



T0 Interpoladores determinísticos espacio-temporales, series de tiempo y análisis de datos funcionales para el estudio y predicción de la precipitación en Cundinamarca y Bogotá D.C. 2017-2020

Deterministic spatio-temporal interpolators, time series and analysis of functional data for the study and prediction of precipitation in Cundinamarca and Bogotá D.C. 2017-2020

Carlos Eduardo Melo Martínez¹, Diego Alejandro Malagón Márquez² & Dilson David Ramírez Forero³

Para citar este artículo: Melo, C., Malagón, D. y Ramírez, D. (2019). Interpoladores determinísticos espacio-temporales, series de tiempo y análisis de datos funcionales para el estudio y predicción de la precipitación en Cundinamarca y Bogotá D.C. 2017-2020. *Revista de Topografía Azimut*, 10), pp: XX-XX..

Fecha de recepción 1ro de marzo de 2019 – Fecha de aceptación: 10 de mayo de 2019

Resumen

En este trabajo se abordó una metodología para el estudio y la predicción de la precipitación en Cundinamarca y Bogotá D.C. a partir de registros de 133 estaciones meteorológicas del IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales), comprendidos entre enero del 2010 y diciembre del 2016 con una periodicidad mensual, utilizando interpoladores determinísticos espaciotemporales, series de tiempo y análisis de datos funcionales como alternativa a los modelos numéricos y geoestadísticos más comunes. Se obtienen entonces predicciones de precipitación total mensual entre 2017 y 2020 (48 meses), incorporando series de

¹ Doctor en Estadística, Universidad de Barcelona; magíster en Economía, Pontificia Universidad Javeriana; especialista en Estadística, Universidad Nacional de Colombia. Correo electrónico: cmelo@udistrital.edu.co

² Ingeniería Catastral y Geodesta, Universidad Distrital Francisco José de Caldas. Correo electrónico: damalagonm@correo.udistrital.edu.co

³ Ingeniería Catastral y Geodesta, Universidad Distrital Francisco José de Caldas. Correo electrónico: ddramirezf@correo.udistrital.edu.co

tiempo como un complemento al proceso de interpolación; adicionalmente, el análisis de datos funcionales permite obtener datos de velocidad y aceleración de la precipitación, los cuales, junto con las predicciones realizadas, se convierten en una herramienta fundamental como insumo en la planeación, el ordenamiento territorial y la gestión y prevención del riesgo.

Palabras clave: análisis de datos funcionales, interpoladores determinísticos espacio-temporales, mapeo espacio temporal en R, predicción de precipitación, series de tiempo.

Abstract

In this research was used a methodology for the study and prediction of the precipitation in Cundinamarca and Bogota DC, Colombia, from 133 records of meteorological stations of the IDEAM (Institute of Hydrology, Meteorology and Environmental Studies), between January 2010 and December 2016 with a monthly frequency, it was used spatial-temporal deterministic interpolators, time series and functional data analysis, as an alternative to the most common numerical and geostatistical models. Space-time predictions are then obtained for total monthly precipitation in the years 2017 to 2020 (48 months), also incorporating time series, as a complement to the interpolation process. Additionally, the functional data analysis allows to obtain data on the speed and acceleration of precipitation, which together with the predictions made become fundamental tool as an input in planning, territorial ordering and risk management and prevention.

Keywords: functional data analysis, prediction of precipitation, spatial-temporal deterministic interpolation, spatial-temporal mapping in R, time series.

T1 Introducción

Predecir la precipitación no es un tema del todo novedoso, pues incluso en el ámbito espacio temporal existen diferentes enfoques, como los métodos numéricos WRF y MM5 usados actualmente por el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales; Arango y Ruiz, 2012), y los métodos geoestadísticos modelados desde variogramas espacio temporales como el Kriging-ST (Bergonzoli, Bernards, ter. Borg y Bookmaker, 2011; Cheng, Hsieh y Wang, 2007; Phillips, Dolph y Marks, 1992).

Estos enfoques han servido para establecer predicciones de corto plazo, las cuales han sido valiosas para generar reportes climáticos y alimentar los sistemas de alerta temprana; sin embargo, actualmente se enfrentan a un nuevo reto, su aplicación como insumo en planes territoriales, de desarrollo y de gestión de riesgos (Ruiz y Escobar, 2012), ante el que estos enfoques suelen presentar limitantes importantes (Zaw y Naing, 2009) en cuanto a su alta complejidad, cortos intervalos de predicción y escala (Dillon, García y Nicolini, 2013).

En respuesta a esta problemática, se usarán interpoladores determinísticos espaciotemporales como lo son IDW-ST (distancia inversa ponderada espacio-tiempo) y RBF-ST (funciones de base radial espacio-tiempo), para generar predicciones de precipitación total mensual (en un periodo administrativo), haciendo uso de *software* libre y con funcionalidades de mapeo como lo es R; además, se desarrolla una metodología que usa los modelos autorregresivos de series de tiempo, complementando a los interpoladores determinísticos, con el objetivo de mejorar las estimaciones y la variación a través del tiempo (Zhang, 2003). Finalmente, el análisis de datos funcionales incorpora una novedosa herramienta que responde a la naturaleza funcional de los datos, derivando y obteniendo información cambio y aceleración de precipitación en el tiempo (Zhang, Wang y Mei, 2017).

T1 Materiales y métodos

T2 Zona de estudio

Cundinamarca y Bogotá D.C., localizada en la parte central de Colombia, cuenta con una superficie promedio de 22 633 km², limitando con los departamentos de Boyacá, Meta, Huila, Tolima y Caldas (Figura 1). Durante la mitad del año se encuentra bajo una temporada lluviosa, con un comportamiento de tipo bimodal (Jaramillo y Chaves, 2000), además, tiene una buena porción del territorio en pendientes bajas y limita con accidentes geográficos como la cordillera Oriental y el Valle del Magdalena; por ello, es una región con riesgo de inundaciones esporádicas, desbordamientos, saturación del suelo y colapso del sistema de alcantarillado. La región presenta un promedio anual de lluvias de 797 mm, distribuido en dos temporadas secas y dos temporadas lluviosas (IDEAM, 2010).

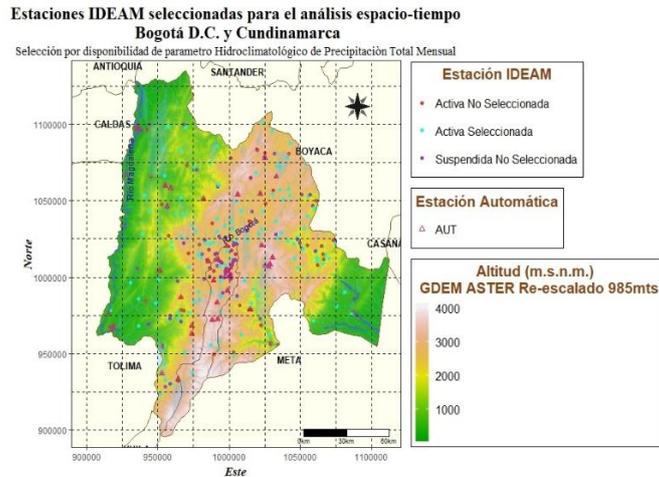


Figura 1. Estaciones IDEAM seleccionadas para el análisis espacio-tiempo
Fuente: IDEAM, 2010.

T2 Datos

Los datos de precipitación fueron obtenidos a través de 133 estaciones meteorológicas del IDEAM. Datos mensuales durante siete años con valores de precipitación entre 0 y 1109 mm y faltantes de 48.8% (Tabla 1); además, para obtener datos de elevación se usó el DEM 30mts Global Aster GDEM en formato ráster.

	Mínimo	Medio	Máximo
Coord. Norte	923674.3	1018092	1128535.
Coord. Este	4	.64	73
Altura	915544.7	1002217	1086158.
Precipitación	9	.90	75
No. Faltantes	235.00	2099.08	3195.00
%Faltantes	0.00	129.18	1109.00
	1.00	8.79	41.00
	1.19	10.46	48.81

Tabla 1. Características generales de la muestra de estaciones meteorológicas seleccionadas

T2 Software

Se realizaron rutinas computacionales en el *software* R, usando paquetes como: Geosptdb, para la optimización, predicción, interpolación y validación cruzada para IDW-ST y RBF-ST (Melo y Melo, 2015); el paquete Forecats para realizar el modelado de series de tiempo (Hyndman y Shahid, 2007); Fda para el análisis de datos funcionales (Ramsay y Silverman, 2005); Ggplot2 y Spacetime para la generación de gráficos y mapas.

T2 Metodología

Este trabajo se desarrolló en cinco fases, que, a continuación, serán explicadas. Debe considerarse que se han utilizado seis interpoladores determinísticos espacio-temporales, funciones de base radial trigonométrica (TRI), *spline* con tensión (ST), completamente regularizada *spline* (CRS), multicuadrática (M) y multicuadrática inversa (IM) que optimizan los parámetros de Eta y Rho, por tratarse de un enfoque del grado de suavizado de las superficies estimadas; de otra manera, el interpolador distancia inversa ponderada (IDW, por sus siglas en inglés) optimiza los parámetros de C y P, por tratarse de un enfoque de extensión del tiempo como una equivalente dimensión espacial.

Por otra parte, las optimizaciones de los interpoladores se basan en la selección óptima de los vecinos más cercanos (Knn), los cuales son seleccionados con base en las distancias desde cada observación hasta la observación vecina más cercana, y que junto con la determinación del error medio cuadrático de predicción (RMSPE, por sus siglas en inglés) se convierten en los criterios para la selección de parámetros óptimos.

- Preprocesamiento: la existencia de valores faltantes en los registros de precipitación implica que en esta fase los datos faltantes se imputen para obtener una base de datos completa, para ello se comienza con una etapa de depuración, seleccionando únicamente aquellas estaciones con más del 70% de registros capturados; después, mediante interpoladores determinísticos, se imputan los registros faltantes mediante el modelo con mejores resultados.
- Método puro: esta fase se compone de tres etapas, en la primera se optimizan los parámetros de las funciones de interpolación con un Knn de hasta 304 y minimizando el RMSPE. Con dichos valores se procede a realizar cada

interpolación, para finalizar con una etapa de comparación de los resultados, haciendo uso de criterios de error (RMSPE), variabilidad temporal y distribución espacial.

- **Método híbrido:** esta fase inicia con el modelado de series de tiempo mediante modelo estacional autorregresivo integrado de medias móviles (SARIMA, por sus siglas en inglés), generando para cada estación de muestreo pronósticos en los 48 meses del rango de predicción, permitiendo a los interpoladores determinísticos trabajar en un entorno de interpolación y no de extrapolación. Tras este proceso se surten las mismas tres etapas del método puro, es decir, la optimización de los parámetros, la interpolación y, finalmente, la comparación.
- **Comparación:** en esta fase se comparan, para cada uno de los métodos, los mejores resultados generados, haciendo uso del criterio de error, de variabilidad temporal y de distribución espacial, determinando la utilidad y calidad de los pronósticos. Al final de esta fase se determinan las ventajas y desventajas que presentan los dos métodos utilizados.
- **Análisis de datos funcionales:** sobre las mejores estimaciones de las anteriores fases se realiza un análisis desde el aspecto funcional, se reconstruyen las formas funcionales, se realiza primera y segunda derivada, por lo cual se encuentra información de la tasa de cambio y de la aceleración en el tiempo de la cantidad de precipitación en cada punto, para finalmente mapear cada uno de estos comportamientos en los 48 meses estimados.

T1 Resultados

T2 Preprocesamiento

En la Tabla 2 se encuentran los resultados de la interpolación que, junto con la combinación de los parámetros óptimos, se ha realizado para la imputación de datos.

Interpolador	RMSPE	Máximo	Mínimo	Variación Temporal
IDW	94,2489	1109,00	0,00	Aceptable
TRI	93,3094	1109,00	0,00	Aceptable
ST	51,5898	1109,00	-203,00	Aceptable
CRS	67,4570	1109,00	0,00	Insuficiente

M	91,4791	1733,87	-926,00	Insuficiente
IM	50,4293	1109,00	0,00	Aceptable

Tabla 2. Resultados obtenidos en imputación

De esta manera, la función de base radial IM es el mejor interpolador; por su bajo error de predicción, variación temporal adecuada y correcta estimación de valores ha sido seleccionado para la imputación de datos faltantes.

T2 Método puro

Una vez imputados los datos faltantes utilizando IM como óptimo interpolador, se realiza nuevamente la optimización de parámetros con la base de datos completa. Se puede evidenciar en la Tabla 3 que la función IM genera nuevamente los mejores resultados asociados a RMSPE.

Tabla 3. óptimos método IM generó espaciales y inadecuados con

Método	KNN ⁽⁴⁾	RMSPE	Eta ⁽⁵⁾	Rho ⁽⁶⁾	
CRS	12	77.8845	0.0953	0.1412	Parámetros puro resultados temporales
M	4	58.4179	2	2	
IM	6	47.3938	0.0776	1.7506	
TRI	12	114.0048	0.113565	0.09958	
ST	12	77.88925	0.045278	0.18603	
Método	KNN	RMSPE	C ⁽⁷⁾	P ⁽⁸⁾	
IDW	54	92.55627	0.040073	0.56037	6 Knn, pero

presentó notorias mejorías al aumentar la cantidad de estos. Gracias a que se genera una variación mínima del error, fue posible evaluarlo con 204 Knn, obteniendo mejores resultados y con un aumento del error de solo 0.1212 mm; en cuanto a la variación temporal, IM bajo la nueva condición es el mejor interpolador, pues no presenta cambios mínimos o nulos como sí los presentan las demás funciones, su variación espacial es también la mejor de todas, seguida de IDW que presenta uno de los RMSPE

⁴ Número de vecinos más cercanos

⁵ Parámetro de suavizado de funciones radiales

⁶ Parámetro de distancia espacio-tiempo entre dos puntos

⁷ Factor de unidad de distancia espacial/unidad de tiempo

⁸ Exponente para la ponderación en la estimación

más bajos y una buena variación espacial, aunque su variación temporal no sea la mejor. Los resultados se aprecian de mejor manera en la Figura 2, donde se observan las predicciones para tres periodos en los cuatro años, así como en el resumen de resultados de la Tabla 4.

Interpolador	RMSPE	Mínimo	Máximo	Variación Temporal	Distribución Espacial
IDW	92.5563	44.91	473.16	Insuficiente	Adecuada
TRI	114.0048	- 64239.0	11504.75	Insuficiente	Insuficiente
ST	77.8893	- 2321.60	3357.60	Inadecuada	Aceptable
CRS	77.8845	- 1118.18	1654.72	Insuficiente	Insuficiente
M	58.4179	- 3515.16	4620.85	Inadecuada	Aceptable
IM	47.4851	6.2130	575.275	Adecuada	Adecuada

Tabla 4. Resultados finales método puro

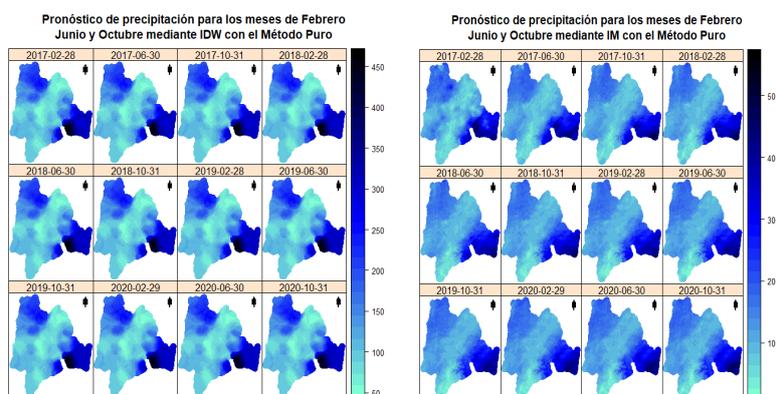


Figura 2. Pronóstico para los meses de febrero, junio y octubre en los cuatro años de predicción para los modelos con mejor distribución espacial

T2 Método híbrido

Haciendo uso de la función Forecast de R, se generaron modelos automáticos SARIMA minimizando los criterios de AIC, BIC, AICc y σ^2 , obteniendo en su mayoría modelos con parámetros entre 0 y 2, en los que al menos 41 de los 133 creados presentan media diferente de cero, dieciocho poseen tendencia en la media y, en general, los modelos tienen relevancia estacional y buena distribución temporal, de modo que es posible encontrar modelos sin componente ARIMA (siglas en inglés de

Interpolador	RMSPE	Mínimo	Máximo	Variación Temporal	Distribución Espacial
--------------	-------	--------	--------	-----------------------	--------------------------

modelo autorregresivo integrado de medias móviles), pero ninguno sin los elementos estacionales (Figura 3).

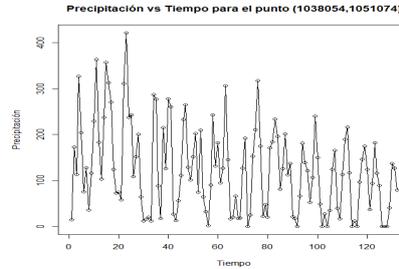


Figura 3. Pronóstico series de tiempo para un punto aleatorio

Una vez construidas las series de tiempo para los 48 meses, se complementó la base de datos y se optimizaron los parámetros. En la Tabla 5 se evidencia claramente que por criterio de error el

es nuevamente

Método	KNN	RMSPE	Eta	Rho
CRS	20	78.7077	0.107	0.1015
M	4	51.1251	2	2
IM	4	40.6685	0.0772	0.7823
TRI	4	51.1227	1e-05	1.9999
ST	4	51.1227	0.00001	0.1841
Método	KNN	RMSPE	C	P
IDW	63	84.19135	0.35215	0.43817

mejor interpolador

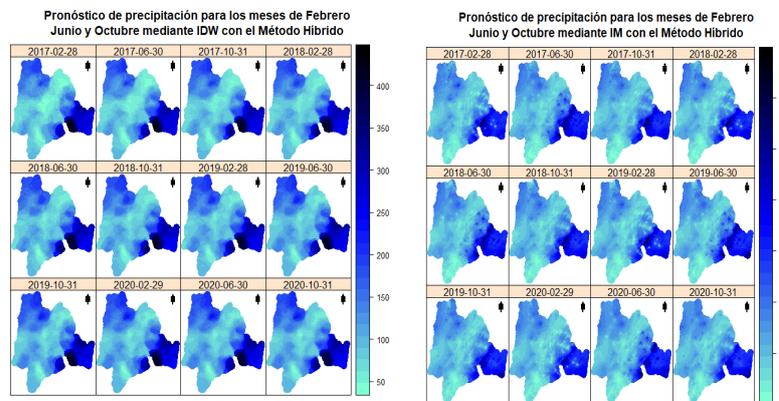
IM.

Tabla 5. Parámetros óptimos método híbrido

Como sucedió en el método puro, IM presenta mejores resultados aumentando el Knn, y gracias a la poca sensibilidad que tiene el RSMPE respecto al Knn se pudo evaluar hasta con 504 de ellos, mejorando los resultados y aumentando el error en tan solo 0.0899 mm. En cuanto a la variabilidad temporal, hay una mejora significativa, eliminándose el problema de periodos sin variación que se observaba en el método puro (Figura 4); IDW, de nuevo, después de IM, es la función con mejor fluidez temporal y espacial, aun cuando presente el mayor error de estimación. Todas las funciones, excepto IM e IDW, presentan una gran subestimación de los valores de precipitación como se ve en la Tabla 6.

IDW	84.1914	34.38	453.11	Adecuada	Adecuada
TRI	51.1227	-	7073.59	Aceptable	Aceptable
		2475.79			
ST	51.1227	-	7073.59	Aceptable	Aceptable
		2475.79			
CRS	78.7077	-	8347.36	Aceptable	Aceptable
		2226.75			
M	51.1251	-	7073.58	Aceptable	Aceptable
		2842.20			
IM	40.7585	0.00	736.68	Adecuada	Adecuada

Tabla 6.
finales método



Resultados
híbrido

Figura 4. Pronóstico para los meses de febrero, junio y octubre en los cuatro años de predicción, para los modelos con mejor distribución espacial

12 Comparación

En la Tabla 7 se presentan en resumen los resultados para el mejor modelo del método puro e híbrido (función IM), tienen similitudes en la distribución final de valores en el espacio; sin embargo, existen diferencias en el criterio de variabilidad temporal.

Interpolador	RMSPE	Mínimo	Máximo	Variación Temporal	Distribución Espacial
Híbrido	40,7585	0,000	736,680	Adecuada	Adecuada
Puro	47,4851	6,213	575,275	Adecuada	Adecuada

Tabla 7. Comparación métodos puro e híbrido

Por otro lado, se evidencia una limitante importante respecto al método híbrido, viene dada por el tipo de enfoque global, pues, mientras en el método puro 204 Knn son suficientes, son necesarios 504 Knn en híbrido, lo cual genera aumentos en tiempo de procesamiento.

12 Análisis de datos funcionales

El acercamiento de datos funcionales se hace desde bases de Fourier con 300 funciones base, para reconstruir el comportamiento de la precipitación se generan 51 472 curvas, correspondientes a los puntos de la malla espacial de predicción y con un rango de continuidad de 48 meses para el método híbrido. Enseguida, se calcula la primera derivada de las funciones, encontrando gran variación a lo largo de todo el rango de pronóstico (Figura 5) y, como se observa en la Figura 6, es posible identificar puntos de velocidad ascendente y descendente; luego, con el cálculo de la segunda derivada, se encuentran resultados con menor variación (Figura 5), pero se logran identificar cambios en las dinámicas y aceleración de precipitación (Figura 6).

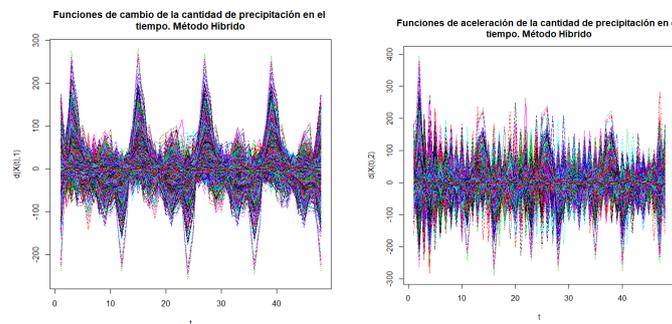


Figura 5. Funciones de velocidad y aceleración de la precipitación

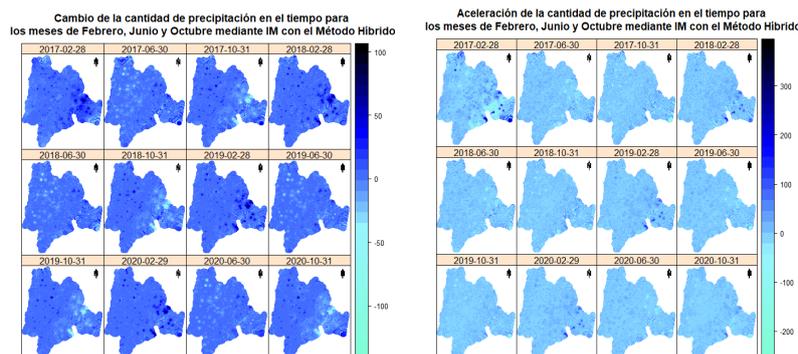


Figura 6. Mapas de velocidad y aceleración de la precipitación **T1 Conclusiones**

Los interpoladores determinísticos espacio-tiempo son adecuados para realizar pronósticos de precipitación en periodos de tiempo prolongados y a escalas adecuadas para implementar en planes de ordenamiento territorial, de desarrollo y de gestión de riesgo de desastres. Además, el uso de series de tiempo dentro del proceso reduce el error y mejora la variabilidad temporal, generando así predicciones más confiables.

Para las predicciones de precipitación en la zona de estudio es recomendable hacer uso de IM e IDW, con un número considerable de vecinos más cercanos ($KNN > 100$); también la existencia de valores negativos y atípicos es una posibilidad constante en estos métodos de interpolación, por lo que siempre es aconsejable realizar procesos de filtrado y análisis de dichos valores con el fin de determinar si las predicciones son o no adecuadas.

Aunque en el presente trabajo se incorporaron variables como x , y (localización espacial), t y altura, es recomendable alimentar el modelo con otras covariables que pueden explicar el modelado mediante la aplicación de tendencia.

Debido a los buenos resultados al aplicar modelado de series de tiempo, para futuros estudios se recomienda explorar las máquinas de vectores de soporte (SVM, por sus siglas en inglés), pues en la literatura se evidencia que ha superado a los modelos ARIMA, por lo que puede mejorar aún más las predicciones. Finalmente, cabe mencionar que el análisis de datos funcionales es una herramienta útil para generar información secundaria a partir de las predicciones de precipitación, y es posible usar la información obtenida mediante derivación como insumo para todos los meses de pronóstico en el método híbrido.

T1 Referencias

Arango, C. y Ruiz, J. (2012). Implementación del Modelo WRF para la Sabana de Bogotá. Recuperado de http://www.ideam.gov.co/documents/21021/21132/Modelo_WRF_Bogota.pdf/f1d34638-e9f8-4689-b5f4-31957c231c46

- Bergonzoli, F., Bernards, W., ter. Borg, R., & Bookmaker. (2011). Grenzeloze technologie 1985-1995. UD y la GEOMÁTICA (Vol. 0). The Reader's Digest. Retrieved from <http://revistas.udistrital.edu.co/ojs/index.php/UDGeo/article/view/4407/7027>
- Cheng, S., Hsieh, H. and Wang, Y. (2007). Geostatistical Interpolation of Space-Time Rainfall on Tamshui River Basin, Taiwan. *Hydrological Processes*, 21(23), 3136–3145.
- Dillon, E., García, A. y Nicolini, M. (2013). Desempeño del pronóstico de modelos de alta resolución, en un área limitada: análisis de la estación de verano 2010-2011. *Meteorológica*, 38(2), 69-89.
- Hyndman, R. and Shahid, M. (2007). Robust Forecasting of Mortality and Fertility Rates: a Functional Data Approach. *Computational Statistics and Data Analysis*, 51(10), 4942–4956.
- Instituto de Hidrología, Meteorología y Estudios Ambientales [IDEAM] (2010). Características Climatológicas de Ciudades Principales y Municipios Turísticos. Recuperado de <http://www.ideam.gov.co/documents/21021/21789/1Sitios+turisticos2.pdf/cd4106e9-d608-4c29-91cc-16bee9151ddd>
- Jaramillo, Á. y Chaves, B. (2000). Distribución de la precipitación en Colombia analizada mediante conglomeración estadística. *Cenicafé*, 51(2), 102–113.
- Melo, C. y Melo, O. (2015). Geosptdb: spatio-temporal radial basis functions with distance-based regression and inverse distance weighting. Recuperado de <https://rdrr.io/cran/geosptdb/man/geosptdb-package.html>
- Phillips, D., Dolph, J. and Marks, D. (1992). A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology*, 58(1–2), 119–141.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis* (2ª ed.). Nueva York: Springer-Verlag.

- Ruiz, J. y Escobar, O. (2012). Alteraciones de la precipitación y la temperatura ante la variabilidad y cambio climático para la ciudad de Bogotá. Recuperado de <http://modelos.ideam.gov.co/media/dynamic/escenarios/escenarios-cambio-climatico-bogota.pdf>
- Zaw, W. and Naing, T. (2009). Modeling of Rainfall Prediction over Myanmar Using Polynomial Regression. Trabajo presentado en *International Conference on Computer Engineering and Technology*, Buenos Aires.
- Zhang, P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 50.
- Zhang, R., Wang, J. and Mei, Y. (2017). Search for Evergreens in Science: a Functional Data Analysis. *Journal of Informetrics*, 11(3), 629–644.