

# Probabilidad de granizo derivada de la atenuación producida en un haz de radar

Santiago Banchemo<sup>1,2</sup>, Romina N. Mezher<sup>1</sup>

<sup>1</sup> Instituto de Clima y Agua - INTA

<sup>2</sup> Universidad Nacional de Luján

{banchemo.santiago, mezher.romina}@inta.gob.ar

**Resumen** En este trabajo se presentan los resultados preliminares de una herramienta para mapear probabilidades de granizo en tiempo real. Se utilizó un abordaje de aprendizaje automático con generación de *features* a partir de los parámetros de un modelo lineal ajustados con datos de  $Z_h$  obtenidos de las mediciones del radar meteorológico de INTA Paraná observando zonas afectadas por atenuación. Además, se utilizaron datos de campo provistos por empresas aseguradoras que cubren daños por granizo. Los resultados son alentadores con ajustes de los modelos que alcanzaron una exactitud de 79 %.

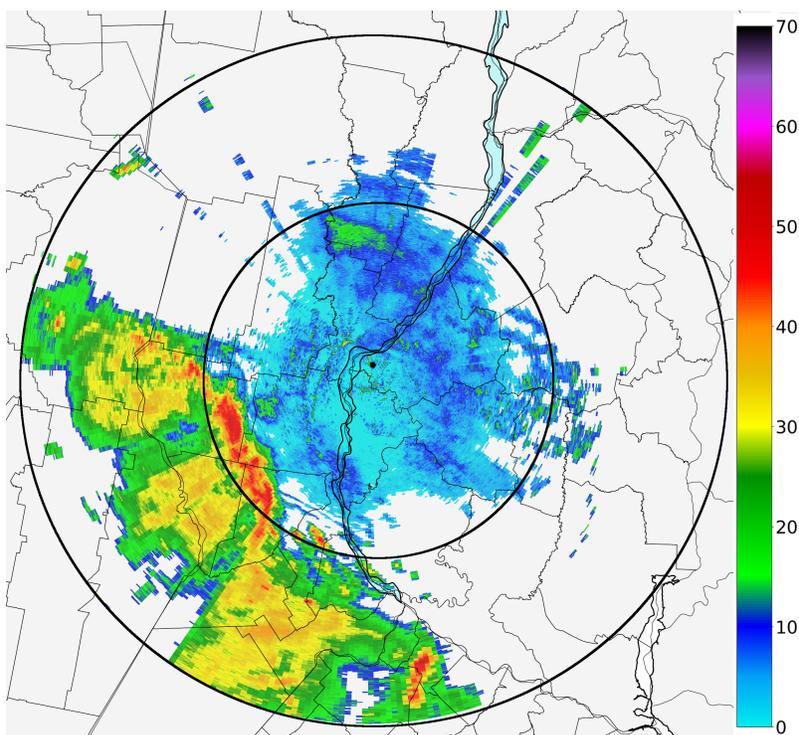
**Keywords:** Granizo, Reflectividad, Machine Learning, Random Forest

## 1. Introducción

El radar meteorológico es una de las herramientas fundamentales para la detección de granizo en celdas de tormentas. Es un sensor remoto activo transmitiendo energía en forma de ondas electromagnéticas en la frecuencia de las microondas a través de un camino de propagación, que en este caso será la atmósfera. Cuando la energía emitida encuentra un objetivo, que en este caso serán *hidrometeoros*, ésta será absorbida, dispersada en todas las direcciones o retrodispersada hacia el radar. La potencia retrodispersada al radar será una medida llamada usualmente *Reflectividad* y es la variable que comúnmente se observa en las imágenes que muestran la intensidad de los ecos precipitantes [10]. La reflectividad, en adelante  $Z_h$ , se define como la suma de los diámetros de gotas a la sexta dentro de un volumen de medición y se mide en unidades de  $mm^6 m^{-3}$ , pero debido a su orden de magnitud usualmente se utiliza la escala logarítmica y  $Z_h$  se medirá en unidades de dBZ. Se define a la *Atenuación* como la pérdida de potencia debido a la absorción o dispersión debido a la presencia de gases o hidrometeoros y se mide en unidades de  $dB km^{-1}$ . La atenuación debida a los gases presentes en la atmósfera o debido a gotas de lluvia o nube se dará principalmente por absorción de los mismos; en cambio en presencia de granizo la atenuación es producto tanto de absorción como de dispersión de energía [6]. Cuando se produce atenuación en el camino de propagación se observa una disminución de  $Z_h$  por detrás del eco de precipitación principal, la cual aumenta en presencia de lluvia (o sea agua líquida) a medida que disminuye la longitud

de onda de emisión del radar, por lo tanto un radar de banda C (longitud de onda de 5.4cm) presentará mayor atenuación que uno de banda S (longitud de onda de 10cm).

Diversos estudios [7] [4], entre otros, estudiaron la atenuación causada por radares de simple polarización en diferentes bandas mostrando que la misma tiene una relación de tipo ley de potencia con la precipitación. En radares de doble polarización [11] muestran la disminución de  $Z_h$  luego de ser afectada por atenuación en presencia de lluvia intensa en radares de banda C en un dado acimut de medición. En [1] muestran que la atenuación puede alcanzar valores de  $0.8 \text{ dB km}^{-1}$  en lluvia, obtenido por medio de datos de disdrómetros y entre  $1 \text{ y } 4 \text{ dB km}^{-1}$  en presencia de granizo en radares de banda C y [12] también muestran que la atenuación y disminución de  $Z_h$  es muy importante incluso en radares de banda S cuando las tormentas tienen granizo incluso pequeño.

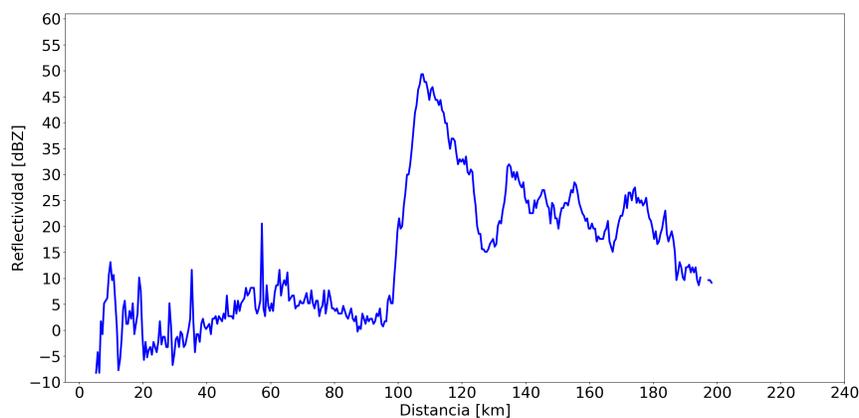


**Figura 1.** PPI de  $Z_h$  a  $0.5^\circ$  del día 01 de Enero de 2015 a las 09:40UTC

En Argentina los radares meteorológicos operativos son radares de banda C por lo que el efecto de atenuación es significativo. Se muestra un ejemplo en una imagen PPI (*Plan Position Indicator*) de  $Z_h$  en la elevación de antena de  $0,5^\circ$  del

día 01 de enero de 2015 a las 09:40UTC del radar de Paraná (Figura 1) donde los valores de  $Z_h$  superiores a los 40dBZ corresponden a áreas con precipitaciones más intensas.

Se observa que hacia el sudoeste se están registrando unas tormentas organizadas en forma de línea con precipitaciones significativas en tonos rojos. Por detrás de las mismas se muestran dos claras regiones, en tonos verdes o azules, donde  $Z_h$  tiene valores menores a los 25 dBZ y son las zonas que se encuentran mayormente afectadas por la atenuación. Se realizó un perfil radial en el acimut 220 grados que corresponde al sudoeste del radar y se muestra en la figura 2. Se observa que desde el centro hasta aproximadamente los 100km de distancia hay  $Z_h$  muy débiles correspondientes a ecos no meteorológicos; luego comienza a aumentar hasta alcanzar un máximo de 50 dBZ a los 105km de rango y posteriormente en los siguientes 10km se produce una abrupta disminución de  $Z_h$ .



**Figura 2.** Perfil radial de  $Z_h$  en el acimut 220 grados para el 01 de Enero de 2015 a las 09:40UTC

El objetivo de este trabajo es analizar la disminución de  $Z_h$  que se produce en el sector posterior de ecos intensos asociados a la presencia de fuertes precipitaciones (efecto de atenuación) en cada acimut donde se registro caída de granizo en superficie para luego poder tomarlo como proxy de detección de áreas posiblemente afectadas por granizo. Si bien con metodologías que utilizan datos de radares de doble polarización la detección de granizo es más eficaz [5,2], se quiere utilizar solo los datos de  $Z_h$  para poder analizar los eventos que se registran dentro de los radares de simple polarización o cuando existan errores en los datos polarimétricos.

## 2. Materiales y Métodos

### 2.1. Área de Estudio

El área de estudio fue la zona cubierta por el radar meteorológico del INTA ubicado en la Estación Experimental Agropecuaria de Paraná (Figura 3). En este trabajo se utilizaron los barridos de 240 km, lo que da una cobertura aproximada de 180,000  $km^2$  que cubre gran parte de la zona de producción agropecuaria núcleo de la República Argentina.

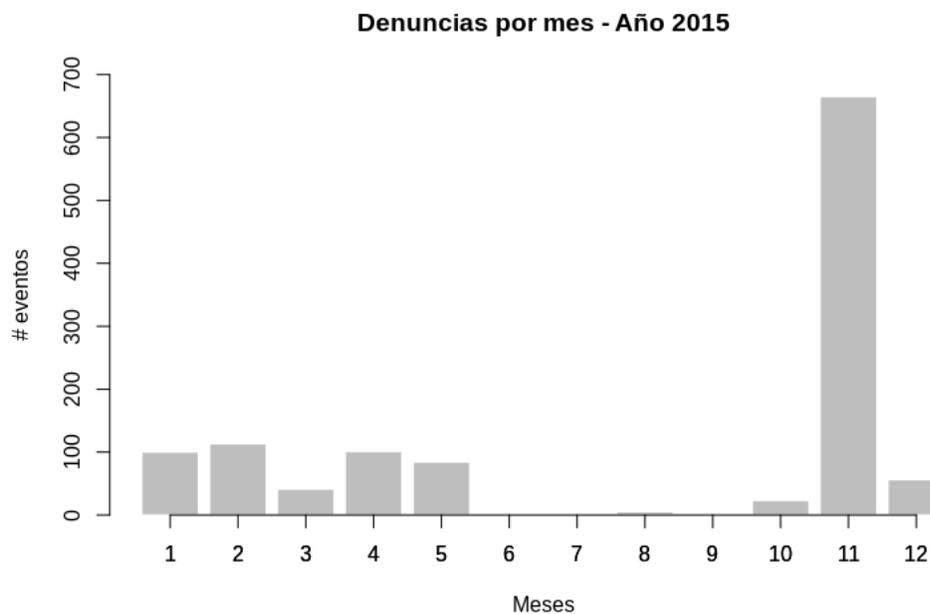


**Figura 3.** Área de estudio comprendida por el radio de 240 km que cubre el radar del INTA ubicado en Paraná.

### 2.2. Conjunto de datos

Los datos utilizados en el trabajo corresponden a siniestros por caída de granizo verificados por una empresa aseguradora de los cuales utilizó la fecha del evento y las coordenadas geográficas en latitud y longitud. En total se cuenta con 1191 casos en el área de estudio que ocurrieron en 73 días diferentes para

el año 2015. La distribución de casos por mes se muestra en la Figura 4 donde puede observarse una cantidad importante de eventos en el mes de noviembre que registró poco menos de 700 casos. En los meses de enero a mayo y octubre a diciembre se registraron la mayor cantidad de eventos, esto tiene que ver en parte con el fenómeno de granizo y además con las campañas agrícolas donde se opera en mayor medida con las organizaciones de seguros. Los meses centrales del años que corresponden a cambios de años agrícolas la demanda de seguros es menor por lo tanto no hay registros. En [8] se realizó una climatología de granizo en Argentina mostrando que la mayor ocurrencia del fenómeno en el norte de la región Pampeana es en primavera con un máximo en octubre.



**Figura 4.** Casos registrados por mes la empresa aseguradora en el área de estudio para el año 2015.

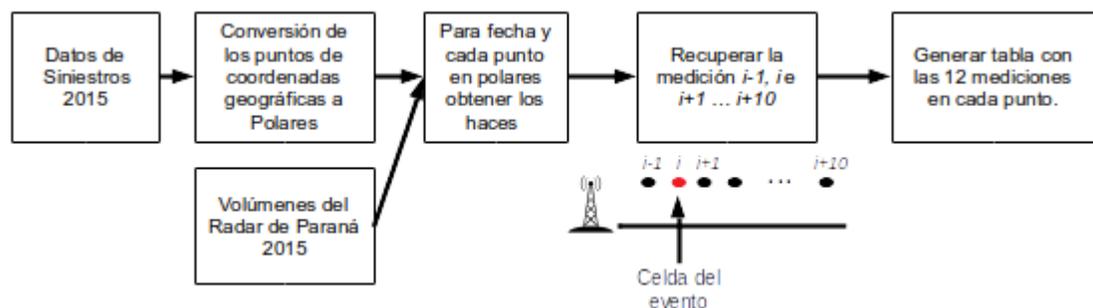
Por otro lado, se utilizaron las mediciones de la variable  $Z_h$  realizadas por el radar de meteorológico de INTA en la EEA Paraná durante el año 2015. Puntualmente se utilizó el PPI de  $Z_h$  en dBZ que corresponde a la primer elevación con un ángulo de  $0.5^\circ$ . Este radar tiene una resolución temporal de diez minutos y realiza escaneos en la atmósfera con diez ángulos de elevación diferentes desde  $0.5^\circ$  hasta  $15.1^\circ$  en el rango de 240km de medición.

La estructura de datos interna de un archivo de radar (un volumen) almacena los datos en matrices de  $360 \times 480$ , ya que mide con una resolución entre acimut de  $1^\circ$  y una resolución espacial dentro del haz de 500 metros. En total se tienen 144 archivos diarios de mediciones de  $Z_h$ .

### 2.3. Metodología

La metodología de trabajo está compuesta por una etapa de preprocesamiento de los datos, para permitir la generación de un espacio de características apropiado para las tareas de descubrimiento de conocimiento por medio de exploración de datos y aprendizaje automático. En este sentido se definió una *pipeline* de procesamiento para cruzar los datos de campo (siniestros) con los datos de radar. Ese esquema se presenta en la Figura 5, allí se muestra el flujo de tareas que incluyen una transformación de la ubicación de un siniestro (latitud y longitud) a coordenadas polares (ángulos y distancias). Esta conversión se realiza en dos pasos, un primero transformando de coordenadas geográficas a coordenadas planas (*Azimuthal Equal Distance* - AEQD) centrada en la posición del radar y posteriormente se calcula el punto en polares.

El siguiente paso es recuperar los valores de  $Z_h$  en dBZ para cada punto dentro del haz y para cada fecha, y luego se recupera el dato que se encuentra a la distancia de la coordenada polar ( $i$ ) y luego se recuperan los valores  $i - 1$  e  $i + 1 \dots i + 10$ . Así se construye una matriz de datos con 12 mediciones de  $Z_h$  y un *timestamp* de la medición.



**Figura 5.** Esquema de preprocesamiento para generación del dataset.

A partir de la tabla de mediciones se genera la variable a modelar, considerando los valores observados en el punto  $i$  se calcularon los percentiles del valor de  $Z_h$  y se tomaron 5 intervalos. Esto fue necesario para manejar el desbalanceo de casos con reflectividad en valores críticos ( $Z_h \geq 40dBZ$ ) y compensar el exceso de casos de observaciones negativas por sobre las positivas. Los rangos de cada clase se muestran en la Tabla 1.

**Tabla 1.** Rangos de  $Z_h$  para las 5 clases definidas a partir de los percentiles

Límite inferior (dBZ)	Límite superior (dBZ)	Clase	Granizo	Observaciones
0.25	4.22	1	No	13579
4.22	8.19	2	No	15535
8.19	14.1	3	No	13322
14.1	39.9	4	No	13572
39.9	60.3	5	Si	580

Cabe destacar que todos los puntos son casos donde se registró caída de granizo y aún así existen registros que no intersectan con reflectividad igual o superior al umbral propuesto. Esto sucede porque al tener una resolución de 10 minutos, dependiendo de la velocidad de movimiento de la tormenta, el siniestro puede quedar entre 2 tiempos, o sea la tormenta que produjo el granizo se observa antes de llegar al punto de siniestro en un tiempo  $k$  y en tiempo posterior  $k+1$  la tormenta se observa por delante del siniestro. En total se obtuvieron 580 casos de puntos coincidentes con mediciones de  $Z_h \geq 40dBZ$  que se atribuyen con presencia granizo y el resto de los casos distribuidos en las 4 clases de *no granizo* (Tabla 1).

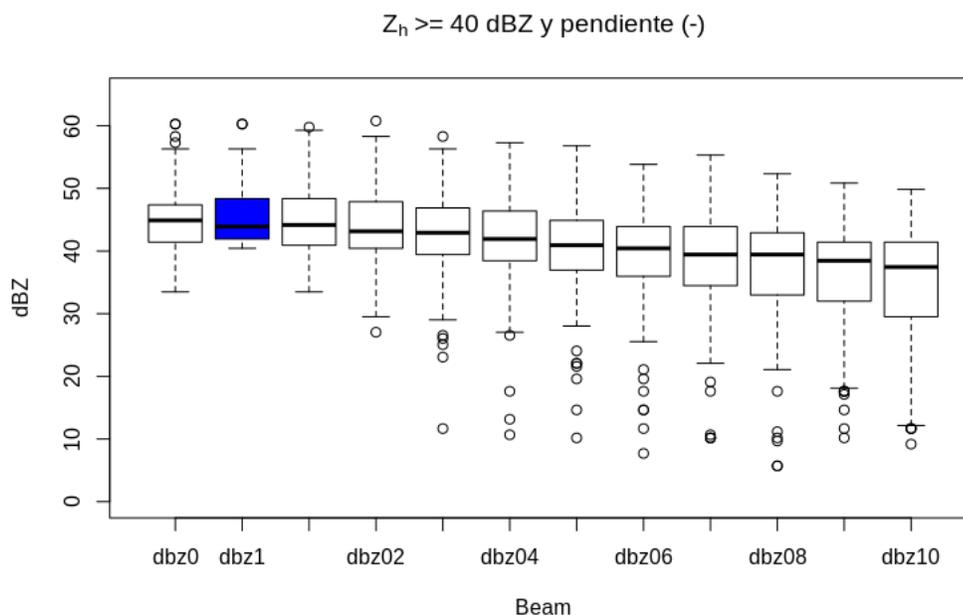
El espacio de características fue generado con un modelo de regresión lineal simple sobre la serie de 12 mediciones registradas en el haz - para el punto medido en  $i$ . De esos ajustes, para cada registro se obtienen los parámetros del modelo lineal  $\beta_0$  y  $\beta_1$ , que fueron utilizados como variables independientes en una clasificación con el algoritmo Random Forest [3]. Este algoritmo de ensamble [9], genera árboles de decisión al azar a partir de un proceso de muestreo sobre las variables y sobre los datos. En este trabajo se utilizó una parametrización con 1000 árboles sin restricciones de altura ni cantidad de individuos por árbol.

Se realiza una división del conjunto de datos en proporciones 80/20 para entrenar y validar respectivamente. Sobre el conjunto de entrenamiento se realiza una validación cruzada con 10-Fold y 3 repeticiones. Como métrica de evaluación se utilizó *exactitud* con los datos de validación independientes.

Para la construcción de los mapas de probabilidades se utilizó un filtrado de ruido utilizando la variable Coeficiente de Correlación que solo se obtiene de los radares polarimétricos la cual es capaz de separar de modo simple ecos meteorológicos de los no meteorológicos. Esta variable tiene datos entre 0 y 1 y se tomó un umbral inferior de 0.6 para el filtrado de datos.

### 3. Resultados

A continuación se presentan resultados preliminares que se obtuvieron del análisis exploratorio. Allí se realizó una primer comparación entre los casos que se etiquetaron como granizo positivo con  $Z_h \geq 40dBZ$  y las pendientes positivas y negativas. En el primer gráfico de cajas (Figura 6) puede observarse que luego del  $Z_h$  en el punto donde se registra el siniestro (caja color azul), los valores de reflectividad comienzan a bajar mostrando la pendiente negativa que se mencionaba previamente relacionándola con la pérdida de potencia en  $Z_h$  producto de atenuación presente por el granizo. En este análisis se observa que el 80% de los etiquetados como granizo verdadero tienen pendiente negativa de  $Z_h$  en los siguientes 5 km del haz.



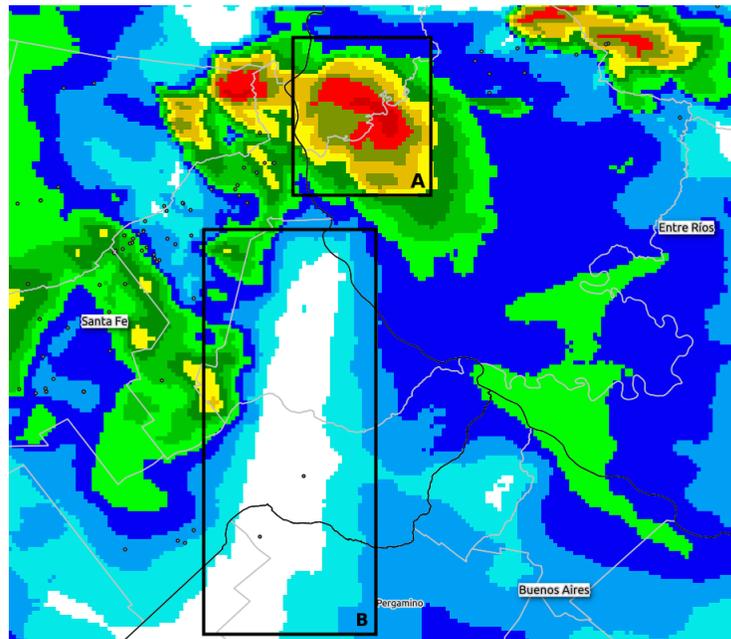
**Figura 6.** Gráfico de cajas para las mediciones (caja color azul) con  $Z_h \geq 40dBZ$  y pendiente de las mediciones posteriores negativa.

Los modelos ajustados con *Random Forest* mostraron resultados promisorios para estas primeras pruebas. El ajuste para las 5 clases dio como resultado un ajuste de exactitud de 0.79 y kappa de 0.72 para el conjunto de *testing*. En la matriz de confusión (Tabla 2) se observa que existe mucho ruido en las clases de 1 a 4 correspondientes a *granizo=No* y en menor

**Tabla 2.** Matriz de confusión de Random Forest para las cinco clases de las cuales solo la número 5 es *granizo=Si*.

Observados	Predichos				
	1	2	3	4	5
1	2163	441	63	18	0
2	422	2229	382	45	0
3	75	349	2050	186	0
4	21	35	223	2422	12
5	2	1	0	30	83

En la Figura 8, se muestra un ejemplo un evento intenso ocurrido el 18 de noviembre de 2015 donde se registraron más de 250 puntos de daño verificados presentes en el área de estudio. Las probabilidades de ocurrencia de granizo derivada de disminución de reflectividad por atenuación producida en un haz de radar indican que existe una coherencia entre los eventos registrados y los sitios con mayor probabilidad de ocurrencia de granizo que se obtuvo con *Random Forest*.



**Figura 7.** Reflectividad producida por una celda de tormenta (A) a las 19:50 UTC donde se destaca el área atenuada en (B).

Además, se destacan dos momentos con reflectividades elevadas ( $Z_h \geq 55dBZ$ ) que fueron las 18:10 y 18:20 horas UTC que muestra una buena correlación espacial entre el mapa de probabilidades y los puntos observados. Notese también que en las horas 19:40 y 20:10 UTC se observa reflectividad intensa en la celda de tormenta (Figura 7 A) ubicada entre Santa Fe y Entre Ríos mostrando de este modo la intensa atenuación producida (Figura 7 B) ya que no se observa reflectividad al sur.

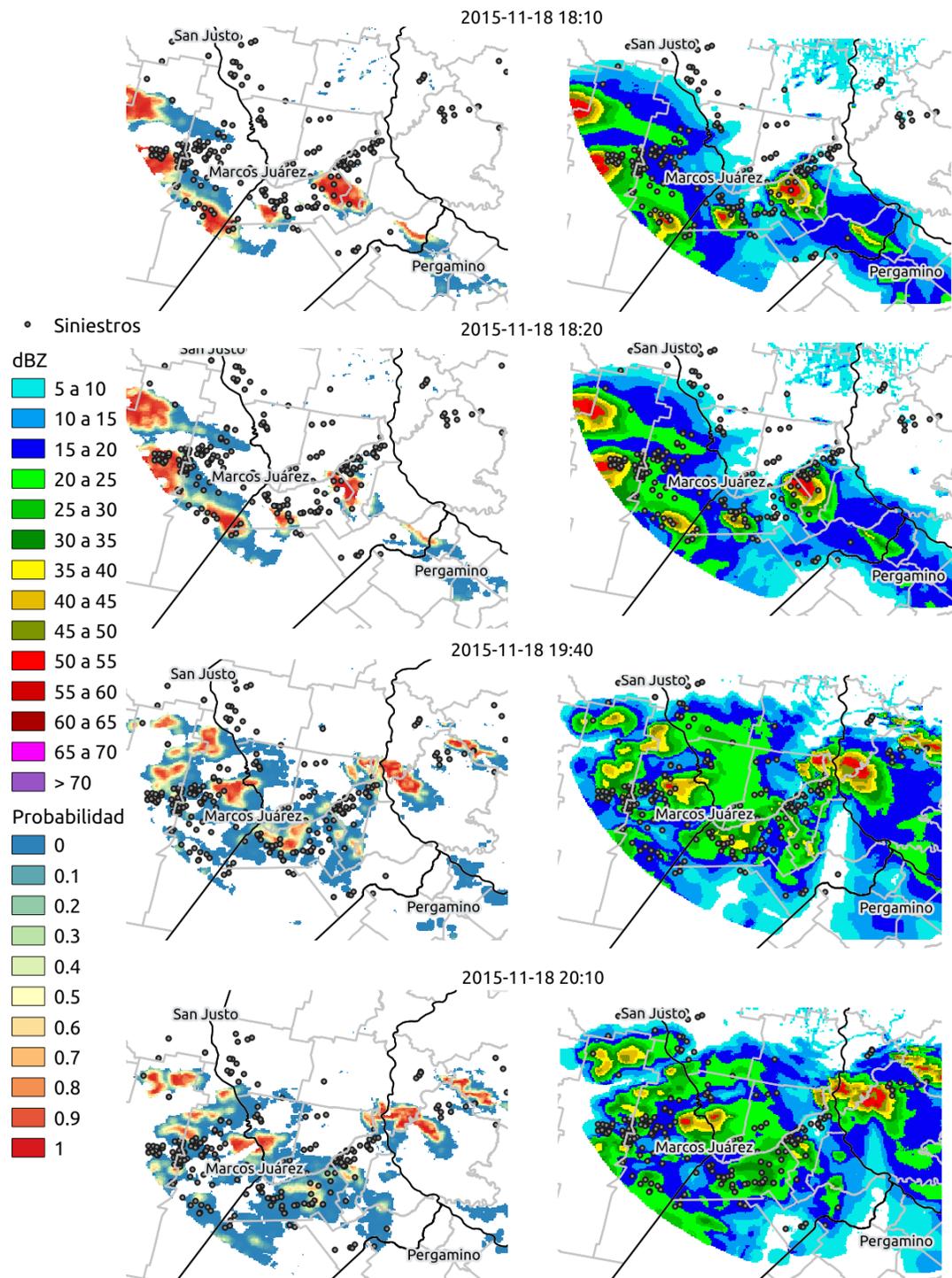
#### 4. Comentarios Finales

Se probaron técnicas de aprendizaje automático para la elaboración de mapas de probabilidades de granizo donde los resultados son promisorios. Los parámetros del modelo lineal ajustados para  $Z_h$  asociados a la atenuación mostraron ser buenos predictores de eventos de granizo, donde para los más intensos es posible generar un mapa con exactitudes de 0.79 que resulta muy alentador para el mapeo en tiempo real de probabilidad de granizo con buena resolución espacial y temporal.

#### 5. Agradecimientos

Este trabajo fue realizado gracias al apoyo de:

- Convenio de Colaboración Técnica N° 25104 INTA – Paraná Sociedad Anónima de Seguros.
- Convenio de Colaboración Técnica N° 22934 INTA – SANCOR Cooperativa de seguros Limitada.



**Figura 8.** Eventos registrados durante el día 18 de noviembre de 2015. Los mapas de la derecha corresponden a la variable  $Z$ , medida en dBZ, mientras que los mapas de la izquierda se corresponden con los mapas de probabilidad de granizo que se obtuvieron a partir de los parámetros del modelo lineal.

## Referencias

1. BOROWSKA, L., RYZHKOV, A., ZRNIC, D., SIMMER, C., AND PALMER, R. Attenuation and differential attenuation of 5-cm-wavelength radiation in melting hail. *Journal of Applied Meteorology and Climatology* 50, 1 (2011), 59–76.
2. BRANDES, E. A., AND RYZHKOV, A. Hail detection with polarimetric radar. In *Preprints, 11th Conf. on Aviation, Range, and Aerospace Meteorology, Hyannis, MA, Amer. Meteor. Soc., CD-ROM P* (2004), vol. 5.
3. BREIMAN, L. Random forests. *Machine learning* 45, 1 (2001), 5–32.
4. DELRIEU, G., ANDRIEU, H., AND CREUTIN, J. D. Quantification of path-integrated attenuation for x-and c-band weather radar systems operating in mediterranean heavy rainfall. *Journal of Applied Meteorology* 39, 6 (2000), 840–850.
5. DEPUE, T. K., KENNEDY, P. C., AND RUTLEDGE, S. A. Performance of the hail differential reflectivity (h dr) polarimetric radar hail indicator. *Journal of Applied Meteorology and Climatology* 46, 8 (2007), 1290–1301.
6. DOVIK, R. J., ET AL. *Doppler radar and weather observations*. Courier Corporation, 2006.
7. HITSCHFELD, W., AND BORDAN, J. Errors inherent in the radar measurement of rainfall at attenuating wavelengths. *Journal of Meteorology* 11, 1 (1954), 58–67.
8. MEZHER, R. N., DOYLE, M., AND BARROS, V. Climatology of hail in argentina. *Atmospheric research* 114 (2012), 70–82.
9. POLIKAR, R. Ensemble learning. In *Ensemble machine learning*. Springer, 2012, pp. 1–34.
10. RAUBER, R. M., AND NESBITT, S. W. *Radar Meteorology: A First Course*. John Wiley & Sons, 2018.
11. RYZHKOV, A., AND ZRNIC, D. Radar polarimetry at s, c, and x bands: Comparative analysis and operational implications, paper presented at 32nd conference on radar meteorology. *Am. Meteorol. Soc., Norman, Okla* (2005).
12. RYZHKOV, A. V., GANSON, S., KHAIN, A., PINSKY, M., AND POKROVSKY, A. Polarimetric characteristics of melting hail at s and c bands. In *Preprints, 34th Conf. on Radar Meteorology, Williamsburg, VA, Amer. Meteor. Soc. A* (2009), vol. 4.