

# Predicción de granizo utilizando índices atmosféricos

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

<sup>1</sup> Instituto de Clima y Agua (INTA). De los Reseros y N. Repeto s/n, Hurlingham.

<sup>2</sup> Facultad de Agronomía. UBA

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

soria@agro.uba.ar

**Resumen** En la actividad agrícola existen muchos factores de riesgo que condicionan los resultados donde algunos pueden ser tratados por el hombre, como por ejemplo enfermedades o plagas pero también existen otros que no pueden ser controlados y que de suceder ponen en peligro la seguridad de las personas y producen grandes trastornos económicos. Este conjunto de factores de riesgo tienen que ver con el comportamiento de la atmósfera y los eventos extremos que surgen de ella. En particular, la ocurrencia de granizo constituye uno de los fenómenos meteorológicos capaces de infligir daños cuantiosos al deteriorar seriamente cultivos, edificios, medios de transporte, entre otros. En este trabajo se somete a discusión parte del trabajo de una tesis de maestría en minería de datos y descubrimiento de conocimiento donde el objetivo es evaluar cuál es la capacidad de predicción de ocurrencia de granizo de un conjunto de índices de inestabilidad utilizando técnicas de aprendizaje automático en un entorno de Big Data.

**Keywords:** Global Forecast System, Radares Meteorológicos, Big Data

## 1. Introducción

En la actividad agrícola existen muchos factores de riesgo que condicionan los resultados, la ocurrencia de granizo constituye uno de los fenómenos meteorológicos capaces de infligir daños cuantiosos al deteriorar seriamente cultivos, edificios, medios de transporte, etc. En Argentina la precipitación de granizo o hidrometeoros tiene una distribución espacio-temporal heterogénea, según [9],[8],[10] la mayor frecuencia en el año se presenta durante el verano y la primavera mientras que durante el otoño y el invierno la frecuencia de ocurrencia es menor y en algunas regiones como en Cuyo o el extremo Noroeste que prácticamente la frecuencia es nula.

La ocurrencia de granizo es un fenómeno que a diferencia de otras variables meteorológicas, como la temperatura que es más homogénea, es difícil de pronosticar a largo plazo dado su comportamiento incierto y su escala de medición. Por tratarse de un fenómeno convectivo, está definido en una mesoescala donde la escala temporal del pronóstico se mide en horas y la resolución espacial se fija entre 2 y 20 kilómetros [12].

Los índices de inestabilidad, como: CAPE, Lift, K-Index, Total Totals, etc. [5] se obtienen a partir de variables atmosféricas como presión, altura, temperatura, temperatura de rocío, entre otras. Su principal utilidad es brindar una herramienta de exploración de condiciones de convección. Tradicionalmente estos índices se calculan utilizando radiosondas que son lanzadas desde aeropuertos internacionales, el problema es que la densidad de puntos de relevamiento es muy baja y el área de cobertura de una sonda es de 200km [1] por ese motivo se utilizaran los datos de GFS.

En este trabajo se utilizaran para el cálculo de los índices datos pronosticados por el *Global Forecast System* (GFS) [3], que es un sistema de predicción numérica del tiempo y que es corrido diariamente por el US National Weather Service (NWS) para generar pronósticos a corto plazo. Estos datos están disponibles cuatro veces al día con cobertura global en formato de grilla con una resolución espacial de 0.25 grados [11].

Como variable objetivo de nuestros modelos se utilizaran las observaciones de la red de radares del INTA. Este instrumental conforma una red híbrida con sensores de polarización simple (Radar de Pergamino) y doble (Radares de Paraná y Anguil) que permiten la observación detallada de la atmosfera con un radio de 240 kilómetros. Puntualmente se utilizará la variable dBZ (o reflectividad) ya que se conocen los umbrales [13] de dicha variable para determinar la presencia de granizo en las nubes.

Las tareas a realizar serán construir modelos a partir de técnicas de aprendizaje automático, utilizando como datos de entrada los índices de inestabilidad y realizar validaciones con datos observados por los radares meteorológicos. A su vez se realizará una validación con datos relevados en campo donde se verificó la caída de granizo.

*Big Data* es el término que se utiliza para conjuntos de datos masivos con estructuras variadas y complejas, donde el almacenamiento, el análisis y la visualización es un problema difícil de manejar[14]. También se lo asocia con la capacidad de extraer información útil de grandes conjuntos de datos o de *stream* de datos, estas nuevas técnicas de *mining* son necesarias dado el volumen, la variabilidad y la velocidad de los datos que se manejan [4].

Se han encontrado trabajos relacionados con la problemática de pronosticar ocurrencia de granizo. En el caso de [7] utiliza índices de estabilidad obtenidos de radiosondas además de otros datos que son combinados con técnicas de aprendizaje automático como regresión logística para pronosticar. Otro trabajo como [6] utiliza las salidas GFS para pronosticar granizo pero no utiliza técnicas de aprendizaje automático, sino que realiza un análisis estadístico sobre un conjunto de variables que son procesadas para calcular índices de inestabilidad. Por último, [2] utiliza datos de reanálisis de NCEP para simular a través de datos modelados y corregidos cómo se distribuye el granizo a nivel global.

## 2. Motivación e importancia del campo

En la actualidad, el sector agropecuario cuenta con un afluente de datos provenientes de redes de sensores de variada resolución espacial y temporal. El relevamiento y disponibilización de información climática es una de las demandas más fuertes del sector que exige datos de calidad y en tiempo real o casi real.

También existe una fuerte demanda de productos derivados de estas redes de sensores para utilizar en los procesos de toma de decisiones. En este caso, la motivación está ligada a la posibilidad de brindar un producto de ayuda a los pronosticadores para utilizar como insumo de su actividad. Esto va a permitir obtener pronósticos más precisos y mejorar la información disponible para la región.

Por otro lado, el advenimiento de las nuevas tecnologías de cloud computing ligadas al procesamiento masivo de datos permite que se pueda incrementar la capacidad de cómputo escalando horizontalmente cada vez con mayor facilidad. La incorporación de tecnologías como Apache Hadoop a los procesos de cómputo y almacenamiento masivo permiten alcanzar objetivos de cómputo de alto desempeño con mayor facilidad.

## 3. Comentarios Finales

En este trabajo se evaluará cuál es la capacidad de predicción de ocurrencia de granizo de un conjunto de índices de inestabilidad utilizando algoritmos de *machine learning* en un entorno de Big Data. Consideramos que debe ser tratado como un problema de Big Data desde el punto de vista de la variedad, ya que se integran fuentes provenientes de diferentes plataformas de redes de sensores y la velocidad en los tiempos de respuestas del sistema para generar los pronósticos. En relación al volumen, en principio se estará trabajando con unos 4GB de datos crudos diarios provenientes de las grillas de pronósticos GFS y de los volúmenes de radar aún no está definida la ventana de tiempo para saber con cuantos datos hacia atrás se deberá contar.

Para finalizar, la existencia de datos cuatro veces al día - con pronósticos hasta 15 días con intervalos de tres horas - va a permitir que se puedan generar prototipos para aplicaciones operativas de predicción de granizo, ya que a partir de los datos grillados la construcción de mapas de riesgo de ocurrencia de un evento es una tarea simple. Este es un aporte muy importante para los pronosticadores y para los diferentes sectores agro-productivos del país.

## Referencias

1. Guide to Meteorological Instruments and Methods of Observation WMO. WMO-No. 8 (2008 edition, Updated in 2010), 2008 edn., <http://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html>

2. Brooks, H.E., Lee, J.W., Craven, J.P.: The spatial distribution of severe thunderstorm and tornado environments from global reanalysis data. *Atmospheric Research* 67, 73–94 (2003), <http://www.sciencedirect.com/science/article/pii/S0169809503000450>
3. Center, E.M.: The global forecast system (gfs) - global spectral model (gsm) (gfs version 11.0.6). Tech. rep., Environmental Modeling Center [Disponible en <http://www.emc.ncep.noaa.gov/GFS/doc.php>] (2003)
4. Fan, W., Bifet, A.: Mining big data: current status, and forecast to the future. *ACM SIGKDD Explorations Newsletter* 14(2), 1–5 (2013)
5. Gottlieb, R.: Analysis of stability indices for severe thunderstorms in the northeastern united states. Ph.D. thesis (2009)
6. Kagermazov, A.K.: The forecast of hail based on the atmospheric global model (t254 ncep) output data. *Russian Meteorology and Hydrology* 37(3), 165–169 (2012)
7. López, L., García-Ortega, E., Sánchez, J.L.: A short-term forecast model for hail. *Atmospheric research* 83(2), 176–184 (2007), <http://www.sciencedirect.com/science/article/pii/S0169809506001086>
8. Mezher, R.N., Mercuri, P.A.: Análisis espacial y temporal de la ocurrencia de eventos de granizo sobre argentina. XV Congresso Brasileiro de Meteorologia (Aug 2008)
9. Mezher, R.N., Mercuri, P.A., Gattinoni, N.N.: Distribucion espacio-temporal del granizo en argentina. Reunión Argentina de Agrometeorología. 12. 2008 10 08-10, 8 al 10 de octubre de 2008. San Salvador de Jujuy. AR. (2008)
10. Mezher, R.N., Doyle, M., Barros, V.: Climatology of hail in argentina. *Atmospheric research* 114, 70–82 (2012)
11. NOAA - National Weather Service: GFS half degree documentation. [http://nomads.ncep.noaa.gov/txt\\_descriptions/GFS\\_half\\_degree\\_doc.shtml](http://nomads.ncep.noaa.gov/txt_descriptions/GFS_half_degree_doc.shtml), accedido el 4 de Noviembre de 2014
12. Orlanski, I.: A rational subdivision of scales for atmospheric process. *Bull. Am. Meteorol. Soc.* 56, 527–530 (1975)
13. Rinehart, R.E.: Radar for meteorologists. University of North Dakota, Office of the President (1991)
14. Sagioglu, S., Sinanc, D.: Big data: A review. In: *Collaboration Technologies and Systems (CTS), 2013 International Conference on*. pp. 42–47. IEEE (2013)