



II CONGRESO CHILENO DE INGENIERÍA AMBIENTAL EN SISTEMAS ACUÁTICOS, CChIASA

MODELO DE PRONÓSTICO DE CRECIDAS EN LA CUENCA DEL RÍO MAPOCHO

Andrés Vivallo Speer¹
Rodrigo Fuster Gómez²

RESUMEN EXTENDIDO

Desde la época de la colonia que las crecidas son un tema de preocupación en la ciudad de Santiago, debido a los desbordamientos que el Río Mapocho ha presentado históricamente. En el año 1987, la zona centro del país se vio afectado por intensas lluvias, las que provocaron grandes inundaciones en la cuenca del Río Mapocho (ONEMI, 1995). En la actualidad sigue siendo un tema de preocupación, en el año 2016 por intervenciones del proyecto Santiago Centro Oriente y 77 mm de agua que cayeron sobre la Capital en 36 h se produjo la última crecida del Mapocho, lo que provocó el corte de la Av. Andrés Bello y parte de la Costanera Norte, además de afectar al sector comercial de Providencia (El Mercurio, 2016). Es por estos motivos que se hace necesario investigar acerca de las crecidas del Río Mapocho y actualizar el modelo del año 1983 de la Dirección General de Aguas (DGA) existente en la estación Río Mapocho en los Almendros (Peña & Nazarala, 1983).

Este trabajo se centró en la modelación del volumen total de agua drenada en eventos de crecidas pluviales con datos meteorológicos diarios, a través de dos métodos de modelación estadística: Generalized Linear Model (GLM) y Random Forest (RF). El objetivo principal de la investigación fue seleccionar entre estos dos métodos, aquel que se ajuste de mejor manera para la predicción de crecidas en la estación Río Mapocho en los Almendros.

El método RF es estocástico no paramétrico, en este, los valores de entrada se van clasificando según su condición, lo que va creando distintas ramas que pertenecen a una misma categoría raíz y generan árboles de decisiones, que en conjunto forman un sólo árbol para dar con la variable a modelar, en la **Figura 1** se observa una representación gráfica (Breiman, 2001; Vizcaino, 2008). Por otro lado, el GLM es un método de regresión paramétrico y determinístico, este se diferencia de las regresiones lineales clásicas por ser más flexible (Mangeaud & Videla, 2005).

¹ Laboratorio de Análisis Territorial, U. de Chile /Andrés Vivallo S., vivallo.cl@gmail.com

² Laboratorio de Análisis Territorial, Depto. Cs. Ambientales y RNR, U. de Chile./Rodrigo Fuster G., rfuster@uchile.cl

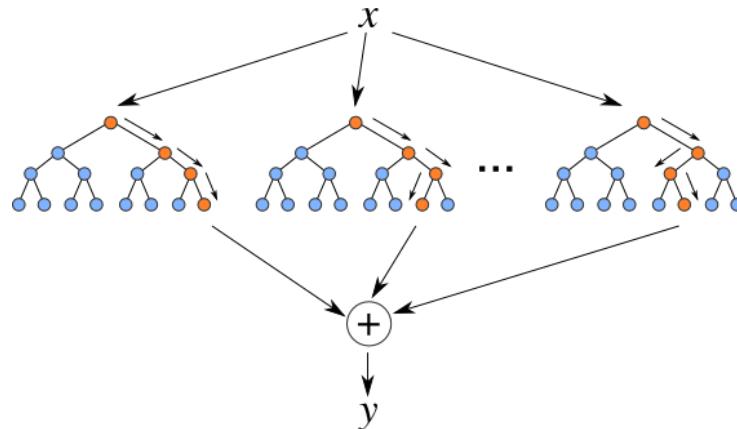


Figura 1: Representación gráfica del modelo Random Forest (Vizcaino, P. 2008).

La cuenca en estudio se encuentra delimitada por el punto de control de la estación fluviométrica de la DGA en los Almendros, siendo parte de la subcuenca Mapocho Alto. Se registraron 104 crecidas en el periodo 1987-2016 durante los meses de abril a octubre inclusive y se obtuvieron 10 posibles variables predictoras para la modelación, las que se derivan de datos de precipitación y temperaturas extremas diarias de estaciones meteorológicas DGA para cada uno de estos eventos. Se utilizó el algoritmo Recursive Feature Elimination (RFE) (Lin *et al.*, 2012) y el índice de correlación entre las variables para seleccionar las que mejor responden a los modelos, y finalmente para comparar y validar los modelos, se incorporaron individualmente en un bootstrap con 500 iteraciones (Efron, 1979; Lopatín *et al.*, 2016), en cada una de estas iteraciones los datos fueron divididos en dos subconjuntos, uno de entrenamiento y uno de prueba, esto se hizo de manera aleatoria y con reposición (Fasnacht *et al.*, 2014). La modelación que se llevó a cabo con los dos métodos, se hizo tanto para el conjunto total de observaciones como el subconjunto de crecidas de caudales mayores. Este subconjunto viene dado por aquellas crecidas que alcanzaron un volumen total de agua drenada igual o superior a 3,5 millones de m³ durante la tormenta. Se utilizó dicho valor, ya que de esta manera se incorporaron todas aquellas crecidas que superan el nivel de pre-alerta en cuanto al caudal máximo alcanzado por éstas, y, aquellas que potencialmente podrían haber llegado a dicho caudal máximo (RODHOS, 2010).

En las modelaciones se obtuvo un R² igual a 0,59 con el método GLM y 0,48 con RF para el conjunto total de observaciones (**Figura 2**), y un R² igual a 0,65 con GLM y 0,31 con RF para el subconjunto de crecidas mayores (**Figura 3**). La variable que más influyó en los modelos fue el Índice de Precipitación Antecedente (IPA), este índice estima las condiciones iniciales de humedad en la cuenca para los eventos de crecida, se utilizó la ecuación 1 para su cálculo (Zimmermann *et al.*, 2000).



Ecuación del Índice de Precipitación Antecedente.

$$IPA_D = 0,95 * P_1 + 0,95^2 * P_2 \dots \dots + 0,95^n * P_n \quad (1)$$

Donde IPA_D es el Índice de Precipitación Antecedente (mm) en el día del evento y $P_{1,2,\dots,n}$ es la precipitación diaria (mm) en los días anteriores 1, 2, ..., n.

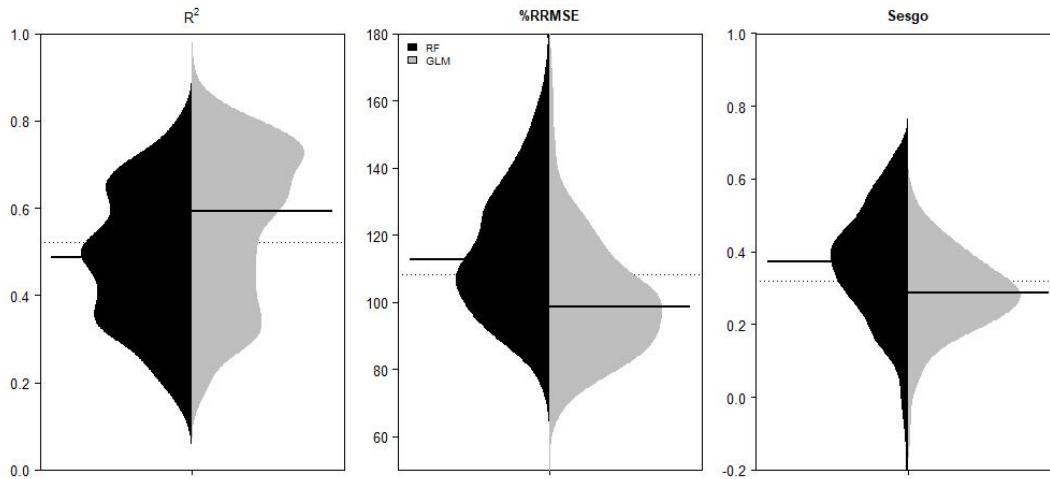


Figura 2: Índices de los modelos con el conjunto total de observaciones de crecidas, RF (negro) y GLM (gris): R^2 , RRMSE y sesgo. Los Beanplots muestran la distribución de los resultados de las 500 iteraciones del bootstrap. Las líneas negras horizontales indican la mediana de las distribuciones. Elaboración propia, 2018.

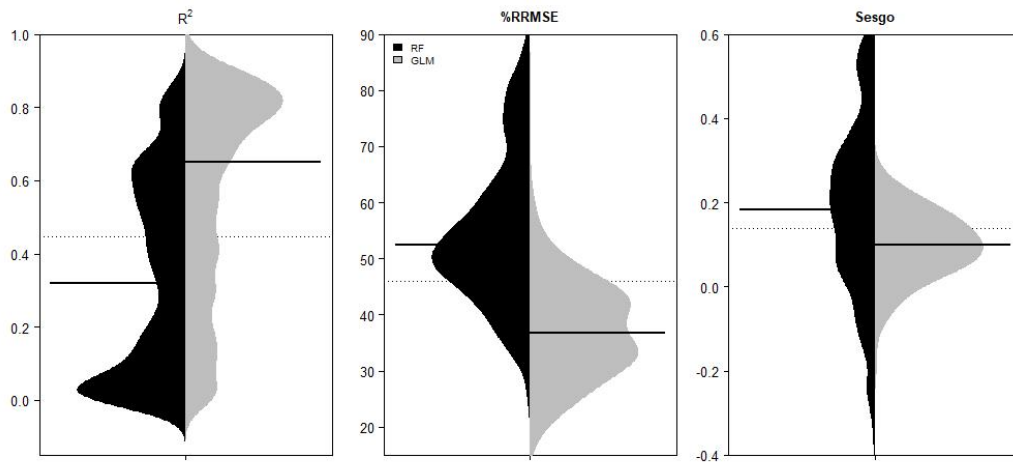


Figura 3: Índices de los modelos con el subconjunto de crecidas mayores, RF (negro) y GLM (gris): R^2 , RRMSE y sesgo. Los Beanplots muestran la distribución de los resultados de las 500 iteraciones del bootstrap. Las líneas negras horizontales indican la mediana de las distribuciones. Elaboración propia, 2018.



Al comparar los métodos RF y GLM, tanto para la modelación de todos los eventos de crecidas como para la modelación de las crecidas máximas, el método GLM alcanzó los resultados más óptimos en cuanto a su precisión y el error asociado. Por último, cabe destacar que el GLM con datos de crecidas mayores presentó mejoras con respecto al R^2 , RRMSE y sesgo, en comparación al GLM con todas las observaciones de crecidas. Este estudio puede considerarse una incursión en la utilización de la herramienta RF en la modelación de volúmenes drenados por crecidas pluviales, ya que, no se documentan estudios previos, por otra parte, el método GLM ha tenido una utilización más extensa en las áreas de la hidrología.

Referencias

Breiman, L. 2001. Random forests. *Machine Learning*, Vol. 45: 5–32 pp.

Efron, B. 1979. Bootstrap methods: Another look at the Jackknife. *Annals of Statistics*, 7: 1–26.

El Mercurio. Santiago. 18 de abril de 2016. 11 p.

Fassnacht, F., Neumann C., Förster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K. & Koch, B. 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, 7: 2547–2561.

Lin, X. et al. 2012. A support vector machine-recursive feature elimination feature selection method based on artificial contrast variables and mutual information. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 910: 149–55.

Lopatín, J.; K. Dolos; H.J. Hernández; M. Galleguillos and F.E. Fassnacht. 2016. Comparison generalized linear model and random forest to model vascular plant species richness using LiDAR data in a natural forest in central Chile. *Remote Sensing of Environment*, 173: 200–210.

Mangeaud, A. & Videla, M. 2005. En Busca de la Independencia Perdida: La Utilización de Modelos Lineales Generalizados Mixtos en Pruebas de Preferencia. *Ecología austral*, 15(2): 199-206.

Oficina Nacional de Emergencia del Ministerio del Interior y Seguridad Pública (ONEMI). 1995. Serie Estudio de Caso, Temporales de Julio – Agosto 1987. [en línea]. Chile: Repositorio Digital ONEMI. Recuperado en: <http://repositoriodigitalonemi.cl/web/bitstream/handle/2012/151/ONE0976.pdf?sequence=1>. Consultado el 28 de Junio de 2018.



Peña, H. & Nazarala, B. 1983. Estudio de Previsión de Crecidas Río Mapocho. (Publicación Interna E.H. N° 83/6), Sub-depto. Estudios Hidrológicos, Depto. de Hidrología, Dirección General de Aguas, Ministerio de Obras Públicas. Santiago, Chile: DGA. 24 p.

RODHOS Asesorías y Proyectos LTDA. 2010. Determinación de los Umbrales de Alerta de Caudales, Lluvias y Temperaturas del Sistema de Transmisión de Datos de la DGA. (S.I.T. N° 202) Dirección General de Aguas, Ministerio de Obras Públicas. Santiago, Chile: DGA. 198 p.

Vizcaino, P. 2008. Aplicación de Técnicas de Inducción de Árboles de Decisión a Problemas de Clasificación Mediante el Uso de WEKA (Waikato Environment for Knowledge Analysis). [en línea]. Bogotá, Colombia: Fac. de Ingeniería de Sistemas, Fundación Universitaria Konrad Lorenz. Recuperado en: http://www.konradlorenz.edu.co/imagenes-1torieess-luma_digital_sistemas/2009_01/final_paula_andrea.pdf Consultado el: 25 de Septiembre de 2017.

Zimmermann, E., Basile, P., López, C., Riccardi, G., Stenta, H., Scuderi, C. & Renteria, J.P. 2000. Condiciones Antecedentes para la Tormenta de Diseño en el Área del Gran Rosario. [en línea]. Rosario, Santa Fe, Argentina: Departamento de Hidráulica, Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario. Recuperado en: https://www.fceia.unr.edu.ar/curiham/Publicaciones/CNA091_ZBL.pdf. Consultado el: 3 de Abril de 2018.