

Modelado de la precipitación diaria en el sistema de chorro de bajo nivel de las Grandes Llanuras Americanas

Gleisis Alvarez-Socorro^{1*}, Luis Gimeno-Sotelo^{2,3}, Javier Roca-Pardiñas⁴, Luis Gimeno^{1,5}

¹ Centro de Investigación Mariña, Environmental Physics Laboratory, Universidad de Vigo, Ourense, España

² Departamento de Ciências Matemáticas, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

³ CEAUL - Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

⁴ Departamento de Estadística e Investigación Operativa, Universidad de Vigo, 36310, Vigo, España

⁵ Centro de Supercomputación de Galicia (CESGA), Avenida de Vigo, 3, 15705, Santiago de Compostela, España

Resumen

Los modelos de regresión son una herramienta muy útil en las ciencias atmosféricas, tanto para el trabajo científico como el operativo ya que permiten describir y cuantificar las relaciones entre variables meteorológicas. Este estudio pretende utilizar un conjunto de diferentes modelos de regresión para analizar la relación entre la precipitación diaria en las Grandes Llanuras Americanas y sus principales determinantes: la humedad transportada, el vapor de agua total en columna y la inestabilidad atmosférica. Para ello se cuenta con una serie de valores diarios de humedad transportada, inestabilidad atmosférica, agua total en la columna y precipitación, de los meses junio-julio-agosto, entre 1980 y 2017. Se empleó el criterio BIC para la selección de las variables predictoras más significativas. Los cuantiles utilizados en el análisis fueron 0.25, 0.50, 0.75, 0.95, 0.99. Se utilizaron diferentes modelos de regresión, desde los más simples a los más complejos, siendo el modelo de localización y escala el que presentó los mejores resultados. Este modelo representó correctamente la distribución de la precipitación para los diferentes cuantiles y permitió identificar las variables predictoras que más influyeron en la variabilidad de dicha precipitación.

1. Introducción

Los métodos estadísticos juegan un papel central en la meteorología y las ciencias del clima, ya que proporcionan herramientas poderosas para identificar patrones, cuantificar las relaciones y predecir las condiciones futuras. El análisis de regresión, en particular, ha demostrado ser invaluable para vincular variables meteorológicas y comprender sus efectos combinados en los procesos atmosféricos (**Wilks, 2011**). Los avances como la regresión cuantil y los modelos flexibles como los modelos aditivos generalizados (GAM) han mejorado la capacidad de capturar comportamientos no lineales, estacionales y extremos en los datos climáticos (**Hastie y Tibshirani, 1990; Wood, 2006**). Estos enfoques son especialmente adecuados para los estudios de precipitación, donde la variabilidad y los extremos a menudo dominan y los supuestos lineales rara vez se mantienen. Numerosas aplicaciones han explorado las relaciones entre la precipitación y los fenómenos climáticos a gran escala, los controles topográficos y los impulsores físicos, subrayando la versatilidad de los modelos de regresión en este campo (**Rodó et al. 1997; Ranhao et al., 2008; Gimeno-Sotelo et al., 2024a**).

La precipitación está fuertemente influenciada por tres impulsores principales: inestabilidad atmosférica, vapor de agua de columna total y transporte de humedad (**Gimeno-Sotelo et al., 2023**). Sin embargo, sus efectos combinados son altamente no lineales y varían según la escala de tiempo y la región. Una de las regiones más estudiadas afectadas por chorros de bajo nivel, es la de las Grandes Llanuras Americanas, y juega un papel fundamental al transportar grandes cantidades de humedad del Golfo de México (**Gimeno et al., 2016**). Su presencia mejora la intensidad de la precipitación, la frecuencia y la distribución espacial, mientras que su ausencia puede desencadenar eventos de sequías (**Algarra et al., 2019**). En trabajos anteriores se ha demostrado que los extremos de humedad transportados por el chorro de bajo nivel de esta región, modula fuertemente la dependencia entre la inestabilidad y la precipitación (**Gimeno-Sotelo et al., 2022**). Sobre la base de esto, el presente estudio investiga la distribución completa de la precipitación diaria y los cuantiles importantes, a través de modelos de regresión que vinculan la precipitación con la humedad transportada, el vapor de agua de la columna y la velocidad vertical. Al combinar enfoques lineales y

basados en GAM, el estudio ofrece una evaluación integral de la variabilidad de la precipitación en relación con el chorro de bajo nivel y sus impulsores clave.

2. Datos y metodología

2.1 Datos

Esta investigación utiliza datos diarios para la temporada de verano (junio, julio y agosto) durante un período de 38 años (1980-2017; 3496 observaciones) para la precipitación que ocurre en la región del sumidero del chorro y tres de sus principales conductores: 1. La humedad transportada (TM, *mm/día*) ($E-P > 0$, la evaporación predomina sobre la precipitación) desde la región fuente de humedad a la región de la región sumidero del chorro de bajo nivel; 2. Vapor de agua total en la columna (TCWV, *kg/m²*) y 3. Velocidad vertical (VV, *Pa/S*) en la región del sumidero del chorro.

Los datos de precipitación (*mm/día*) se obtuvieron del Centro de Predicción Climática (CPC), mientras que la humedad transportada se derivó del modelo Lagrangiano FLEXPART, rastreando las ganancias de humedad de la parcela de aire de la fuente a las regiones sumidero. La VV se calcula como la media de omega (ω) (componente vertical de la velocidad en las coordenadas de presión) a 850 hPa. Siguiendo la metodología de investigaciones previas, la variable $-\omega$ (-Omega) se utilizó como métrica para la velocidad vertical en esta investigación. Es la medida que se utiliza para representar la inestabilidad atmosférica (Holton, 1973).

2.2 Modelos de regresión

El objetivo principal de las técnicas estadísticas es evaluar cómo las variables explicativas (covariables) influyen en una variable de respuesta, a menudo a través de modelos de regresión de complejidad variable. Cuando están involucrados múltiples predictores, el análisis se lleva a cabo utilizando regresión multivariada (o múltiple), que examina la relación entre una variable dependiente y varias variables independientes. En este estudio, dicho enfoque se aplica para analizar la precipitación (variable de respuesta) en relación con tres variables explicativas: TM, -Omega y TCWV.

2.2.1 Modelos lineales

La regresión, en términos generales, se formaliza como la media condicional de la variable respuesta Y en función del valor de las p variables explicativas $X = (X_1, \dots, X_p)$. En este contexto se puede considerar el modelo $Y = \mu(X) + \varepsilon$, donde se supone que los residuos ε son variables aleatorias de media cero. La estructura elegida para $\mu(X) = E[Y | X]$ determina el modelo de regresión considerado. El modelo más simple es que el supone que el efecto de las covariables es lineal y puede asumirse que (Nester et al., 1996):

$$\mu(X) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p,$$

donde $(\beta_0, \dots, \beta_p)$ son coeficientes desconocidos que necesitan ser estimados en base a un simple datos $n \{(X_i, Y_i)\}_{i=1}^n$.

2.2.2 Modelos aditivos generalizados

Los modelos lineales son limitados porque asumen una relación constante y estrictamente lineal entre la variable de respuesta y sus predictores, lo que restringe su capacidad para capturar dinámicas más complejas. Para abordar esto, los modelos aditivos (Hastie y Tibshirani, 1990; Wood, 2017) extienden modelos lineales al permitir que el efecto de cada predictor se exprese a través de funciones suaves $f_i(X_i)$. Este tipo de modelos permite una mayor variedad de efectos de covariables que son continuas, que incorporan relaciones no lineales entre las diferentes variables explicativas y la media de la variable respuesta, así como las interacciones entre las variables dichas. En estos modelos, el predictor lineal se reemplaza por un predictor aditivo, lo que resulta en el siguiente modelo

$$Y = \mu(X) + \varepsilon = \beta_0 + f_1(X_1) + \dots + f_p(X_p) + \varepsilon$$

2.2.3 Modelos de localización y escala

Para tener un modelo de regresión más flexible, con una mejor capacidad para capturar la posible relación que existe entre las variables predictoras a tener en cuenta y la variable respuesta, se propone el uso de los modelos de ubicación y escala, ya que incluyen la varianza como parte del modelado. Por esta razón, son particularmente útiles cuando se sospecha que la varianza de la variable respuesta no es homogénea, sino que depende del nivel de uno o más predictores. Los modelos de ubicación y escala expresan la media condicional y la varianza condicional como funciones no paramétricas aditivas de las covariables (Silva et al., 2016). Se supone un modelo de ubicación y escala:

$$Y = m(X) + \sigma(X)\varepsilon$$

donde m y σ son funciones desconocidas y ε es la variable de error cuya distribución no depende de X . En resumen, cuando ambos aspectos se combinan en este tipo de modelo, hace posible describir no solo cómo la variable respuesta cambia en una función de las variables predictoras, sino también cómo cambia la variabilidad de dicha variable respuesta (Viechtbauer y López-López, 2022).

2.4 Criterio de información bayesiano

Existen varios criterios para comparar diferentes modelos, basado en algún principio estadístico subyacente. En este trabajo se utilizó el criterio BIC, el cual fue introducido por Gideon E. Schwarz en 1978 y está estrechamente relacionado con el criterio AIC, publicado en 1974. El BIC se deriva para servir como una aproximación asintótica a una transformación de la probabilidad bayesiana a posteriori de un modelo candidato. El cálculo del BIC se basa en la probabilidad logarítmica empírica y no requiere la especificación a priori, de ahí que sea de gran utilidad para problemas de modelados donde los antecedentes son difíciles de establecer con precisión (Neath y Cavanaugh, 2012).

En este trabajo, para la selección de las variables en los modelos se aplicó un algoritmo que probaba todas las posibles combinaciones entre dichas variables a incorporar. La combinación de las variables en cada uno de los modelos se evalúa, tanto de una, dos o la cantidad total posible y se calcula su valor de BIC correspondiente. Seguidamente, se ordenan los modelos según el valor de BIC obtenido en forma ascendente y se selecciona como mejor modelo aquel que presente el menor valor de BIC. En resumen, con este criterio de BIC, se proporciona un equilibrio entre la complejidad del modelo y el ajuste de los datos.

3. Resultados y discusiones

En esta sección, se aplicarán los enfoques de los modelos de regresión explicados anteriormente, para analizar la relación entre la precipitación asociada con el chorro de bajos niveles de las Grandes Llanuras Americanas, la inestabilidad atmosférica (-Omega), el vapor de agua total en la columna (TCWV) y la humedad transportada (TM) desde la región fuente. El lenguaje de programación utilizado para analizar esta relación es R (R Core Team, 2022).

3.1 Modelos lineales

Se realizó un primer análisis con el modelo más sencillo: el modelo lineal, con la precipitación como variable dependiente en función del resto de las variables predictoras. Para ello se utilizó la función “lm()” del paquete “stats”. El ajuste de dicho modelo presentó un R^2 de 0.2126 y un R^2 ajustado de 0.2119. Esto significa que el modelo ajustado no es capaz de representar correctamente la relación entre dichas variables y que las variables independientes sólo explican aproximadamente el 21% de la variabilidad de la precipitación. Por otra parte no se cumplieron las hipótesis en la diagnosis del modelo. Además se analizó la autocorrelación de los residuos a partir de las gráficas ACF para comprobar si el modelo planteado ha captado toda la información relevante de los datos analizados. En la **Figura 3.1** se muestra el gráfico de autocorrelación para este modelo lineal, demostrando la fuerte autocorrelación que hay en los primeros retardos.

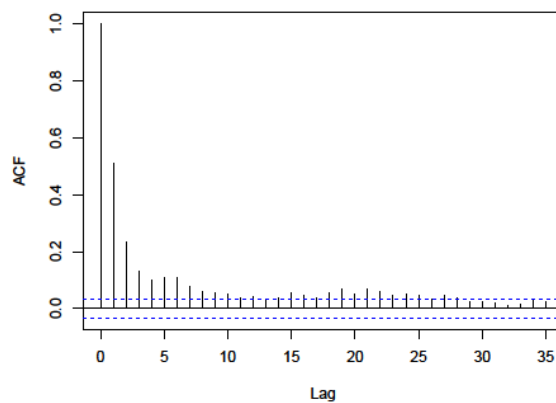


Figura 3.1: Autocorrelación de los residuos del modelo lineal original de la precipitación en función de la inestabilidad atmosférica, la humedad transportada y el vapor de agua total en la columna.

Se pueden aplicar diferentes métodos para solucionar este problema de autocorrelación, por ejemplo, aplicar transformaciones a los datos, ajustar un modelo diferente o más complejo, o introducir términos de retardos adicionales. Teniendo en cuenta que el análisis realizado es de variables meteorológicas, en las cuales, en muchas ocasiones las condiciones de un día se ven influenciadas por el día anterior, se consideró que la última cuestión mencionada puede ser una solución. Para corregir esta autocorrelación en los residuos del modelo, se incorporaron como nuevas variables predictoras, las mismas variables independientes pero con retardos igual a 1, incluso de la propia variable precipitación. Se pasa de contar con 3 variables para analizar la precipitación a tener 7 variables. Se debe tener en cuenta, que en el momento de incorporar estas nuevas variables, la base de datos utilizada se modifica ligeramente, ya que para incorporar el primer retardo, la serie tendrá que comenzar el 2 de junio de 1980. En resumen, el modelo de regresión lineal sería la precipitación de un día, en función de la inestabilidad, el transporte de humedad y el agua total en la columna de ese mismo día, así como de la inestabilidad, el transporte de humedad, el agua total en la columna y la precipitación del día anterior.

Separación de la base de datos

La nueva base de datos modificada ligeramente con la nueva incorporación de las variables con retardo igual a 1, es considerada entonces, como la base de datos final para realizar el análisis. Es por ello que, teniendo todas las variables a considerar en el modelo, se decidió separarla en dos muestras. Una de ellas será la muestra de entrenamiento para ajustar los modelos de regresión que se van a utilizar y una muestra test, para evaluar la precisión de las predicciones. Para ello se consideró como muestra de entrenamiento el 80% de los datos y como muestra test el 20% restante.

Selección de variables

Al incorporar nuevas variables predictoras en el modelo, surge la necesidad de aplicar un criterio de selección para reducir la complejidad del mismo y a su vez tener en cuenta solo las variables más significativas. El criterio aplicado para la selección fue el BIC, el cual busca un equilibrio entre la bondad de ajuste del modelo y su complejidad, imponiendo una penalización más fuerte por el número de parámetros estimados, en comparación con el criterio AIC. Se analizaron todas las combinaciones posibles de variables explicativas, incluyendo las que tienen lag 1. En la **Tabla 3.1** se presentan las variables a incluir de los primeros 5 mejores modelos según este criterio.

Tabla 3.1: Variables incorporadas de los 5 mejores modelos lineales según el criterio de selección de variables BIC.

Orden	-Omega	TM	TCWV	-Omega1	TM1	TCWV1	Prec1	BIC
1	x			x	x		x	9150.2
2	x		x	x	x		x	9152.4
3	x	x		x	x		x	9153.7
4	x			x	x	x	x	9154.3
5	x			x	x	x	x	9156.6

Se destaca que en todos los modelos se incorporaron las variables inestabilidad atmosférica, tanto con retardo igual a 1 como sin retardo, así como la humedad transportada desde la región fuente y la precipitación, ambas del día anterior. El mejor modelo, según este criterio, resultó tener un valor de BIC igual a 9150.2, incorporando las variables mencionadas anteriormente en su ajuste. Con estas nuevas variables se ajusta nuevamente el modelo de regresión lineal, obteniéndose un valor de R^2 de 0.4972 y un R^2 ajustado de 0.4965. Este resultado significa que cerca del 50% de la variabilidad de la precipitación, puede ser explicada por la inestabilidad atmosférica, del mismo día y del día anterior, así como por la precipitación y el transporte de humedad, ambos con retardo igual a 1. Además se destaca que, en este modelo, todas las variables incluidas son significativas para cualquier nivel de significancia.

Los valores estimados de las variables en el ajuste de este modelo lineal fueron los siguientes:

- El valor esperado de la precipitación cuando todas las variables predictoras son 0, es ~ 0.69 mm/día. Dicho valor es bajo, teniendo en cuenta que en este caso se considera que no hubo ningún contenido de humedad transportado desde el Golfo de México, que no precipitó el día anterior y que hay baja inestabilidad. Es decir, no existían condiciones atmosféricas en la región de interés para que precipitara, lo que se puede traducir a un ambiente de estabilidad.
- Una de las contribuciones mayores es cuando aumenta el valor de la inestabilidad atmosférica, ya que manteniendo constantes el resto de las variables predictoras, por cada unidad que aumenta la inestabilidad en esa región, se espera que aumente la precipitación aproximadamente 32 unidades. Esto demuestra la importancia de la existencia de inestabilidad en la atmósfera para que precipite.
- Un incremento en una unidad de la inestabilidad del día anterior, significa un incremento de aproximadamente 6 unidades en la precipitación. Este incremento de la precipitación es menor, lo que puede estar relacionado con que la influencia de la inestabilidad del día anterior no es de forma directa.
- Cuando hay un aumento de una unidad del transporte de humedad del día anterior y la precipitación del día anterior, el valor de la precipitación se espera que aumente, 0.17 y 0.5 unidades, respectivamente. Estos valores bajos de precipitación demuestran que estas variables influyen en menor medida en la variable respuesta y confirman la necesidad de presencia de movimientos verticales en la atmósfera para que se genere inestabilidad.

Uno de los aspectos importantes es comprobar si al incorporar las variables predictoras con lag 1 se corrige la autocorrelación presente en los residuos del modelo. El gráfico ACF para este modelo se muestra en la **Figura 3.2**, donde se puede apreciar la corrección de la autocorrelación en los primeros retardos. Esto demuestra que la incorporación de las variables predictoras con retardos adicionales solucionó el problema de la autocorrelación en los residuos del modelo, lo que posibilita contar con un modelo más adecuado que captura mayor cantidad de información de los datos analizados. A pesar de ello, se realizó la diagnosis del modelo para comprobar el resto de hipótesis, las cuales no se cumplieron satisfactoriamente.

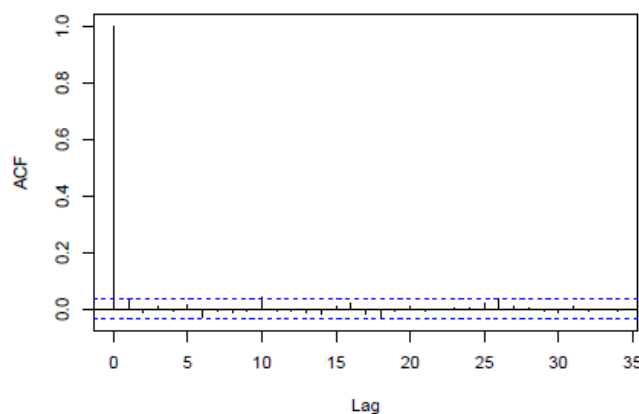


Figura 3.2: Autocorrelación de los residuos del modelo lineal de la precipitación en función de la inestabilidad atmosférica con retardo igual a 1 y sin retardo, la humedad transportada del día anterior y la precipitación del día anterior.

De esta manera se puede pasar a un análisis más profundo, para diferentes cuantiles con este modelo de regresión lineal. Para el cálculo de las nuevas predicciones se utilizó la muestra test mencionada anteriormente. Los cuantiles utilizados fueron: 0.25, 0.50, 0.75, 0.95 y 0.99, los cuales permitirán encontrar valores extremos que pueden estar presentes en los datos analizados. Estos cuantiles pueden proporcionar una herramienta para la planificación y la gestión de recursos hídricos, así como para evaluar los posibles riesgos que pueden existir, ya sea por problemas de sequías o por precipitaciones intensas.

Análisis individual de las variables predictoras

Para un análisis más específico y con el objetivo de ver cuánto puede influir la diferencia de las condiciones de las variables explicativas, se varió solo una variable predictora, desde el valor mínimo hasta el valor máximo, manteniendo fijas, con su valor medio, el resto de las variables explicativas. Esto permitiría además, representar el comportamiento de la precipitación en función de cada una de esas variables predictoras, así como los cuantiles utilizados utilizando la muestra test. Para el análisis se emplearon los cuantiles mencionados anteriormente. A continuación se mostrarán los resultados para cada una de las variables analizadas.

En la **Figura 3.3** se muestra la representación de la precipitación en función de cada una de las variables predictoras. De forma general el modelo de regresión utilizado en este caso es un modelo lineal, por lo que las líneas de los cuantiles, no se ajustan del todo correcto a los datos, ya que no es lo suficientemente flexible para captar la relación que existe entre las dos variables y por tanto la distribución de los puntos. En cuanto a la inestabilidad atmosférica (**Figura 3.3 a**) se observa como al aumentar los valores de la inestabilidad atmosférica en la región de estudio, la precipitación tiende a aumentar en cada uno de los cuantiles. Esto demuestra una vez más que, la existencia de inestabilidad en la atmósfera es una condición necesaria para la ocurrencia de precipitaciones.

Por otra parte, en la **Figura 3.3 b**, se observa que el aumento de la humedad transportada desde el Golfo de México, produce también un incremento de la precipitación para cada uno de los cuantiles, pero mucho más ligero en comparación con la inestabilidad atmosférica. Estos resultados demuestran que puede haber gran contenido de humedad en la atmósfera, transportada desde una región fuente, pero es necesario la existencia de condiciones inestables en la región de interés para que se generen movimientos verticales, el vapor de agua ascienda y se condense para que se genere la precipitación. La inestabilidad atmosférica del día anterior tiene menos influencia en el comportamiento de la precipitación (**Figura 3.3 c**), y es que tiene una relación menos directa, ya que puede estar influenciada por la presencia de algún sistema meteorológico que presenta evento de precipitación y que perdure en el tiempo. Por último, considerando la variable explicativa de la precipitación con lag 1 (**Figura 3.3 d**), se observa como el aumento de esta, produce un aumento también en la precipitación que se está analizando. Esta relación entre la precipitación de ambos días está condicionada también con la presencia de inestabilidad atmosférica el día anterior y de sistemas que generen varios días de precipitación.

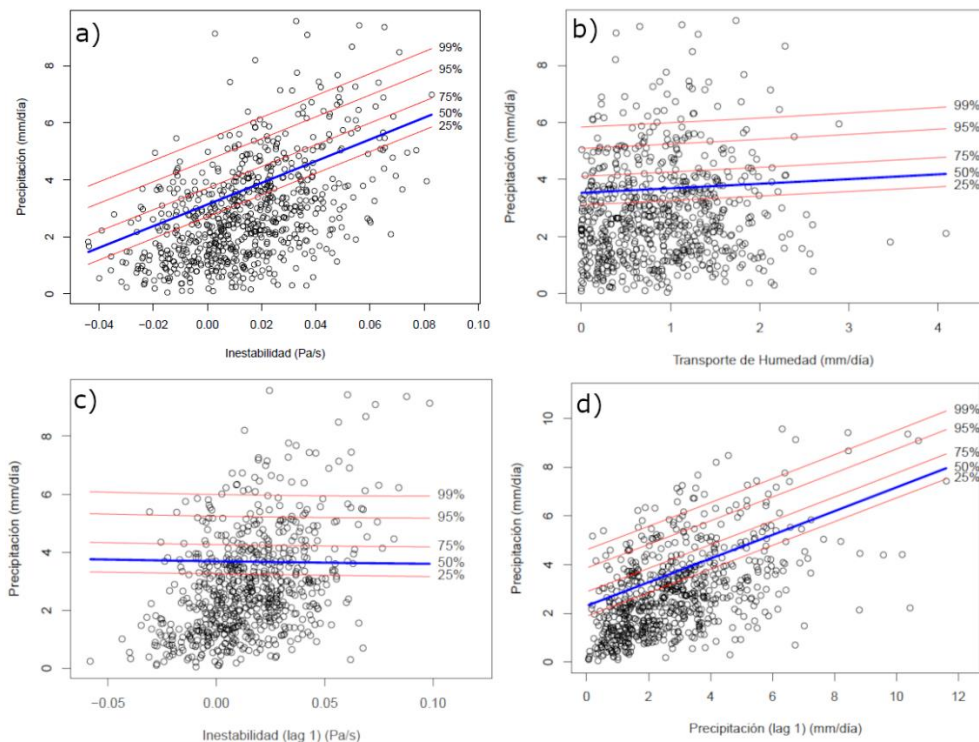


Figura 3.3: Cuantiles de la precipitación en función de las variables predictoras del ajuste del modelo lineal. a) Precipitación en función de la inestabilidad atmosférica, b) precipitación en función de la humedad transportada del día anterior, c) precipitación en función de la inestabilidad atmosférica del día anterior, d) precipitación en función de la precipitación del día anterior.

3.2 Modelos aditivos generalizados

Para realizar un análisis aún más complejo y lograr representar relaciones más flexibles entre las variables predictoras y la respuesta se utilizaron los modelos de regresión aditivos generalizados. Teniendo en cuenta que es otro modelo, con otras características, se aplicó el mismo procedimiento que para el modelo lineal. Es decir, primeramente se ajustó el modelo para las variables originales sin ningún retardo; la precipitación como variable respuesta en función de $-\Omega$, TM y TCWV. Para ello, se utilizó la función “*gam()*” del paquete “*mgcv*” (Wood, 2011, 2017; Wood et. al., 2016). Se obtuvo un valor de R^2 ajustado bajo, igual a 0.226, lo que significa que es capaz de explicar aproximadamente el 23% de la variabilidad de la precipitación. En cuanto a la diagnosis del modelo, tampoco se cumplieron todos los criterios de validación, y al igual que en el caso anterior del modelo lineal, el gráfico ACF también mostró la autocorrelación de los residuos en los primeros retardos (Figura 3.4).

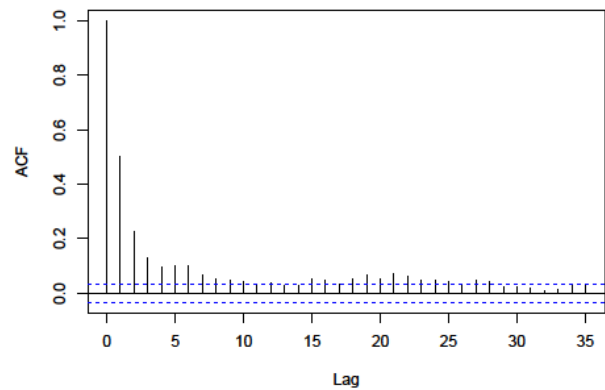


Figura 3.4: Autocorrelación de los residuos del modelo aditivo generalizado original de la precipitación en función de la inestabilidad atmosférica, la humedad transportada y el vapor de agua total en la columna.

Selección de variables

Se incluyeron nuevamente las mismas variables con retardo igual a 1 y se aplicó el criterio de selección BIC para incorporar sólo las más significativas. Los variables incorporadas en los 5 mejores modelos GAM, según este criterio, se muestran en la **Tabla 3.2**.

Tabla 3.2: Variables incorporadas de los 5 mejores modelos aditivos generalizados según el criterio de selección de variables BIC.

Orden	$-\Omega$	TM	TCWV	$-\Omega_{lag1}$	TM ₁	TCWV ₁	Prec ₁	BIC
1	x			x	x		x	9155.4
2	x			x			x	9157.8
3	x	x		x			x	9159.4
4	x						x	9160.8
5	x						x	9165.3

De forma general, estos modelos aditivos incorporan menos variables para realizar el ajuste en comparación con el modelo lineal. Sin embargo si fue similar en ambos modelos que se incluyeran en todos los casos las variables inestabilidad atmosférica y precipitación del día anterior. El mejor modelo, según este criterio, resultó ser el que presentó un valor de BIC igual a 9155.4, incluyendo las variables predictoras: $-\Omega$, $-\Omega_{lag1}$, TM₁ y Prec₁. La incorporación de estas variables coincidió también con el mejor modelo lineal, lo que demuestra la importancia de estas variables en el estudio de la precipitación.

Se ajustó un nuevo modelo GAM con las variables predictoras incorporadas anteriormente, obteniéndose un mejor ajuste que el modelo con las variables originales. En este caso se obtuvo un valor de R^2 ajustado igual a 0.513, lo que permite explicar casi el 52% de la variabilidad de la precipitación en función del resto de las variables predictoras. Si bien este valor no es muy alto, teniendo en cuenta la complejidad de trabajar con datos climáticos y meteorológicos, se considera un buen resultado. Por otra parte, todas las variables incluidas son estadísticamente significativas para cualquier nivel de significancia dado.

La diagnosis del modelo no arrojó tampoco los mejores resultados, sin embargo la incorporación de estas variables sí fue suficiente para corregir la autocorrelación en los residuos como se observa en la **Figura 3.5**.

En resumen, este modelo GAM, a pesar de no cumplir todas las condiciones necesarias para la validación del mismo, teniendo en cuenta la complejidad de los datos, se puede decir que, muestra un mejor ajuste con relaciones no lineales significativas entre las variables predictoras y la respuesta. La significatividad de todas las variables y el porcentaje de desviación explicada sugieren que el modelo puede ser adecuado brindando una representación de las relaciones que se establecen entre las variables analizadas. Sin embargo, el no poder validarse, es una cuestión a tener en cuenta a la hora de extraer conclusiones de los resultados. Después de haber realizado todo este análisis con el modelo GAM, se decide continuar con el estudio y se considera que puede ser utilizado en modelos de regresión más complejos.

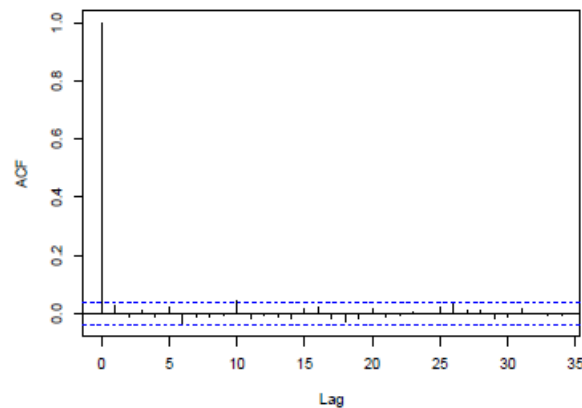


Figura 3.5: Autocorrelación de los residuos del modelo aditivo generalizado de la precipitación en función de la inestabilidad atmosférica con retardo igual a 1 y sin retardo, la humedad transportada del día anterior y la precipitación del día anterior.

Modelos de localización y escala

Para implementar un análisis más completo de la relación entre las variables predictoras y la precipitación se utilizaron los modelos de localización y escala, que incluyen la varianza como parte de la modelación. Ambos términos, el de localización y escala, en la modelación se representan como un modelo aditivo generalizado incluyendo las variables predictoras seleccionadas según el criterio BIC anterior. Además se utilizaron los cuantiles anteriores para un análisis más completo de la distribución de la precipitación, así como realizar un análisis individual de la influencia de cada una de las variables explicativas en la precipitación con la muestra test. En la **Figura 3.6** se muestran los ajustes para cada una de las variables. De forma general, el comportamiento de los cuantiles se ajusta mejor a la distribución de los datos, en comparación con el modelo lineal anterior. En primer lugar se aprecia, en la **Figura 3.6 a**, el aumento de la precipitación en función de la inestabilidad atmosférica, manteniendo el valor medio del resto de las variables predictoras. Además, se observa cómo varía la dispersión de la variable respuesta a medida que aumenta la inestabilidad. Para valores más pequeños de inestabilidad la diferencia entre los cuantiles es ligeramente menor, mientras que para los mayores valores estos están más separados. Si bien es cierto que estos valores no son elevados, es un aspecto a tener en cuenta a la hora de planificar y entender las condiciones bajo las cuales se podría esperar una precipitación significativa.

Los resultados relacionados con la humedad transportada del día anterior, también son similares a los obtenidos en el modelo lineal. En la **Figura 3.6 b**, se evidencia que el aumento de esta variable, con presencia de valores medios del resto, no influye considerablemente en el aumento de la precipitación para ese día. Esto demuestra que un gran contenido de humedad en la atmósfera no es suficiente por sí solo para que ocurra la precipitación, ya que se necesitan otras condiciones como la presencia de núcleos de condensación y de movimientos verticales ascendentes. En este sentido al contar con valores de inestabilidad medios, tanto para ese mismo día como para el día anterior, pueden no ser suficientes para que la humedad se condense y precipite. En cuanto a la inestabilidad atmosférica del día anterior (**Figura 3.6 c**), sólo para sus valores más altos, se considera un ligero aumento de la precipitación. En este caso la variabilidad de los cuantiles de la precipitación es mucho mayor para los valores más altos de inestabilidad. Por último, con el aumento de la precipitación del día anterior, la precipitación del mismo día analizada, tiende también a aumentar de forma más ligera (**Figura 3.6 d**). El mayor aumento se observa en los cuantiles más altos, manteniéndose más constante para los mayores valores de la precipitación con lag 1. Además, para los valores más bajos de precipitación con lag 1, hay menos variabilidad, ya que las líneas de los cuantiles de la precipitación están más próximas. Sin embargo, para los valores mayores ocurre lo contrario, la diferencia es mayor, con una mayor variabilidad.

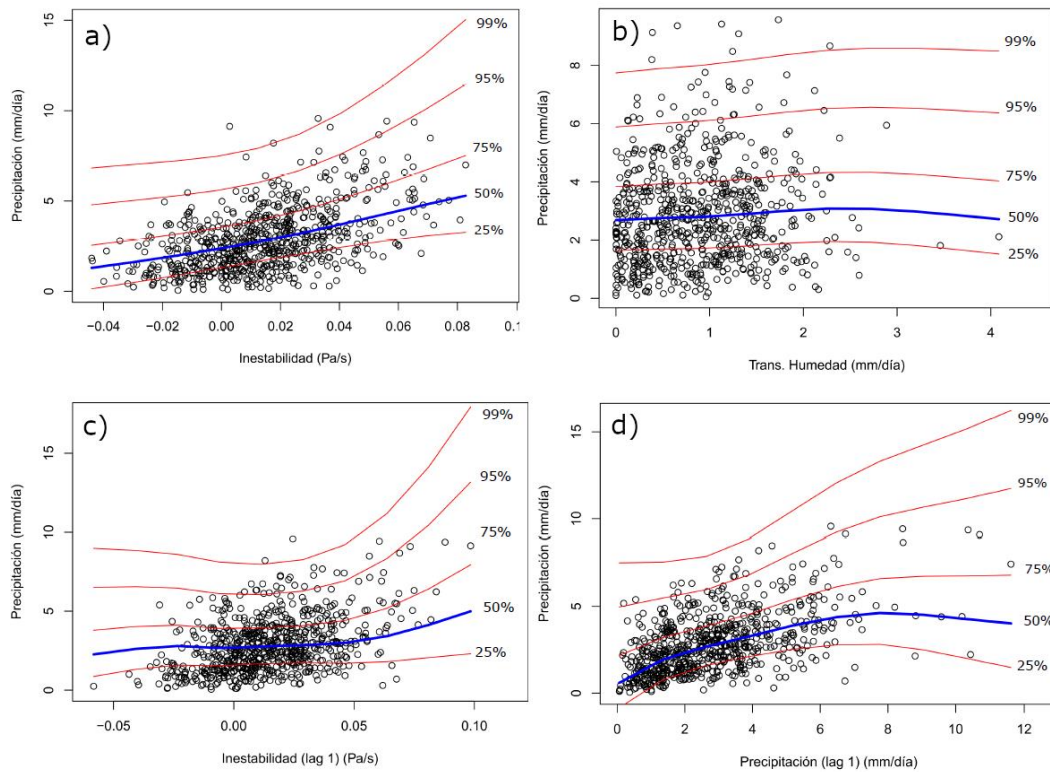


Figura 3.6: Cuantiles de la precipitación en función de las variables predictoras del ajuste del modelo de localización y escala. a) Precipitación en función de la inestabilidad atmosférica, b) precipitación en función de la humedad transportada del día anterior, c) precipitación en función de la inestabilidad atmosférica del día anterior, d) precipitación en función de la precipitación del día anterior.

Conclusiones

La precipitación asociada con la corriente de chorro de bajo nivel en la región de las Grandes Llanuras Americanas ha sido ampliamente estudiada debido a la importancia que esta tiene para la agricultura de la región. Algunos estudios centrados en la precipitación extrema a escala diaria han identificado la inestabilidad atmosférica como el factor principal que contribuye a los altos valores de precipitación, combinados con la presencia de un suministro continuo de grandes cantidades de humedad, como el proporcionado por el chorro de bajo nivel (**Gimeno-Sotelo et al., 2022**). Extender estos resultados a toda la distribución de la precipitación diaria es un problema altamente complejo, debido a los diversos factores atmosféricos que podrían intervenir. Sin embargo, los hallazgos presentados en este artículo, que emplean modelos de regresión que van desde simples hasta más complejos y a través de varios cuantiles de precipitación, parecen aportar importantes resultados al tema.

En este estudio, se emplearon modelos de regresión para analizar la relación entre la precipitación diaria asociada con el chorro de bajo nivel y la inestabilidad atmosférica, el vapor de agua total en la columna de aire y la humedad transportada desde las regiones fuente en el mar Caribe y el Golfo de México al sumidero en las Grandes Llanuras Americanas. Se ajustaron modelos lineales, aditivos generalizados y de localización y escala, utilizando además los cuantiles 0.25, 0.50, 0.75, 0.95 y 0.99. La incorporación de variables predictoras con retardos iguales a 1 corrigió la fuerte autocorrelación de los residuos de los modelos en los primeros lag, lo que demuestra la influencia de las condiciones meteorológicas en días anteriores sobre las condiciones de los días actuales. Las variables predictoras inestabilidad atmosférica con y sin lag 1, la humedad transportada con lag 1 y la precipitación con lag 1, fueron incorporadas en los ajustes de los modelos, por el criterio de selección de variables BIC.

Con el aumento de la complejidad de los modelos de regresión utilizados durante el desarrollo del trabajo, se fueron obteniendo mejores resultados y ajustes en cuanto a la distribución de la precipitación, lo que demuestra la necesidad de modelos más robustos para el estudio de la relación entre las variables analizadas. El modelo de localización y escala fue el que mejores resultados arrojó en cuanto a la representación de la precipitación en función del resto de las variables predictoras. De forma general, se destacó la importancia de la presencia de inestabilidad atmosférica en la región de interés para la ocurrencia de precipitación, por encima del resto de las variables. Y además, que a pesar de contar con humedad en la región de interés, esta

no es suficiente por sí sola para la ocurrencia de precipitación. Los resultados son de gran importancia para comprender el comportamiento de la precipitación en la región de estudio y poder actuar con mayor eficacia ante la ocurrencia de posibles riesgos, ya sea de precipitaciones extremas o sequías.

En trabajos futuros se pretende ampliar el estudio de la precipitación en otras regiones de interés, incorporando de ser necesario nuevas variables predictoras con el objetivo de obtener un mejor ajuste de los modelos. Además, aplicar otras técnicas estadísticas que cuenten con una mayor capacidad predictiva y que representen con más robustez la posible relación que existe entre las variables predictoras y la respuesta.

Referencias bibliográficas

- Algarra I, Eiras-Barca J, Nieto R, Gimeno L (2019) Global climatology of nocturnal low-level jets and associated moisture sources and sinks. *Atmos Res* 229:39-59. <https://doi.org/10.1016/j.atmosres.2019.06.016>.
- Gimeno L, Dominguez F, Nieto R, Trigo R, Drumond A, Reason CJ, et al (2016) Major mechanisms of atmospheric moisture transport and their role in extreme precipitation events. *ANNU REV ENV RESOUR*, 41(1), 117-141. <https://doi.org/10.1146/annurev-environ-110615-085558>.
- Gimeno-Sotelo L, de Zea Bermudez P, Algarra I, Gimeno L (2022) Modelling hydrometeorological extremes associated to the moisture transport driven by the Great Plains low-level jet. *Stoch Environ Res Risk Assess*, 36(7), 1917-1941. <https://doi.org/10.1007/s00477-022-02199-x>.
- Gimeno-Sotelo L, Bevacqua E, Gimeno L (2023). Combinations of drivers that most favor the occurrence of daily precipitation extremes. *Atmos Res* 294:106959. <https://doi.org/10.1016/j.atmosres.2023.106959>.
- Gimeno-Sotelo, L., Bevacqua, E., Fernández-Alvarez, J. C., Barriopedro, D., Zscheischler, J., & Gimeno, L. (2024a). Projected changes in extreme daily precipitation linked to changes in precipitable water and vertical velocity in CMIP6 models. *Atmospheric Research*, 304, 107413.
- Hastie T, Tibshirani R (1993) Varying-coefficient models. *J R Stat Soc, Series B* 55(4):757-796. <https://doi.org/10.1111/j.2517-6161.1993.tb01939.x>.
- Holton JR, Hakim GJ (2013) An introduction to dynamic meteorology (Vol. 88). Academic press.
- Neath, A. A., y Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2), 199-203.
- Nester J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) Applied linear statistical models. McGraw-Hill/Irwin. ISBN-13 978-0-07-310874-2.
- R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Ranhao S, Baiping Z, Jing T (2008) A multivariate regression model for predicting precipitation in the Daqing Mountains. *MT RES DEV* 28(3):318-325. <https://doi.org/10.1659/mrd.0944>.
- Rodó X, Baert E, Comín FA (1997) Variations in seasonal rainfall in southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation. *Clim. Dyn.* 13(4):275-284. <https://doi.org/10.1007/s003820050165>.
- Silva I, Roca Pardiñas J, Ordóñez C (2016) Forecasting SO2 pollution incidents by means of quantile curves based on additive models. *Environmetrics* 27(3):147-157. <https://doi.org/10.1002/env.2384>.
- Viechtbauer W, López-López JA (2022) Location-scale models for meta-analysis. *Res Synth Methods*, 13(6):697-715. <https://doi.org/10.1002/jrsm.1562>.
- Wilks DS (2011) Statistical Methods in the Atmospheric Sciences (3rd ed.). Academic Press.
- Wood SN (2006) Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC.
- Wood SN (2017) Generalized additive models: an introduction with R. Second Edition. Boca Raton: CRC Press. <https://doi.org/10.1201/9781315370279>.
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J Royal Stat Soc (B)* 73(1):3-36. <https://doi.org/10.1111/j.1467-9868.2010.00749.x>.
- Wood SN, Pya N, Saefken B (2016) Smoothing parameter and model selection for general smooth models (with discussion). *J. Am. Stat. Assoc.* 111:1548-1575. <https://doi.org/10.1080/01621459.2016.1180986>.