012.

- Rocca, G. 2015. Cazadores de Tormentas. Servicio de Información Científica de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires. [Online]. <http://nexciencia.exactas.uba.ar/radares-meteorologicos-sinarame-tormentas-pronostico-paola-salio>. Último acceso abril 20 de 2015.
 - Rundensteiner, M.; W. Mani, M. Wei, 2008. "Utility-driven Load Shedding for XML Stream Processing," in International World Wide Web, Beijing, China, pp. 855-864.
 - Servicio Meteorológico Nacional. (2015, Abril) Imágenes de Radar. [Online]. <http://www.smn.gov.ar/?mod=radar&id=4>. Último acceso abril 20 de 2015.
 - Widom, S. y J. Babu, 2001. "Continuous Queries over Data Streams," ACM SIGMOD Record, pp. 109-120
-

