



Procesamiento de datos meteorológicos del IDEAM con Python y validación de datos del reanálisis ERA5.

Universidad de Antioquia.
Facultad de Ingeniería, Escuela Ambiental.

Daniel Felipe Vásquez Franco. Pregrado en Ingeniería Ambiental.
Asesor: John Alejandro Martínez A., PhD.

Resumen

Mediante el uso de Python y algunas de sus útiles librerías de funciones enfocadas al análisis de datos y al manejo de series de tiempo, se realizó un preprocesamiento y un procesamiento de datos meteorológicos correspondientes a estaciones del IDEAM. A través de esta herramienta computacional, se realizaron análisis que permitieron observar el comportamiento de las variables climatológicas de interés (temperatura, humedad relativa, precipitación y velocidad del viento) en diferentes escalas de tiempo mediante el cálculo de ciclos medios anuales, anomalías, promedios mensuales, entre otros.

En este trabajo se caracteriza la estacionalidad en diferentes localidades de Colombia y la variabilidad climática asociada a fenómenos de gran escala como las fases La Niña 2010-12 y El Niño 2015-16 del ENSO. Se pudieron apreciar grandes cantidades de datos faltantes para la velocidad del viento en todas las estaciones de estudio, además de datos "contaminados" entre los datos de la estación de Cartagena, siendo ambos casos contratiempos negativos que pueden modificar notoriamente la calidad e incertidumbre de los resultados.

Otro de los propósitos principales para este trabajo era el de validar la información climatológica estimada por el reanálisis ERA5 mediante una comparativa con los datos provenientes de estaciones del IDEAM. En general se encontraron sub y sobreestimaciones por parte de ERA5 con respecto a los valores observados, encontrando incluso sobrestimaciones de más del 150% respecto a las observaciones, como ocurrió para algunos meses del ciclo medio anual de velocidad del viento en la estación de Riohacha.

Más allá de esto, ERA5 se mostró como una fuente de información útil para representar la climatología de algunas regiones del país al menos siguiendo adecuadamente el comportamiento de las variables medidas por el IDEAM, al menos en términos de ciclo anual y variabilidad interanual.

Quedan pendientes como tareas para estudios posteriores realizar una revisión rigurosa de la calidad de los datos observados en las estaciones y también encontrar métodos de corrección de sesgo (Bias Correction) que permitan utilizar a ERA5 de manera más amplia para estudios locales. También queda pendiente extender el análisis realizado a otras estaciones meteorológicas del IDEAM.

Agradecimientos

Es para mí fundamental agradecer a mi asesor de trabajo, el docente John Alejandro Martínez, por su continuo apoyo y dedicación, además de su inagotable paciencia, no solo durante el desarrollo de este trabajo sino también en demás momentos de mi aprendizaje.

Agradezco también a la Universidad de Antioquia por forjarme como un profesional con pensamiento crítico y social, y a mis seres queridos por el apoyo brindado constantemente durante mi formación en todas las formas que les fuera posible hacerlo.



Tabla de contenido

Capítulo 1. Introducción, Objetivos y Marco Teórico.	4
1.1. Introducción.	4
1.2. Objetivos.	5
1.3. Marco teórico.	5
Capítulo 2. Datos y metodología.	7
2.1. Estaciones de interés.	7
2.2. Manipulación de datos climatológicos con Python.	8
2.3. Procesamiento de datos del IDEAM.	9
2.4. Selección, descarga y procesamiento de datos de ERA5.	11
2.5. Comparación entre datos del IDEAM y datos de ERA5.	13
Capítulo 3. Resultados y análisis.	14
3.1 Series de tiempo IDEAM.	14
3.1.1. Datos originales y ciclo medio anual con valores diarios.	14
3.1.2. Datos faltantes en las estaciones.	24
3.1.3. Porcentaje total de datos perdidos.	32
3.2. Comparación IDEAM y ERA5.	33
3.2.1. Temperatura.	34
3.2.2. Humedad relativa.	40
3.2.3. Precipitación.	46
3.2.4. Velocidad del viento.	52
Capítulo 4. Conclusiones.	58
Referencias bibliográficas.	60
Anexo 1. Reporte sobre datos meteorológicos ITM.	61

Capítulo 1. Introducción, Objetivos y Marco Teórico.

1.1. Introducción

El sistema climático es una estructura compleja e interactiva constituida por la atmósfera, la superficie terrestre, la nieve y el hielo, los océanos y otros cuerpos de agua, y los organismos vivos. El clima de la tierra es determinado por la interacción entre estos componentes y la influencia directa de un agente externo (radiación solar incidente). Anteriormente, el estudio del clima otorgaba solo información y técnicas básicas para delinear la climatología local, de mesoescala y global. Si bien estos productos son fundamentales, también son poco refinados y por tanto no permitían realizar un análisis a profundidad que se acoplara con otros datos físicos, sociales y económicos. Está bajo los intereses de cada nación aplicar prácticas consistentes en las observaciones climatológicas, realizar un buen manejo de datos históricos y mantener la calidad y utilidad de los servicios provistos (WMO, 2011).

El uso inicial de los datos climatológicos suele ser solo la primera de muchas aplicaciones posibles. Muchos de los usos de esta información no se pronostican durante la planeación y el proceso de adquisición de los datos, por lo que es común que nuevas aplicaciones de los mismos emerjan tiempo después de su obtención. La necesidad de contar con información precisa en menos tiempo continúa incrementando a medida que los requerimientos de los usuarios se siguen diversificando y para mantenerse al día con dicha necesidad es fundamental que la información climatológica (actual e histórica) sea manejada de manera comprensiva y sistemática.

Python es un lenguaje moderno, orientado a objetos y de código abierto que es usado en variados tipos de ingeniería de software. Aunque ha estado disponible por más de dos décadas, el uso de este lenguaje de programación en las ciencias atmosféricas comenzó a incrementar hace unos pocos años, cuando el desarrollo comunitario convergió hacia el manejo estándar de paquetes científicos (manejo de conjuntos de datos) necesario para el trabajo en las ciencias atmosféricas (Lin, J.W-B., 2012).

Este trabajo tiene como propósito realizar un ejercicio de preprocesamiento y procesamiento de variables climatológicas, en el lenguaje de programación Python, a partir de los datos meteorológicos suministrados por la red de estaciones del Instituto de Hidrología, Meteorología y Estudios Ambientales - IDEAM- que deje insumos útiles, al menos en la parte computacional, para estudios futuros. Este interés se extiende a entender la variabilidad climática de diferentes regiones de Colombia y realizar una validación de otra fuente de información climatológica como lo es el reanálisis ERA5.

En este estudio se consideran posibles limitaciones debido a datos faltantes en las estaciones, pero no se realiza un tratamiento de los mismos ni se evalúa rigurosamente la calidad de los datos disponibles.

1.2. Objetivos

Objetivo general:

Procesar y preparar datos meteorológicos del IDEAM para validar datos del reanálisis ERA5.

Objetivos específicos:

1. Realizar un preprocesamiento de datos del IDEAM (Temperatura, Humedad relativa, Precipitación y Velocidad del viento, a escala diaria y mensual, según disponibilidad en las estaciones) para llevarlos a un formato más fácil de manipular y adecuado para su posterior análisis.
2. Consolidar la información obtenida en el nuevo formato mediante gráficos y tablas que muestren la climatología de las variables de interés, incluyendo sus ciclos anuales y su variabilidad.
3. Obtener datos del reanálisis ERA5 que sean representativos de las estaciones, períodos y variables elegidas de la base de datos del IDEAM.
4. Validar los datos de ERA5 mediante una comparación con datos del IDEAM para un conjunto de estaciones seleccionadas.

1.3. Marco Teórico.

El clima, en un sentido estricto, se puede definir como las condiciones climáticas promedio para un lugar y período de tiempo específico. El clima se puede describir en términos de estadísticos descriptivos de las tendencias centrales y la variabilidad de elementos relevantes como temperatura, precipitación, presión atmosférica, humedad y vientos, o a través de la combinación de elementos como fenómenos y tipos de clima que sean típicos de una región, o del planeta como un todo, para cualquier período de tiempo. La climatología es el estudio del clima, sus variaciones y extremos, y sus influencias en distintas actividades que incluyen (pero no se limitan a) la salud humana, la seguridad y el bienestar. Con el tiempo, esta se ha convertido en una rama dinámica de la ciencia con un amplio rango de funciones y aplicaciones. Nuevas técnicas están siendo desarrolladas y distintas investigaciones están siendo acometidas para estudiar las aplicaciones del clima en diversos campos como la agricultura, los ecosistemas, la energía, la industria, entre otros. La meta básica del manejo de información climatológica es preservar, capturar y proveer acceso a datos climáticos y resultados de procesamiento para el uso de planificadores, tomadores de decisiones e investigadores (WMO, 2011).

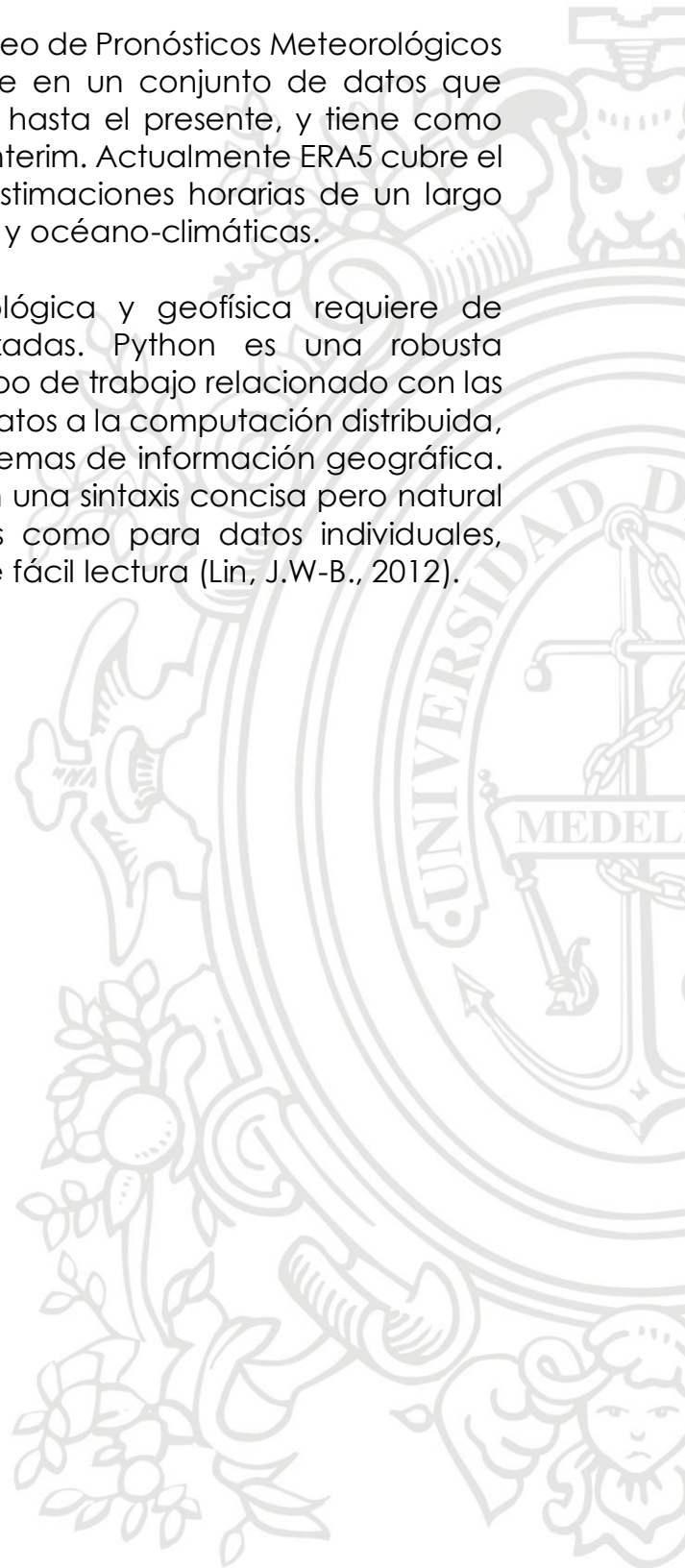
La consistencia y transparencia en la información climatológica y los métodos utilizados facilitan la comparación entre regiones o entre los modelos computacionales. El origen de todos los conjuntos de datos climatológicos son las redes in situ de observaciones meteorológicas, que son mantenidas por agencias meteorológicas alrededor del mundo (Ruane, A.C., 2015).

La cantidad y calidad de las estaciones puede variar ampliamente en el tiempo y en el espacio, por lo que resulta pertinente recurrir a otras fuentes de información climatológica como los reanálisis. Los reanálisis otorgan capturas exhaustivas de las condiciones atmosféricas en intervalos regulares sobre

largos períodos de tiempo (a menudo años o décadas). Son producidos a través de la asimilación de datos, un proceso que depende de las observaciones y de los pronósticos resultantes de modelos para estimar las condiciones atmosféricas (Parker, 2016).

El reanálisis ERA5, producto del Centro Europeo de Pronósticos Meteorológicos (ECMWF – por sus siglas en inglés), consiste en un conjunto de datos que cubrirán el período de tiempo desde 1950 hasta el presente, y tiene como propósito reemplazar a su predecesor ERA-Interim. Actualmente ERA5 cubre el período 2008-2018. Este reanálisis provee estimaciones horarias de un largo número de variables atmosféricas, terrestres y océano-climáticas.

La manipulación de información climatológica y geofísica requiere de herramientas computacionales especializadas. Python es una robusta plataforma de integración para cualquier tipo de trabajo relacionado con las ciencias atmosféricas, desde el análisis de datos a la computación distribuida, y desde interfaces gráficas de usuario a sistemas de información geográfica. Entre sus características, Python cuenta con una sintaxis concisa pero natural tanto para arreglos y conjuntos de datos como para datos individuales, haciendo el software sumamente claro y de fácil lectura (Lin, J.W-B., 2012).



Capítulo 2. Datos y Metodología.

2.1. Estaciones de interés.

Para la realización del estudio se escogió un conjunto de seis estaciones pertenecientes a la base de datos del IDEAM y correspondientes a diferentes regiones de Colombia, con el fin de contrastar la variabilidad climática que existe entre los distintos lugares del país. (Figura 1)

Mediante el servicio de solicitud de información prestado por el IDEAM para pedir datos meteorológicos, se obtuvieron archivos en formato de texto “.txt” que cuentan con una gran cantidad de valores diarios en diferentes años, para diferentes variables y en diferentes estaciones ubicadas en el territorio nacional. Para este estudio, se determinó que las variables de interés a extraer de dichos archivos serían la temperatura, la humedad relativa, la precipitación y la velocidad de los vientos. Siendo los datos de precipitación valores totales diarios y los de las demás variables valores promedios diarios. En cuanto al período de tiempo, se planteó extraer los datos para las estaciones elegidas que abarcaran el período de tiempo entre 1998 y 2017 [01-01-1998 a 31-12-2017] (20 años). Estos 20 años de información permiten analizar las condiciones climatológicas más probables (estacionalidad) y la variabilidad interanual (anomalías) de las estaciones de interés con una buena capacidad predictiva.



Figura 1. Estaciones del IDEAM elegidas para realizar el estudio.

2.2. Manipulación de datos climatológicos con Python.

Luego de elegir las estaciones de interés y el período de tiempo de estudio, fue necesario realizar un preprocesamiento de la información. Los formatos ".txt" enviados por el IDEAM son de fácil lectura para los humanos, pero presentan inconvenientes a la hora de ser leídos y manipulados a través de scripts, por lo que uno de los grandes propósitos de este trabajo era conseguir extraer la información de dichos formatos poco adecuados y llevarlos a formatos más fáciles de manipular a la hora de hacer análisis mediante lenguajes de programación (por ejemplo, archivos separados por comas o ".csv"). La figura 2 muestra un extracto de uno de los archivos originales del IDEAM.

I D E A M - INSTITUTO DE HIDROLOGIA, METEOROLOGIA Y ESTUDIOS AMBIENTALES												SISTEMA DE INFORMACION NACIONAL AMBIENTAL	
VALORES MEDIOS DIARIOS DE TEMPERATURA (oC)													
FECHA DE PROCESO : 2018/05/17						ANO 1998			ESTACION : 23085200 APTO J M CORDOVA				
LATITUD	0610 N	TIPO EST	SP	DEPTO	ANTIOQUIA	FECHA-INSTALACION	1973-OCT						
LONGITUD	7525 W	ENTIDAD	01 IDEAM	MUNICIPIO	RIONEGRO	FECHA-SUSPENSION							
ELEVACION	2073 m. s. n. m	REGIONAL	01 ANTIQUIA	CORRIENTE	RIONEGRO								
DIA	ENERO	FEBRE	MARZO	ABRIL	MAYO	JUNIO	JULIO	AGOST	SEPTI	OCTUB	NOVIE	DICIE	
01	17.4	19.1	18.0	18.9	17.2	16.3	18.4	17.3	16.5	18.6	17.5	15.0	
02	18.2	19.5	17.9	17.7	17.8	17.9	17.8	18.3	16.9	18.1	17.7	15.5	
03	17.4	19.6	19.1	18.3	15.8	17.5	18.0	17.2	16.8	17.1	17.6	15.9	
04	18.7	18.8	18.3	18.3	16.9	17.7	16.6	17.9	17.5	17.4	16.7	17.0	
05	17.2	19.3	18.1	18.9	16.3	18.6	15.5	17.9	17.4	18.5	16.9	16.1	
06	18.4	18.9	19.0	18.0	18.1	18.3	16.9	17.3	18.5	17.8	16.4	16.4	
07	17.7	17.6	19.4	18.6	17.5	18.1	17.2	16.0	18.2	17.1	15.7	15.5	
08	18.1	16.4	18.9	19.3	16.9	17.6	16.7	16.7	16.4	17.1	16.1	15.8	
09	17.8	16.7	17.9	19.1	18.5	18.7	17.4	18.1	17.5	17.9	17.4	15.9	
10	18.2	16.9	17.8	19.1	19.2	18.1	17.5	17.5	17.0	17.7	17.5	16.5	
11	18.6	17.5	18.1	17.0	19.1	19.0	17.1	18.0	16.8	17.9	17.7	16.3	
12	18.2	18.4	17.4	17.9	17.6	17.8	18.0	16.8	15.0	17.9	16.5	16.6	
13	18.0	18.5	18.3	16.9	19.3	17.1	17.6	17.1	15.3	16.5	16.8	16.8	
14	17.3	18.8	18.5	17.6	18.6	18.3	16.8	18.5	16.3	17.0	16.5	17.6	
15	18.4	18.5	18.5	18.6	17.7	16.8	16.9	15.8	17.5	16.8	16.3	16.5	
16	17.4	19.3	17.7	17.8	19.5	16.8	17.0	16.7	17.6	16.5	17.1	16.7	
17	17.5	17.6	19.5	19.6	18.6	17.2	17.8	16.9	16.8	16.6	17.0	16.8	
18	17.9	18.0	19.6	18.8	19.0	18.1	17.0	17.0	18.1	16.5	17.6	16.5	
19	18.9	18.8	19.9	17.5	19.4	17.5	17.2	16.4	18.2	16.9	17.1	16.9	
20	17.9	17.9	18.9	17.5	19.5	17.8	17.5	17.0	18.0	16.7	16.6	16.8	
21	19.4	18.7	18.3	18.9	20.4	18.6	17.7	17.7	17.8	16.0	15.5	16.0	
22	18.5	19.1	19.7	18.6	19.7	17.1	16.8	18.4	16.5	15.4	15.4	15.1	
23	19.3	17.9	17.6	18.5	18.6	16.7	17.4	17.8	15.4	15.9	15.7	15.1	
24	18.5	17.7	17.9	18.3	18.5	17.8	17.9	18.8	17.5	16.2	15.6	15.1	
25	18.6	19.3	18.7	18.5	19.0	18.8	16.8	15.8	17.6	15.9	15.8	14.3	
26	19.4	18.4	19.0	17.5	17.7	18.1	17.8	15.9	16.9	17.5	15.7	14.8	
27	18.0	19.6	17.9	18.2	16.5	17.9	17.5	17.0	16.6	16.7	16.0	16.0	
28	16.8	18.4	17.0	17.0	16.8	18.3	16.9	17.7	17.2	16.5	15.0	15.7	
29	19.5	17.6	18.3	17.4	17.5	17.3	17.3	17.6	17.0	17.0	16.3	15.1	
30	17.5	15.9	19.6	18.0	18.0	16.8	17.1	17.1	17.9	16.6	16.4	15.6	
31	19.1	17.4			17.4		17.4	15.2		17.5		17.3	
MEDIA	18.2	18.4	18.3	18.3	18.1	17.7	17.3	17.2	17.1	17.0	16.5	16.0	
MAXIMA ABSOLUTA	25.8	27.8	26.6	25.2	26.2	24.2	24.2	24.2	24.3	24.8	23.2	22.8	
MINIMA ABSOLUTA	8.2	10.0	10.8	13.0	10.6	7.0	10.0	9.2	11.0	8.8	10.2	9.3	
MEDIA DE MINIMOS	12.8	13.5	13.4	14.8	13.9	12.5	12.5	12.3	12.8	12.4	12.3	11.7	
MEDIA DE MAXIMOS	24.1	24.3	23.9	23.0	23.1	23.0	22.6	22.6	22.5	22.4	21.3	21.0	

Figura 2. Formato de archivo original del IDEAM para datos meteorológicos.

El preprocesamiento consistió en escribir un código o script en el lenguaje de programación Python, principalmente mediante a través del IDE (Integrated Development Environment) Spyder.

El código utiliza algunas librerías útiles de Python, entre ellas:

- Pandas: Una librería especializada en el manejo de series de tiempo.
- Calendar: que provee funciones relacionadas al manejo de fechas.
- NumPy: una poderosa librería enfocada al manejo de arreglos vectoriales o de matrices.

El funcionamiento del código consiste, de manera resumida, en los siguientes pasos:

- Realizar una búsqueda línea por línea del nombre de la variable de interés dentro de los archivos ".txt" del IDEAM.
- Cuando se encuentra la variable, el script procede a extraer información de interés previa a los datos como el código de la estación y el año de la serie.
- A partir de esto, se genera un DataFrame o Tabla de datos con algunas funciones de la librería pandas, donde el índice de las filas corresponde a las fechas diarias del año extraído y el nombre de cada columna indica el código de cada estación. Algunas funciones de la librería calendar se utilizan para tratar la longitud de los meses y la aparición de años bisiestos.
- Se extraen los datos diarios para la variable de la tabla que se encuentra en el formato del IDEAM en forma de serie de tiempo. Esto requiere conocer las posiciones exactas de los valores para poder capturarlos.
- A medida que se encuentran más años para la misma estación, estos se van añadiendo al DataFrame. Cuando se recopila todos los datos de todos los años disponibles para una estación, el código busca la variable en las líneas siguientes extrayendo así datos para demás estaciones.
- Cuando se llega a la última línea del archivo de entrada y no se encuentran más datos diarios para la variable buscada, el script escribe el valor estándar "-999" en los datos faltantes o en blanco y guarda el archivo de salida en un formato .csv (separado por tabulación). En este nuevo archivo se encuentran todos los datos diarios extraídos para la variable de interés donde la primera columna con las fechas sirve como índice del tiempo y cada una de las otras columnas contienen los datos de cada estación encontrada que tenía datos de la variable.

Luego de correr el código utilizando los archivos de entrada disponibles para extraer las variables de interés, se obtienen nuevos archivos para cada variable con una estructura legible para el lenguaje computacional.

Es importante mencionar que antes de realizar todo este proceso de escritura del código y extracción de los datos, ya se habían realizado previamente algunas tareas de manipulación de datos meteorológicos en Python, usando las librerías mencionadas arriba para explorar y resumir información de series de tiempo meteorológicas. (Ver Anexo 1).

2.3. Procesamiento de datos del IDEAM.

Al contar con los datos en el nuevo formato más manejable, fue posible realizar un procesamiento y análisis exploratorio de los datos para cada variable, que tenía como propósito entender el comportamiento de las variables climatológicas en las regiones representadas por las estaciones.

Como los nuevos archivos contienen todas las estaciones que contaban con datos en las variables buscadas, era necesario extraer de la tabla solo las

estaciones elegidas previamente. Para encontrarlas de manera fácil, se escribió otro script cuya función es crear directorios de las estaciones presentes en cada archivo del IDEAM. Este código funciona de manera muy similar al de la extracción de datos, pero obtiene solo la información espacial de cada estación además del nombre y el código respectivo. De los directorios de estaciones creados se obtuvo la información general para las seis estaciones del IDEAM elegidas para el estudio. (Tabla 1).

Tabla 1. Información general de las estaciones del IDEAM escogidas. *: El aeropuerto Antonio Nariño está ubicado en el municipio de Chachagüí, pero por convención se utiliza Pasto.

Nombre	Código	Longitud	Latitud	Municipio	Departamento
AEROPUERTO J.M. CORDOVA	23085200	-75.43	6.17	Rionegro	Antioquia
AEROPUERTO ANTONIO NARIÑO	52045020	-77.28	1.38	Pasto*	Nariño
AEROPUERTO EL CARAÑO	11045010	-76.63	5.68	Quibdó	Chocó
AEROPUERTO LOS GARZONES	13085040	-75.82	8.82	Montería	Córdoba
AEROPUERTO RAFAEL NUÑEZ	14015020	-75.52	10.45	Cartagena	Bolívar
AEROPUERTO ALM. PADILLA	15065010	-72.92	11.53	Riohacha	Guajira

Se elaboró otro script para el procesamiento de la información de modo que, con el código de cada estación de interés, se extrajeran de los archivos .csv únicamente las columnas del índice de fechas y las de cada estación de interés.

Con la información en un nuevo DataFrame, se realizó un “corte” en las fechas de la serie de tiempo ya que solo eran de interés las del período 1998-2017. Luego mediante el uso de funciones de NumPy se indica que los datos con valor “-999” corresponden a datos faltantes y sean tratados como tal durante los cálculos posteriores. Con la función “resample” de la librería pandas se obtuvieron los valores promedios mensuales de cada variable a partir de los datos diarios (valores mensuales totales en el caso de la lluvia).

Como también era importante conocer la cantidad de datos faltantes para cada variable en cada estación, se realizó otro script de funcionamiento similar al anterior que contabilizara y transformara a valores porcentuales la cantidad de datos perdidos por mes por estación para cada variable y que también cuantificara la cantidad total de datos perdidos por variable por estación de manera porcentual. Este script produjo figuras que muestran dichos porcentajes.

Utilizando otra de las librerías más conocidas y útiles de Python, matplotlib, se realizaron las diferentes figuras que permitieran realizar un análisis claro sobre la climatología de las diferentes estaciones; esto incluyó gráficas para los ciclos medios anuales con datos diarios, las series con valores originales (diarios), variabilidad interanual, tendencias, entre otros.

2.4. Selección, descarga y procesamiento de datos de ERA5.

Otro de los propósitos principales era obtener datos del reanálisis ERA5 que fueran representativos de las variables, el período de tiempo y las estaciones elegidas para realizar una validación comparativa frente a los datos observados por las estaciones del IDEAM.

Para la descarga de datos de ERA5 se utilizó la API (Application Program Interface) de la Climate Data Store (CDS) (<https://cds.climate.copernicus.eu>). Esto consistió en realizar un script en Python que descarga de manera automática datos horarios para la variable, el año o período de tiempo y el área que se indique dentro del script. Se delimitó entonces un área correspondiente al territorio colombiano (ver Figura 3), con una resolución de grilla de $0.25^{\circ} \times 0.25^{\circ}$ asegurándose de que la información contendría la ubicación espacial de todas las estaciones. Se descendieron datos horarios para el área delimitada para cada variable de interés (temperatura, humedad relativa, precipitación y vientos) en un período de tiempo de 1998 a 2017 (incluyendo algunos datos de 2018 para hacer correcciones posteriores). Los archivos descendidos de la CDS vienen en formato netCDF, ampliamente utilizado para el empaquetamiento de datos climatológicos.

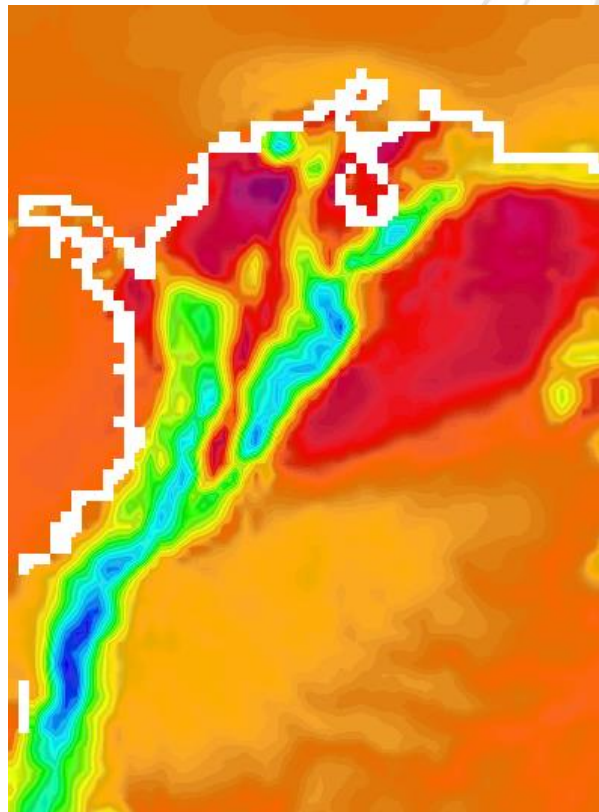


Figura 3. Visualización de campo de temperatura promedio mensual (enero 1998) en ncview. Los colores fríos indican temperaturas bajas y los colores cálidos temperaturas más altas.

Es importante mencionar que las variables humedad relativa y velocidad del viento no están disponibles para descarga directa, y se deben calcular a partir de otras:

- La humedad relativa se calcula luego de obtener la presión de vapor de saturación para la temperatura y para la temperatura de rocío, que se calcula mediante la ecuación:

$$e_{sat} = 0.6108 \exp\left(\frac{17.27 T^c}{237.3 + T^c}\right)$$

Donde T^c es la temperatura en grados Celsius.

Teniendo e_{sat} para ambas temperaturas, se calcula la humedad relativa como:

$$HR = \frac{e_{sat}(T_d)}{e_{sat}(T)} * 100$$

Donde T es la temperatura y T_d la temperatura de rocío.

- La magnitud de velocidad del viento se calcula mediante la ecuación:

$$Vel. viento = \sqrt{u^2 + v^2}$$

Donde u es la componente zonal del viento y v la componente meridional.

Se debe mencionar también que las componentes de viento de ERA5 tienen una altura nominal de 10 m y no de 2 m como las demás variables.

Para llevar las variables de interés a promedios mensuales (o valores mensuales totales en el caso de la lluvia) se utilizó la suite CDO (Climate Data Operators) que consiste en múltiples operadores de línea de comando para manipular y analizar datos climáticos, particularmente útil para archivos netCDF (<https://code.mpimet.mpg.de/projects/cdo/>).

Utilizando CDO se realizó una corrección para llevar las horas de los datos (ERA5 tiene los datos en UTC 00:00) a una hora local (UTC -5) y se realizaron demás cálculos necesarios para obtener valores mensuales.

Usando el software ncview se visualizó la información obtenida y se extrajeron las series de tiempo individuales para cada variable en coordenadas aproximadas a las de cada estación del IDEAM. De aquí se obtuvieron archivos .txt con valores mensuales para los datos descargados de ERA5 en las coordenadas especificadas.

2.5. Comparación entre datos del IDEAM y datos de ERA5.

Teniendo ambos conjuntos de datos en una escala mensual se procedió a realizar figuras en Python mediante la librería matplotlib que sirvieran para comparar la información de las estaciones del IDEAM con las series representativas de las estaciones descargadas de ERA5. Las figuras incluyeron comparación de los promedios mensuales, del ciclo medio anual, series de tiempo de anomalías para cada variable y scatterplots (análisis pareado) para cada variable de cada estación.

Para el cálculo del ciclo medio anual se agruparon todos los valores correspondientes a un mismo mes a lo largo de los 20 años (por ejemplo, todos los enero) mediante la función "groupby" de la librería pandas y se obtuvo su valor promedio. Al final se obtuvieron 12 valores promedio correspondientes a los 12 meses del año.

Luego de calcular el ciclo medio anual, se calcularon las anomalías mensuales. Este cálculo consistió en restar a cada mes particular de todos los años el valor promedio de ese mismo mes (por ejemplo, restar a enero de 1999 el enero promedio del ciclo medio anual).

Para el análisis pareado se cuantificó la correlación entre ambos conjuntos de datos en cada variable de cada estación y también la ecuación de la regresión lineal que representa el mejor ajuste. Esta parte estadística se logró haciendo uso de otra librería de Python muy útil llamada SciPy (cuenta con múltiples herramientas matemáticas, científicas e ingenieriles), usando de forma específica la función "linregress" del módulo "stats".

Capítulo 3. Resultados y análisis.

3.1. Series de tiempo IDEAM.

Luego de realizar el preprocesamiento de los datos meteorológicos provenientes de las estaciones del IDEAM se pasó al procesamiento de la información de interés (ahora en un formato más manejable), extrayendo así las series de tiempo originales aún en escala diaria para cada estación y variable de interés.

3.1.1 Datos originales y ciclo medio anual con valores diarios.

En las series originales diarias no es posible observar de manera clara el comportamiento de las variables climatológicas debido a la gran cantidad de datos en cada serie (Ver por ejemplo figura 2-a). Aun así, estas primeras gráficas sirven como punto de partida para apreciar los valores atípicos de cada variable, el rango de las fluctuaciones, y baches correspondientes a falta de información en las mediciones (datos perdidos o faltantes).

Graficar el ciclo medio anual con valores diarios es útil ya que permite observar la variación estacional de la variable de interés después de promediar sobre varios años (20 años en el presente caso).

Temperatura

Las series diarias de temperatura cuentan con una buena cantidad de datos en general, según se ve de la continuidad de las series de tiempo en la Figura 2. En la estación de Rionegro, los valores de temperatura alcanzan máximos mayores a 20°C y mínimos por debajo de los 14°C, además, no se aprecian datos faltantes a primera vista (Figura 4-a).

En la estación de Pasto se observaron valores máximos superiores a 24°C y mínimos inferiores a 16°C, y no se notan valores faltantes (Figura 4-c).

Los picos máximos de temperatura se aprecian en los años 1998, 2009, y 2015, y los picos mínimos se pueden ver en los años 2007-08, 2011-12 (años La Niña) y 2017.

Para la estación de Quibdó se observaron temperaturas máximas mayores a 31°C, ubicándose los picos máximos en la primera parte de los años 1998 y 2015, y temperaturas mínimas inferiores a 23°C, donde resalta el pico mínimo durante el año 2011. Tampoco se aprecian datos faltantes a partir de la figura (Figura 4-e).

En la serie correspondiente a la estación de Montería se notan inmediatamente baches en la información, en parte de los años 1999, 2016 y 2017 (Figura 4-g). Las mayores temperaturas superan los 30°C, ubicándose el pico máximo en el año 1998 (>32°C) y el siguiente en el año 2015. Los valores mínimos llegan a caer por debajo de los 25°C y sobresalen particularmente los mínimos de los años 2013 y 2015.

En la serie correspondiente a la estación de Cartagena las temperaturas máximas también superan los 30°C y los picos máximos se ubican en los años 2007, 2009 y 2014-15. Los valores mínimos pueden ser inferiores a los 25°C y los

picos mínimos se ubican en los años 2000-01, 2004 y 2007-08. Se aprecian datos faltantes para los años 2016-17 (Figura 4-i).

Finalmente, en la serie de datos diarios para la estación de Riohacha las temperaturas máximas superan los 32°C, ubicándose el pico máximo en el año 2015 y por encima de los 34°C. Las temperaturas mínimas llegan a ser menores a 25°C, resaltando los picos mínimos ubicados en 1998, 2000 y 2008 (<24°C). En cuanto a datos faltantes, se nota un bache en el año 2004 y otro en el año 2017 que se extiende hasta el final de la serie (Figura 4-k).

Continuando con el análisis del ciclo medio anual con valores diarios (columna derecha en Figura 2), para la estación de Rionegro muestra que las mayores temperaturas ocurren a mitad de año, viéndose los valores máximos en el mes de mayo (>17.75°C). La temporada con menores temperaturas se ubica a finales y comienzos del año, alcanzando los valores mínimos en el mes de noviembre (<16.25°C) (Figura 4-b).

Para la estación de Pasto se aprecia que los meses cálidos también ocurren a mitad de año alcanzando los valores máximos (>20.5°C) en el mes de agosto. La época más fría ocurre al final del año y los valores mínimos se observan en los meses de noviembre y diciembre (<18°C) (Figura 4-d).

En el ciclo anual con valores diarios para la estación de Quibdó se puede ver que las mayores temperaturas ocurren en la temporada Marzo-Abril-Mayo (MAM) llegando a valores superiores a 27.5°C, mientras que las temperaturas mínimas se encuentran en el tramo final del año (entre octubre y diciembre) con valores por debajo de los 26°C (Figura 4-f).

De forma similar, en la estación de Montería se aprecian las mayores temperaturas entre marzo y abril, con valores máximos superiores a 28.75°C (que también se alcanzan en el mes de junio) y las temperaturas mínimas correspondientes a la temporada fría se encuentran a final del año, resaltando los valores mínimos que se observan en octubre y noviembre, por debajo de 27.25°C (Figura 4-h).

Para la estación de Cartagena, el ciclo medio anual con valores diarios muestra que el año inicia con la época más fría, con valores mínimos entre enero y febrero por debajo de 27°C, mientras que la temporada cálida se ubica a mitad del año, observándose los máximos de temperatura entre junio y septiembre (>29°C) (Figura 4-j).

Finalmente, en la gráfica para la estación de Riohacha también se puede ver como la temporada cálida se ubica a mitad del año, entre junio y agosto, con valores máximos de temperatura superiores a los 30°C, mientras que la temporada fría ocurre a final y comienzo del año, alcanzando las temperaturas mínimas en Diciembre-Enero-Febrero, DEF, (<27°C) (Figura 4-l).

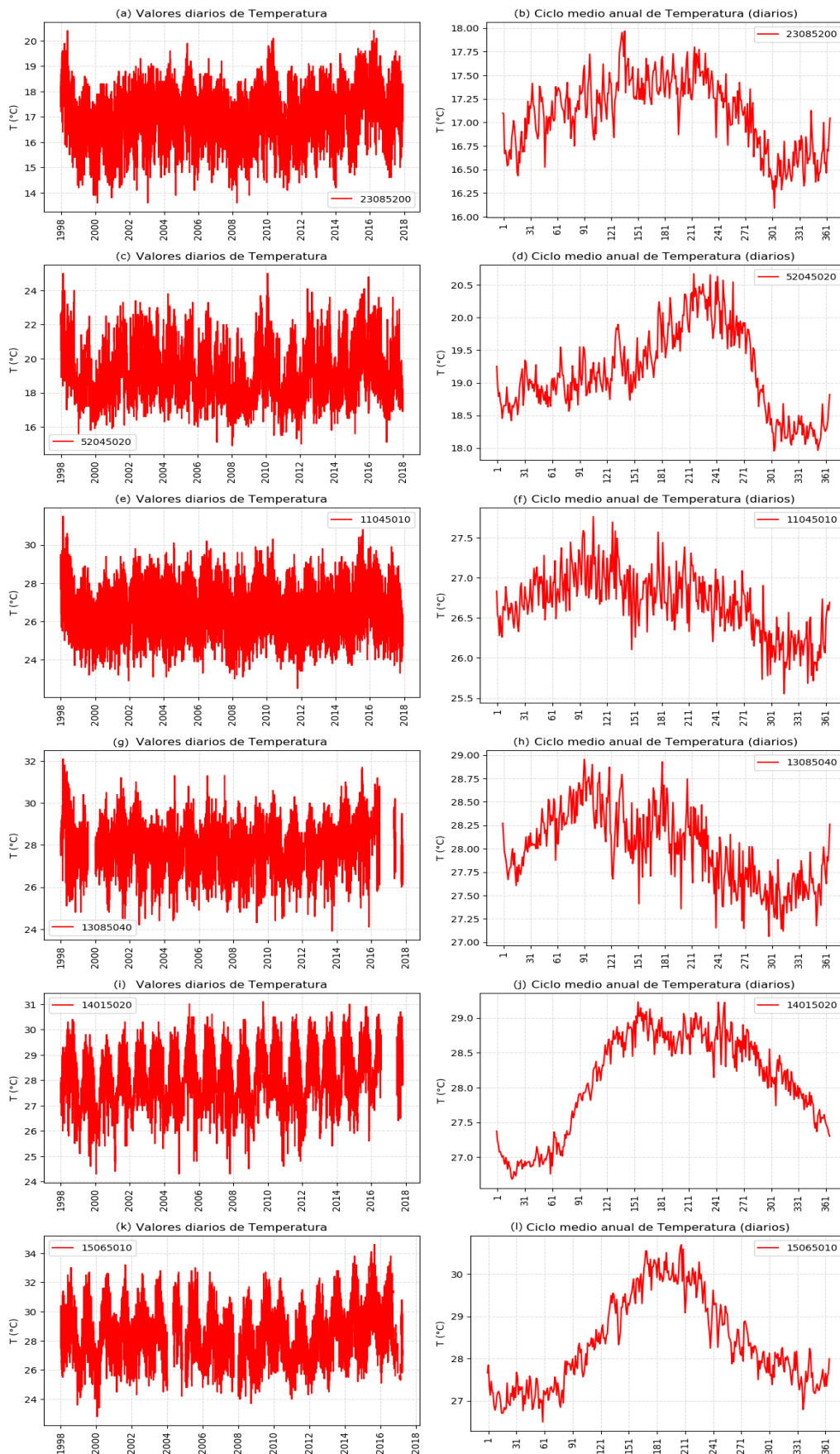


Figura 4. Valores diarios de temperatura (izq.) y ciclo medio anual de temperatura con valores diarios (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Rihacha [15065010].

Humedad relativa

Los valores diarios de humedad relativa para la estación de Rionegro cuentan con una amplitud de las oscilaciones relativamente grande, desde valores máximos superiores al 95% cuyos picos se ubican en los años 2005, 2007 y 2008, y valores mínimos por debajo del 65% donde los picos más notables se encuentran en los años 2011-12 y 2015-16. En cuanto a datos faltantes, no se logran apreciar baches en la información dentro de la gráfica (Figura 5-a).

En la estación de Pasto se observan amplitudes en la variación de la humedad relativa incluso mayores, oscilando entre valores máximos que casi rozan la saturación (100%), ubicándose los picos máximos en 1999, 2001, 2005 y 2011, y valores mínimos que caen hasta por debajo del 40%, siendo los picos mínimos más notables en 2010, 2012, 2014 y 2016. A primera vista, no se notan datos faltantes en la serie de esta estación (Figura 5-c).

La estación de Quibdó cuenta con una amplitud menor en las variaciones de humedad relativa, yendo desde máximos que se aproximan al 100% hasta mínimos que caen por debajo del 75%. Resalta entre todos los picos el mínimo que se presenta en el año 2011 con un valor por debajo del 70%, siendo esta época parte de una fuerte fase La Niña del fenómeno ENSO (El Niño-Southern Oscillation). En esta serie no se aprecian datos faltantes (Figura 5-e).

Para la estación de Montería se nota de manera inmediata los baches debido a datos faltantes ubicados en los años 1999, 2016 y 2017. En cuanto a valores máximos y mínimos, los picos máximos se pueden encontrar entre el período 2008-2015 (>95%) mientras que los mínimos que más resaltan se encuentran en los años 2001, 2004, 2007 y 2015, llegando hasta el 65% de humedad relativa (Figura 5-g).

La serie de valores diarios de la estación de Cartagena tiene valores faltantes entre los años 2016 y 2017. Los valores máximos de humedad relativa superan el 95% y algunos picos como los de 2005 y 2017 llegan hasta el punto de saturación (100%), mientras que los mínimos se pueden encontrar por debajo del 70%, llegando incluso a valores del 65% en 1998 y 2008 (Figura 5-i).

Finalmente, la gráfica para la estación de Riohacha muestra pequeños baches de datos faltantes en los años 2004 y 2007 además de una gran cantidad de datos perdidos en el año 2017. En esta serie también se encuentran valores máximos superiores al 95% cuyos picos incluso alcanzan el 100% de humedad relativa en el período 2007-08. La amplitud de la variación de la humedad es relativamente grande, puesto que se encuentran valores mínimos por debajo del 50%, resaltando algunos en los años 2002 y 2009 donde casi se alcanza el 40% de humedad (Figura 5-k).

El ciclo medio anual con valores diarios para la estación de Rionegro muestra de manera clara como los valores mínimos de humedad relativa se dan durante mitad de año llegando a valores inferiores al 74% en julio, mientras que los valores máximos ocurren en la temporada de finales de año, encontrándose los mayores porcentajes en noviembre (Mayores al 80%, ver Figura 5-b). Este comportamiento se puede asociar al papel que tiene la

temperatura en la capacidad del aire para contener vapor de agua y por tanto modificar el punto de saturación del mismo.

En la estación de Pasto se observa también como los valores mínimos durante el año promedio se alcanzan en la temporada cálida de mitad de año llegando a estar la humedad por debajo del 60% en julio, mientras que los valores máximos ocurren a finales de año en la época donde las temperaturas disminuyen y alcanzando valores superiores al 85% (Figura 5-d).

Para la estación de Quibdó la gráfica muestra como los mayores porcentajes de humedad relativa ocurren en los últimos meses del año (>89% en noviembre y diciembre) mientras los valores mínimos se ubican en el primer cuatrimestre (<84%), aunque también se observan algunos mínimos en el mes de julio. Para esta estación no se observa una estacionalidad tan marcada como en las dos anteriores, es decir, hay una menor amplitud entre los máximos y los mínimos (Figura 5-f).

En el ciclo medio anual con valores diarios correspondiente a la estación de Montería se puede apreciar de manera clara la estacionalidad de la humedad relativa, que alcanza sus valores mínimos en el primer trimestre del año (<75%), mientras que los valores máximos se ven entre los meses de agosto, septiembre y octubre, llegando a valores por encima del 84% de humedad relativa (Figura 5-h).

De manera similar pero con menor amplitud entre los valores se observa la gráfica para la estación de Cartagena, donde los mínimos que ocurren en el comienzo del año caen por debajo del 78% en febrero y lo máximos que ocurren en septiembre-octubre-noviembre llegan a superar el 84% de humedad (Figura 5-j).

Finalmente, la gráfica para la estación de Riohacha muestra como los valores mínimos ocurren a mitad de año llegando a estar por debajo del 65% de humedad relativa, mientras los valores máximos aparecen en el período septiembre-noviembre alcanzando valores por encima del 80% (Figura 5-l).

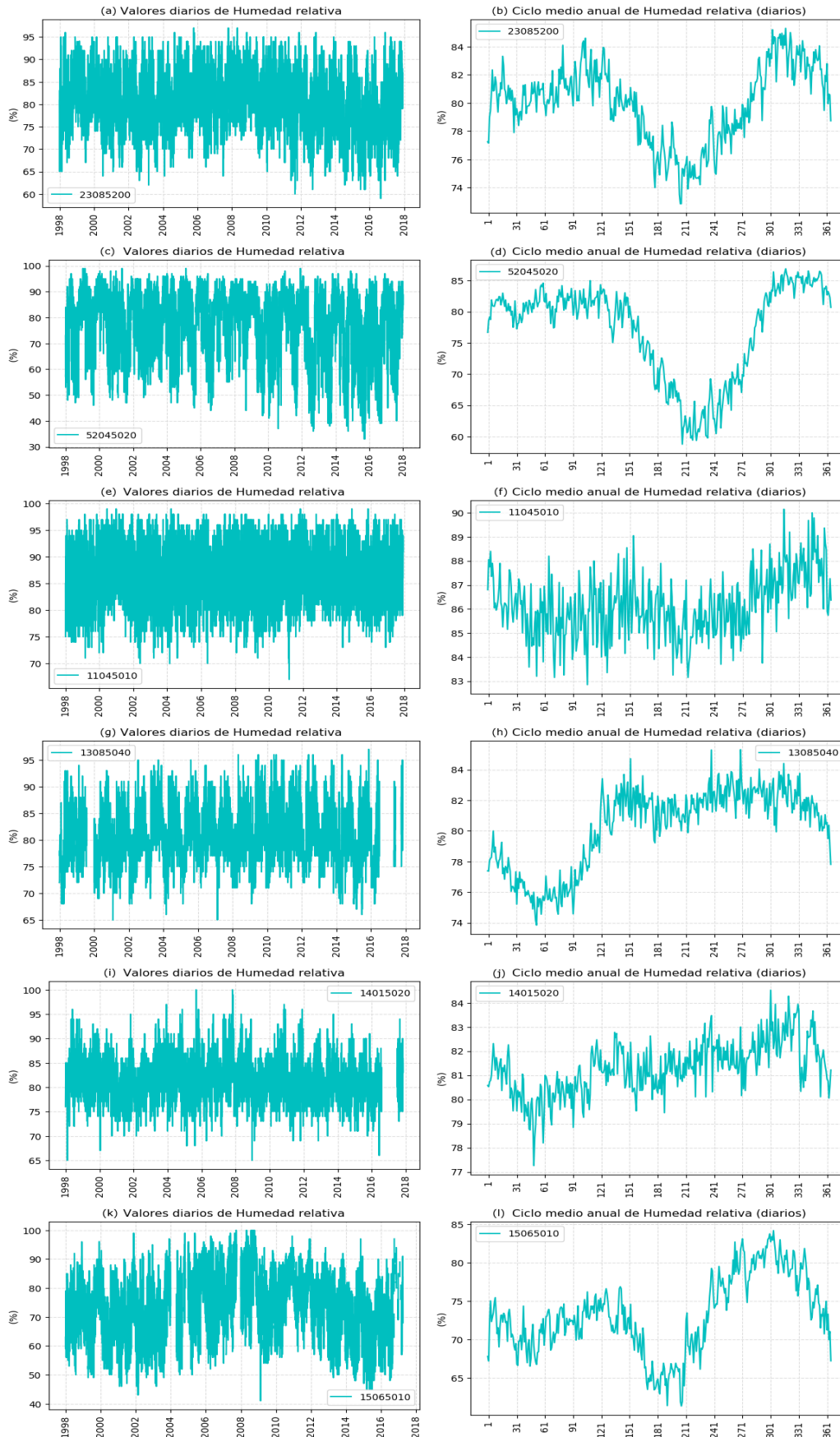


Figura 5. Valores diarios de humedad relativa (izq.) y ciclo medio anual de humedad relativa con valores diarios (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

Precipitación

La serie de datos diarios de precipitación para la estación de Rionegro muestra que gran parte de los valores de lluvia se encuentran por debajo de los 20 mm/día, aunque se observan algunos valores máximos atípicos que superan esta marca e incluso pueden sobrepasar los 80mm, como en los años 1999, 2006 y 2010-11. En cuanto a datos faltantes no se aprecian baches en ninguna parte de la serie (Figura 6-a).

En la estación de Pasto, la mayoría de los datos diarios también se encuentran por debajo de los 60 mm/día de lluvia, pero se pueden apreciar algunos valores máximos por encima de este número, en especial en los años 1999, 2011 y 2012, siendo el último año donde se ubica el pico máximo de precipitación (>80 mm/día). No se observan datos faltantes en la serie diaria (Figura 6-c).

Para la estación de Quibdó se nota a primera vista que los valores de lluvia diarios son mucho mayores en comparación con las anteriores, concentrándose la mayor cantidad de datos por debajo de los 150 mm/día. Los picos máximos más notorios superan incluso los 200 mm/día y se encuentran en los años 2001, 2007 y 2008. Esto muestra de manera cualitativa que valores de humedad relativa similares (como se ven al comparar la estación de Rionegro con la de Quibdó) no indican realmente la cantidad de precipitación que puede presentarse. En la serie no se aprecian baches de datos faltantes (Figura 6-e).

Gran parte de los valores diarios de lluvia en la estación de Montería se encuentran por debajo de los 60 mm/día, pero se pueden observar valores atípicos que alcanzan a superar los 80 mm/día de precipitación. En particular, resalta el pico máximo de lluvia en el año 2006 que alcanza a superar los 120 mm diarios, indicando un evento extremo muy atípico. Además, se aprecian datos faltantes para la serie en el período 2016-17 (Figura 6-g).

En la estación de Cartagena también se observan datos faltantes para los años 2016-17. La mayor parte de los datos se encuentran por debajo de la marca de 50 mm/día, pero se aprecian valores máximos que superan los 125 mm/día en los años 2003, 2004, 2007, 2010 y 2011 (Figura 6-i).

Finalmente, para la estación de Rihacha se observan pequeñas brechas correspondientes a datos faltantes en los años 2004 y 2007. Llamam la atención los valores máximos que superan los 100 mm/día de lluvia y en especial el pico máximo en el año 2011 que supera los 200 mm/día, posiblemente asociado a la fase La Niña del fenómeno ENSO ocurrido en dicho año (Figura 6-k).

En cuanto al ciclo medio anual con valores diarios, en la estación de Rionegro se observa de manera clara el régimen bimodal de lluvias de la región. Se presentan dos temporadas de lluvias durante el año, la primera en Marzo-Abril-Mayo (MAM) y la segunda en Septiembre-Octubre-Noviembre (SON), con la primera siendo más "fuerte" en cuanto a los valores máximos de lluvia (llegando a superar los 14 mm/día). Los valles correspondientes a las temporadas secas, que ocurren en los períodos Diciembre-Enero-Febrero (DEF)

y Junio-Julio-Agosto (JJA) muestran valores mínimos que pueden llegar hasta casi 0 mm/día (Figura 6-b).

Revisando las figuras previas de ciclo medio anual para temperatura (Figura 3, der) se puede ver entonces cómo estas dos variables interactúan entre ellas de manera inversa (al aumentar una disminuye la otra y viceversa) teniendo en cuenta otros factores asociados como la radiación solar entrante y la nubosidad presente durante eventos de lluvia.

Para la estación de Pasto se aprecia un régimen bimodal con épocas de lluvia en los períodos MAM y SON, siendo la primera donde se notan los picos máximos de precipitación que llegan casi hasta los 12 mm/día. La temporada seca más fuerte ocurre a mitad de año (JJA) con valores por debajo de 2 mm/día que alcanzan a llegar hasta los 0 mm por día (Figura 6-d).

La estación de Quibdó muestra un ciclo medio anual con datos diarios donde los valores máximos de lluvia se alcanzan a mitad de año (>40 mm/día) ubicándose el pico máximo en el mes de agosto (>50 mm/día). En general, los valores oscilan entre 10 y 30 mm/día de precipitación y no se observa una estacionalidad clara, coincidiendo con la climatología de la región que cuenta con un régimen de lluvias sin temporada seca definida (Figura 6-f).

En la estación de Montería se nota un régimen monomodal de precipitación, donde solo se ve una temporada seca de diciembre a marzo y una temporada de lluvias continúa desde abril hasta noviembre. Los valores máximos de la temporada de lluvias llegan a superar los 12 mm/día en mayo, junio, agosto y septiembre (Figura 6-h).

A primera vista, la estacionalidad para la estación de Cartagena puede asemejarse a un régimen monomodal, aunque realmente la región cuenta con un régimen bimodal con una temporada seca débil a mitad de año y otra mucho más marcada entre diciembre y abril. La temporada de lluvias más fuerte ocurre entre septiembre y noviembre, alcanzando valores máximos superiores a los 14 mm/día de precipitación (Figura 6-j).

Finalmente, para la estación de Riohacha se aprecia un régimen bimodal en su estacionalidad con dos épocas secas marcadas entre diciembre-marzo y junio-agosto. Las temporadas de lluvia ocurren de abril a mayo y de septiembre a noviembre, siendo la segunda la más lluviosa con valores superiores a los 10mm y que incluso llegan a superar los 15 mm/día en el mes de septiembre (Figura 6-l).

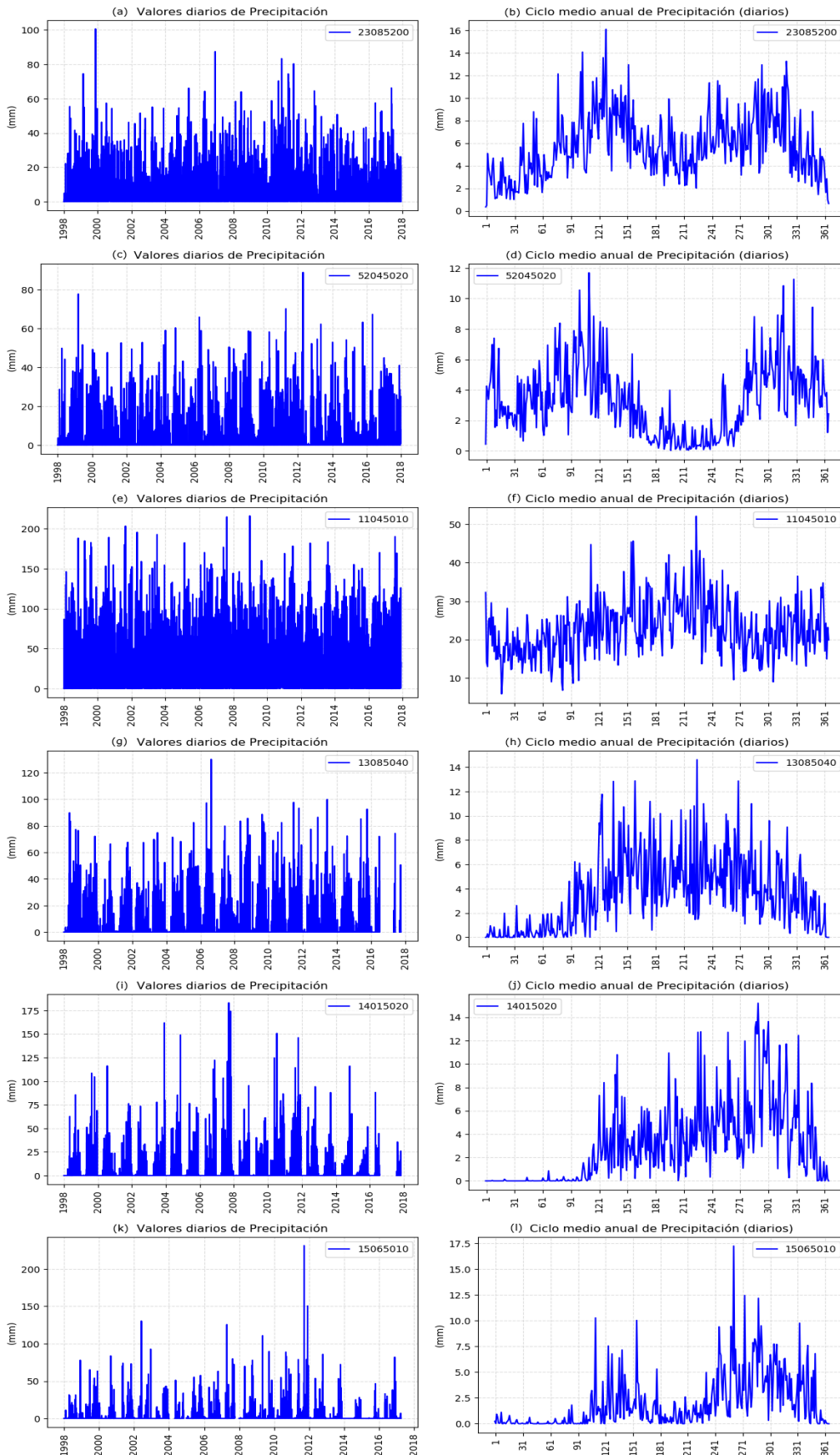


Figura 6. Valores diarios de precipitación (izq.) y ciclo medio anual de precipitación con valores diarios (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Rihacha [15065010].

Velocidad del viento

La serie de valores diarios de velocidad del viento para la estación de Rionegro (Figura 7-a) inicialmente aparenta no tener una gran cantidad de datos faltantes, pero en el eje temporal se puede observar cómo se dejan de tener datos en el año 2010, cuando el período de tiempo elegido abarca hasta 2017. Los valores máximos de velocidad se encuentran en los años 1998, 2002 y 2004 (>4 m/s) y los mínimos se encuentran entre 1.5 y 1 m/s, resaltando particularmente el valor mínimo del año 2010 donde la velocidad del viento cae completamente (0 m/s), aunque este número podría deberse a un error en la medición.

Para la estación de Pasto la serie muestra datos faltantes en varios baches; algunos en el período 1998-2000, otros entre 2003-2005 (el 2004 falta en su totalidad) y del 2014 hacia adelante. Los valores máximos superan los 6 m/s y los picos máximos más notables (>7 m/s) en los años 1998, 2001, 2002 y 2013, mientras que los valores mínimos van por debajo de 1 m/s y se observaron en los años 2000 y 2002 (Figura 7-c).

En la estación de Quibdó se nota inmediatamente una gran cantidad de datos faltantes entre 2004 y 2011, sin contar los que no aparecen en la gráfica (2012 en adelante). Los valores máximos ocurren entre 2001 y 2002 (>2.5 m/s) y los mínimos (<0.5 m/s) ocurren entre 1999 y 2000, alcanzando este último año magnitudes de velocidad nulas (Figura 7-e).

Los datos diarios para la estación de Montería muestran baches correspondientes a datos faltantes en los años 2002, 2003, 2004, 2005, 2006 al 2008 y 2016 al 2018. Los picos máximos de velocidad se ven en los años 2002, 2004 y 2011 con valores superiores a 3 m/s, mientras que los picos mínimos ocurren en el período 1998-2000 donde incluso aparecen valores de 0 m/s (Figura 7-g).

Gran parte de los datos diarios para la estación de Cartagena oscilan entre 0 y 5 m/s de velocidad del viento. Sin embargo, no es posible observar claramente las variaciones de los valores debido a la presencia de algunos valores extremos que superan los 30 m/s. Estos valores tan altos podrían ser espurios, producto de errores en la medición de la variable, y no debidos a eventos extremos reales. En cuanto a datos faltantes, se notan baches en los años 2011, 2012, 2014 y 2016 (Figura 7-i).

Finalmente, en la serie correspondiente a la estación de Riohacha se notan datos faltantes en los años 2004, 2007, 2008, 2012 y 2013 en adelante. Los valores máximos de velocidad del viento se encuentran en los años 1998 y 2001 (>6 m/s), mientras los valores mínimos se ven en los años 2006 y 2009 donde la velocidad cae por debajo de 1 m/s hasta llegar a valores nulos (Figura 7-k).

El ciclo medio anual con valores diarios para la estación de Rionegro muestra como los valores máximos se presentan a mitad del año (>2.8 m/s) mientras que la temporada de vientos más débiles ocurre a final de año con velocidades entre 1.8 y 2 m/s (Figura 7-b).

En la estación de Pasto se observa una estacionalidad con dos épocas de vientos débiles en los períodos de marzo a mayo y noviembre-diciembre, donde los vientos caen hasta velocidades menores a 2 m/s, mientras los valores máximos aparecen entre julio y septiembre con velocidades superiores a los 3.5 m/s (Figura 7-d).

Para la estación de Quibdó no se observa una estacionalidad clara con respecto a los vientos, pero se puede apreciar como los valores máximos ocurren entre agosto y septiembre (>1.8 m/s) mientras los mínimos ocurren en marzo y abril, siendo menores de 1 m/s y cayendo incluso por debajo de 0.8 m/s (Figura 7-f).

La estación de Montería muestra un ciclo medio anual de valores diarios donde los vientos son máximos entre febrero y abril (alcanzando velocidades superiores a 2 m/s) y las velocidades decrecen a lo largo del año hasta encontrar valores inferiores a 1 m/s entre octubre y noviembre (Figura 7-h).

La gráfica para la estación de Cartagena se encuentra completamente condicionada por los valores irreales mencionados anteriormente. El ciclo medio anual sugiere velocidades mucho mayores en los primeros meses del año, pero debido a la alteración causada por dichos valores atípicos no se pueden asumir estos valores como certeros (Figura 7-j).

Finalmente, en la estación de Riohacha se notan dos temporadas de vientos fuertes y dos de vientos débiles. La temporada de vientos más fuertes ocurre entre junio y agosto, con valores máximos por encima de 4 m/s, mientras que la temporada de vientos más débiles ocurre entre septiembre y noviembre con valores mínimos de velocidad del viento inferiores a los 2 m/s (Figuras 7-l).

3.1.2. Datos faltantes en las estaciones.

En la sección anterior se observó que en las series de datos de las estaciones pueden encontrarse datos faltantes en cualquier parte del período de tiempo. En algunas estaciones y para algunas variables es más o menos notoria la presencia de estos baches, o ausencia de datos. Con el fin de cuantificar detalladamente la cantidad de datos faltantes, se realizaron gráficas para cada estación y cada una de sus variables donde se pueden ver los porcentajes de datos perdidos mensualmente con respecto a los datos que debería de tener cada mes (por ejemplo, un mes de 30 días donde solo se registraron 15 datos diarios mostrará un porcentaje de datos faltantes del 50%).

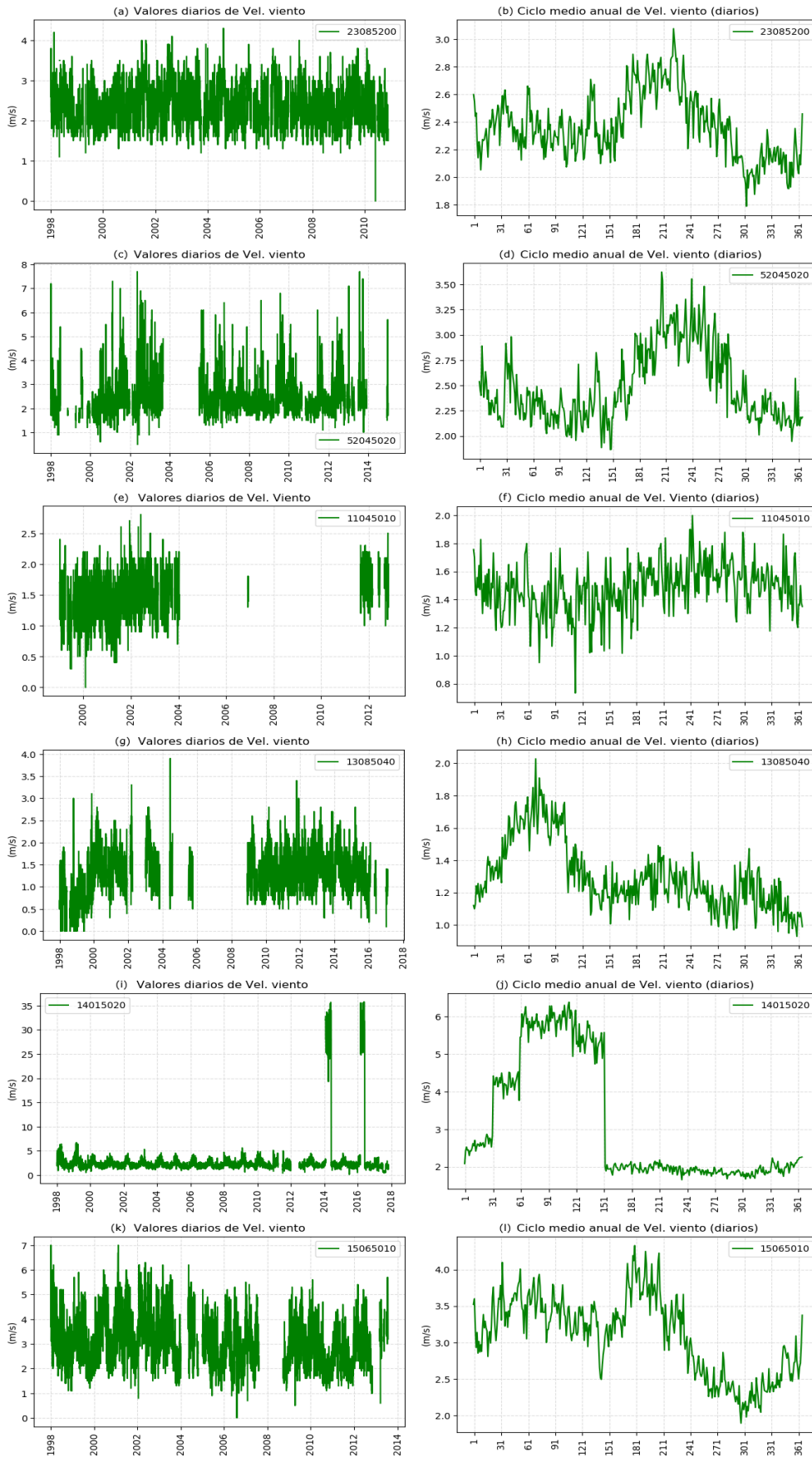


Figura 7. Valores diarios de Vel. viento (izq.) y ciclo medio anual de Vel. viento con valores diarios (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

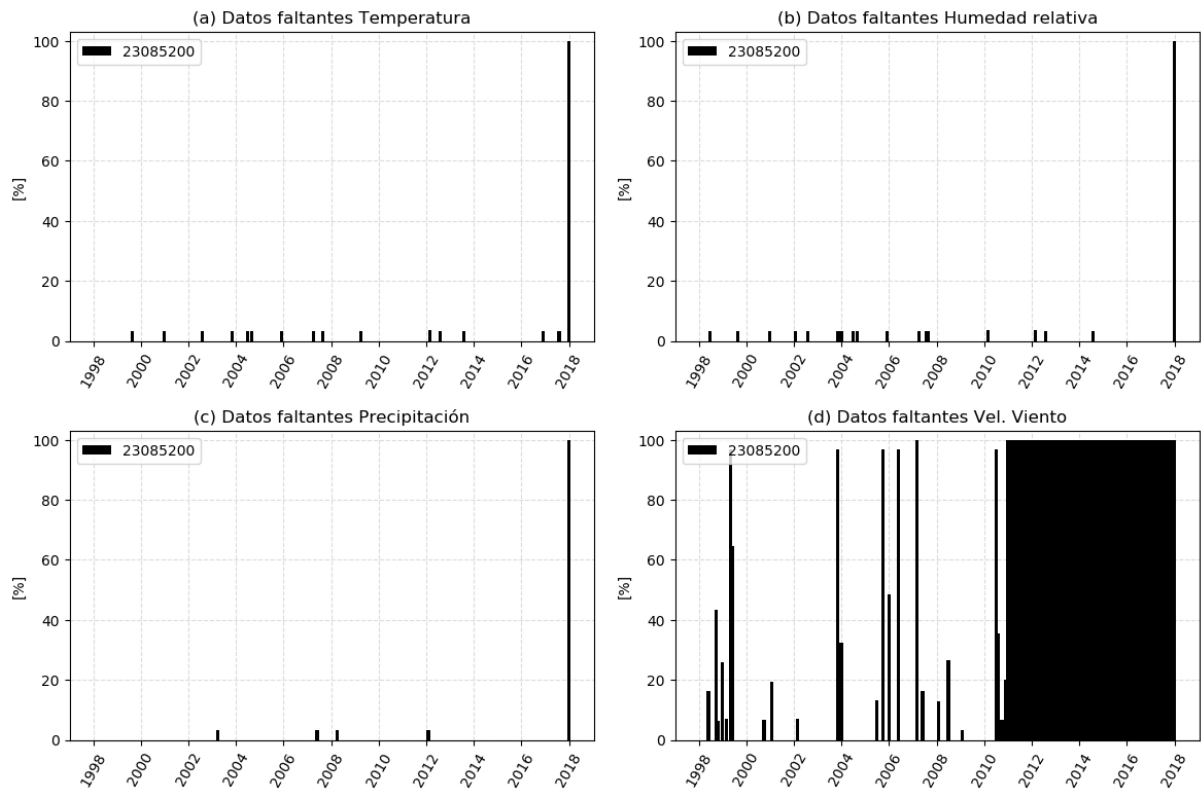


Figura 8. Cantidad de datos faltantes por mes para cada variable en la estación de Rionegro [23085200]. **(a):** Temperatura, **(b):** Humedad relativa, **(c):** Precipitación, **(d):** Vel. Viento.

Comenzando con la estación de Rionegro, en los datos diarios de temperatura se puede ver que en el último mes de 2017 faltan la totalidad de los datos y en algunos meses correspondientes a los años 1999, 2000, 2002, 2003, 2004, 2005, 2007, 2009, 2012, 2013, 2016 y 2017 se observan cantidades no mayores al 5% de datos faltantes (Figura 8-a).

Los datos faltantes para la humedad relativa en la estación de Rionegro se ubican en meses de los años mencionados anteriormente y también en porcentajes no mayores al 5%. Incluso aparece la ausencia total de datos en el último mes de 2017 al igual que en los datos de temperatura, sugiriendo que la medición de estas variables puede realizarse con los mismos equipos y metodologías (Figura 8-b).

La precipitación es la variable con datos más “completos” para la estación, aunque igual se presenta el 100% de datos faltantes en diciembre de 2017. Se notan datos perdidos de lluvia en meses de los años 2003, 2007, 2008 y 2011 en cantidades menores al 5% (Figura 8-c).

Por último, la velocidad del viento es la variable con mayor cantidad de datos faltantes en la estación. El gran bache de datos diarios perdidos ocurre entre los años 2010 y 2017 donde faltan casi la totalidad de datos para este período. En los años anteriores se observa en distintos años a través del rango temporal; desde datos faltantes por mes menores a 5% en el año 2008 hasta valores que superan el 20% e incluso llegan a 100% de datos faltantes por mes entre los años 1998-2000, 2003, 2005, 2006 y 2007 (Figura 8-d).

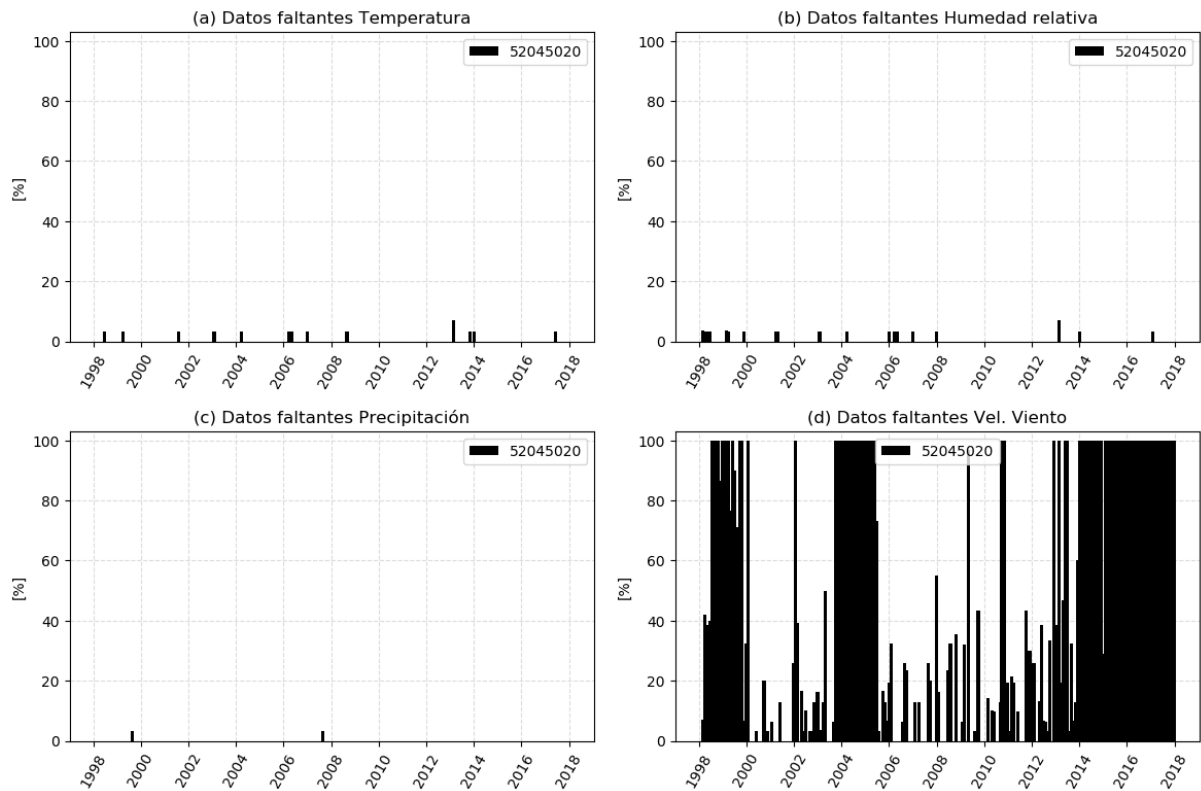


Figura 9. Cantidad de datos faltantes por mes para cada variable en la estación de Pasto [52045020]. **(a):** Temperatura, **(b):** Humedad relativa, **(c):** Precipitación, **(d):** Vel. Viento.

Los datos de temperatura para la estación de Pasto son “buenos” en cuanto a cantidad se refiere, ya que solo en un mes del 2013 se encuentran datos faltantes tímidamente por encima del 5%, mientras los demás meses permanecen por debajo de este umbral o no tienen datos perdidos (Figura 9-a).

En cuanto a humedad relativa, se observan datos faltantes en los mismos meses que la temperatura y en cantidades similares (inferiores a 5%). En el año 1998 se notan datos faltantes para varios meses, pero siempre por debajo del 5% (Figura 9-b).

La variable precipitación diaria para la estación de Pasto es la que menos datos faltantes presenta, con solo pequeñas cantidades faltantes inferiores al 5% en los años 1999 y 2007 (Figura 9-c).

En contraste, para la estación de Pasto, la variable velocidad del viento presenta una gran cantidad de datos faltantes para el período de interés, con múltiples intervalos en los cuales se alcanza el 100% de los datos perdidos por mes. Los grandes tramos de datos faltantes aparecen entre 1998-2000, 2003-2005 y 2014-2018, aunque entre estos hay una gran cantidad de meses que supera el 20% de datos perdidos y no parece haber ningún año que cuente con todos los datos para cada mes (Figura 9-d).

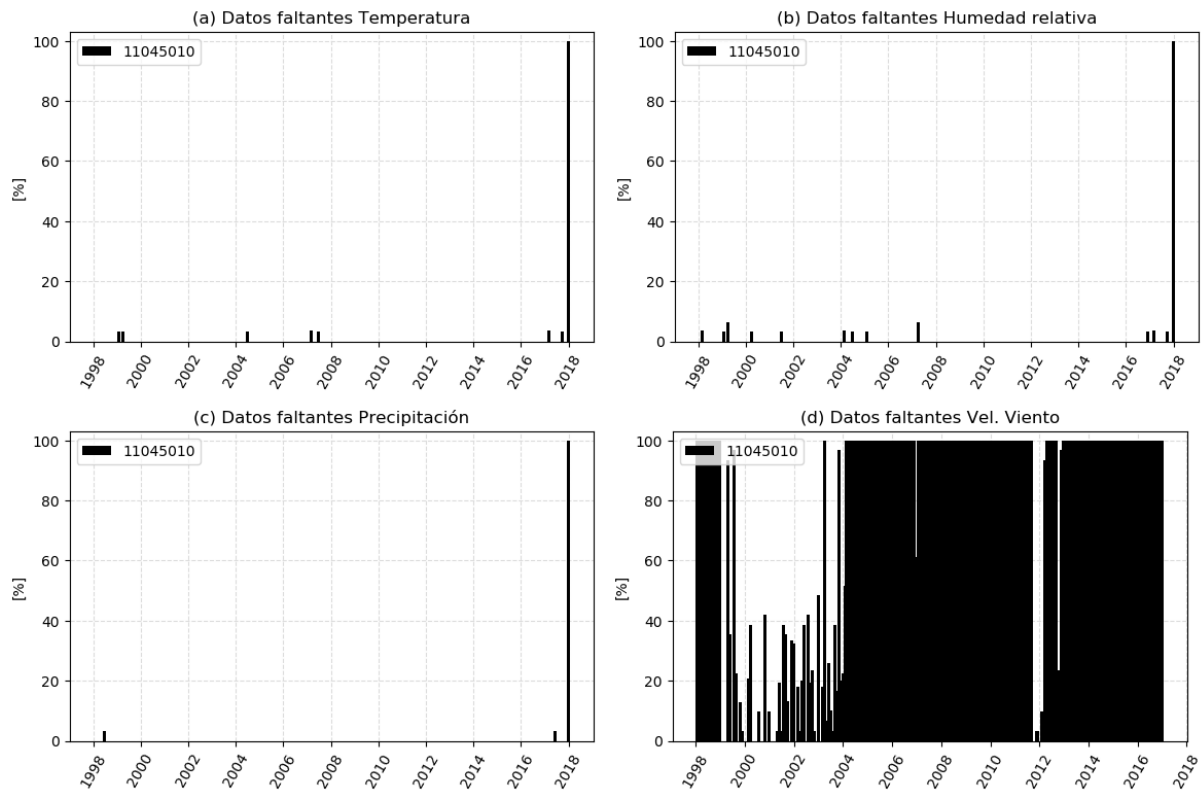


Figura 10. Cantidad de datos faltantes por mes para cada variable en la estación de Quibdó [11045010]. **(a):** Temperatura, **(b):** Humedad relativa, **(c):** Precipitación, **(d):** Vel. Viento.

La estación de Quibdó cuenta con datos faltantes de temperatura en la totalidad del mes de diciembre de 2017 y en pequeños porcentajes menores a 5% en los años 1999, 2004, 2005 y el mismo 2017 (Figura 10-a).

La humedad relativa medida en la estación también carece de la totalidad de los datos del último mes de 2017 y tiene algunos datos perdidos por mes inferiores al 5% en lugares similares a los datos de temperatura. Resaltan un mes de 1999 y otro de 2007 donde los datos faltantes se encuentran entre el 5 y el 10% (Figura 10-b).

En cuanto a la precipitación, esta estación ubicada en la región pacífica cuenta con una muy buena cantidad de datos, haciendo falta únicamente la totalidad de diciembre de 2017 y porcentajes inferiores al 5% en un par de meses en el mismo 2017 y en 1998 (Figura 10-c).

Por último, la cantidad de datos faltantes para la velocidad del viento medida en la estación es tan grande que de la gráfica se observa claramente que es mucho mayor el número de datos faltantes que de observaciones registradas. El único período donde no falta el 100% de datos mensuales se encuentra entre 1999 y 2003 y una pequeña parte de 2011-12, e incluso allí los datos perdidos por mes llegan a superar el 20% (Figura 10-d).

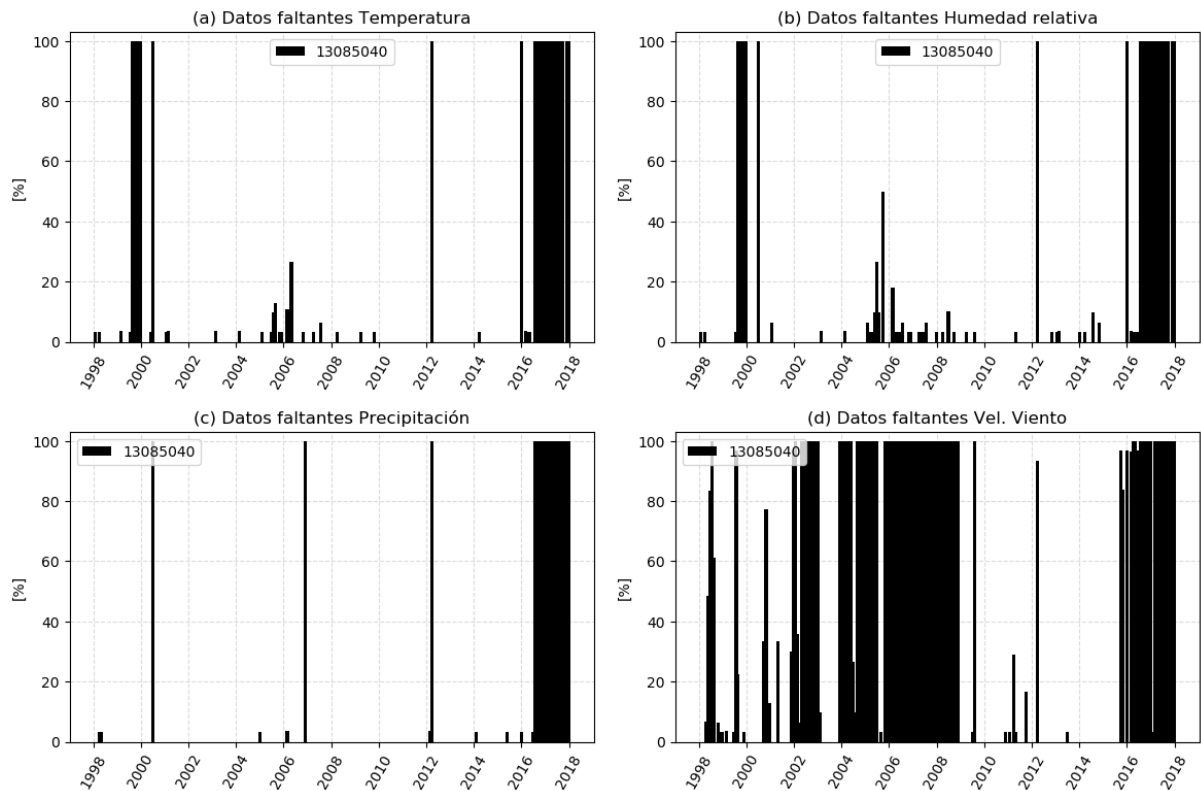


Figura 11. Cantidad de datos faltantes por mes para cada variable en la estación de Montería [13085040]. **(a):** Temperatura, **(b):** Humedad relativa, **(c):** Precipitación, **(d):** Vel. Viento.

En la estación de Montería, se nota una cantidad significativa de datos faltantes de temperatura, principalmente entre 2016 y 2018 donde faltan muchos meses en su totalidad. En los años 1999, 2000 y 2012 también se ven meses donde faltan todos sus datos diarios y entre 2005-2006 algunos meses tienen entre el 5 y el 25% de datos faltantes aproximadamente (Figura 11-a). Para la humedad relativa se notan datos faltantes en proporciones y fechas similares a donde se ubican los de temperatura, apareciendo mayor cantidad de datos faltantes entre 2005-2006 e incluso llegando a superar el 40% (Figura 11-b).

En esta estación se observa una mayor cantidad de datos faltantes de precipitación por mes en comparación con las anteriores, con un gran bache entre 2016 y 2018 y algunos meses completamente vacíos de datos en los años 2000, 2006 y 2012 (Figura 11-c).

De nuevo, la velocidad del viento es la variable con mayor cantidad de datos faltantes también en esta estación. Los agujeros más amplios de datos perdidos (donde incluso faltan años completos) se ubican entre los años 2002-2003, 2004-2005, 2006-2008 y 2016-2017. En el tramo 1998-2002 se encuentran algunos meses que, si bien cuentan con datos, la cantidad de faltantes supera el 60%, por lo que al realizar análisis posteriores estos suelen ser descartados (Figura 11-d).

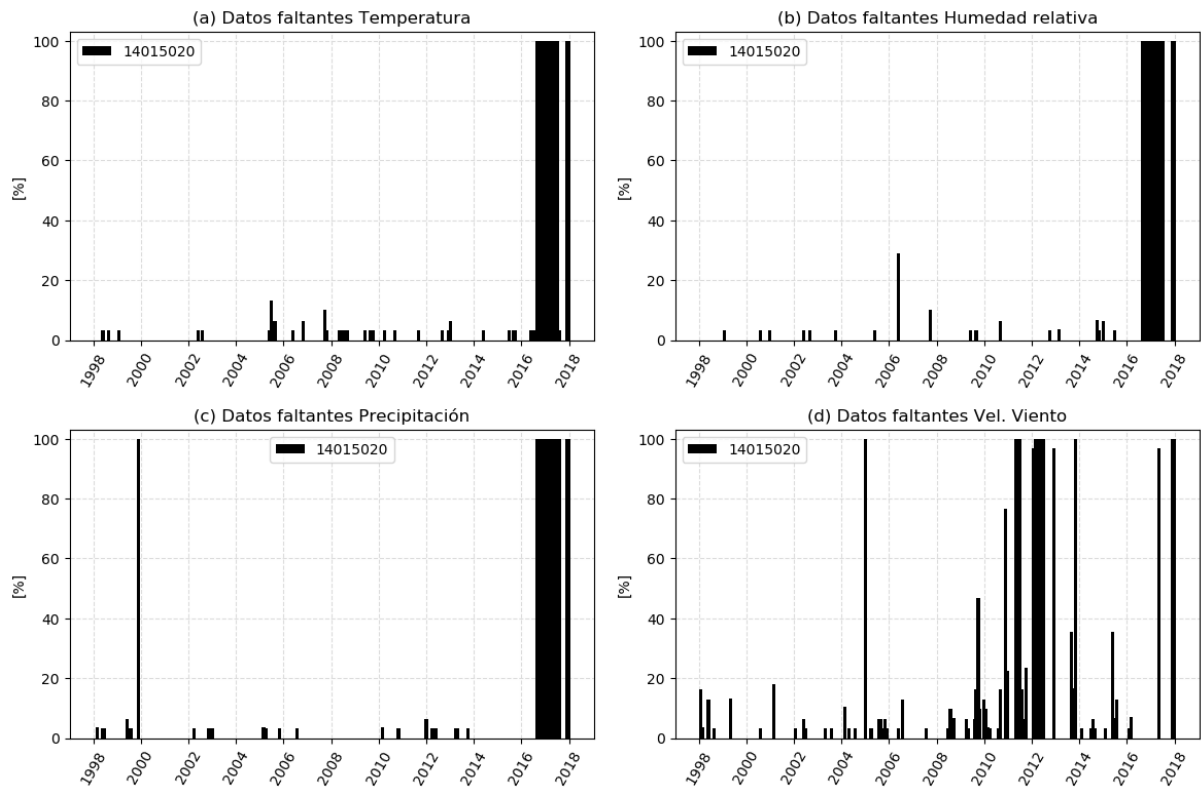


Figura 12. Cantidad de datos faltantes por mes para cada variable en la estación de Cartagena [14015020]. **(a):** Temperatura, **(b):** Humedad relativa, **(c):** Precipitación, **(d):** Vel. Viento.

En las observaciones de temperatura correspondientes a la estación de Cartagena se notan datos faltantes en un porcentaje del 100% para meses de 2016 y 2017, y algunos porcentajes entre 5 y 15% en 2005 y 2007. Algunos meses de otros años también cuentan con datos faltantes, pero bajo un umbral relativamente pequeño (<5% de datos perdidos) (Figura 12-a).

La humedad relativa para esta estación presenta datos faltantes en cantidad porcentual y ubicación temporal similar a la temperatura, resaltando además un mes en 2006 que llega a tener aproximadamente solo el 70% de la información (Figura 12-b).

La cantidad de datos faltantes para precipitación en Cartagena se asemeja a la de la estación de Montería, con un gran bache entre 2016 y 2018. Resalta también un mes del año 1999 del cual faltan la totalidad de datos (100%) (Figura 12-c).

En cuanto a la velocidad del viento, la estación de Cartagena no presenta tantos datos faltantes como se aprecian en las demás, aunque se pueden ver meses en los años 2004, 2010, 2011, 2012, 2013 y 2017 con el 100% de datos faltantes y otros “menores” por encima del 20% de datos faltantes (Figura 12-d).

Curiosamente, en esta estación es donde fueron observados los valores extremos irrealistas de velocidad del viento, implicando que la disponibilidad de datos en un lugar no necesariamente se traduce en datos que representen la realidad y que, por el contrario, los datos “contaminados” pueden afectar negativamente el análisis de la información.

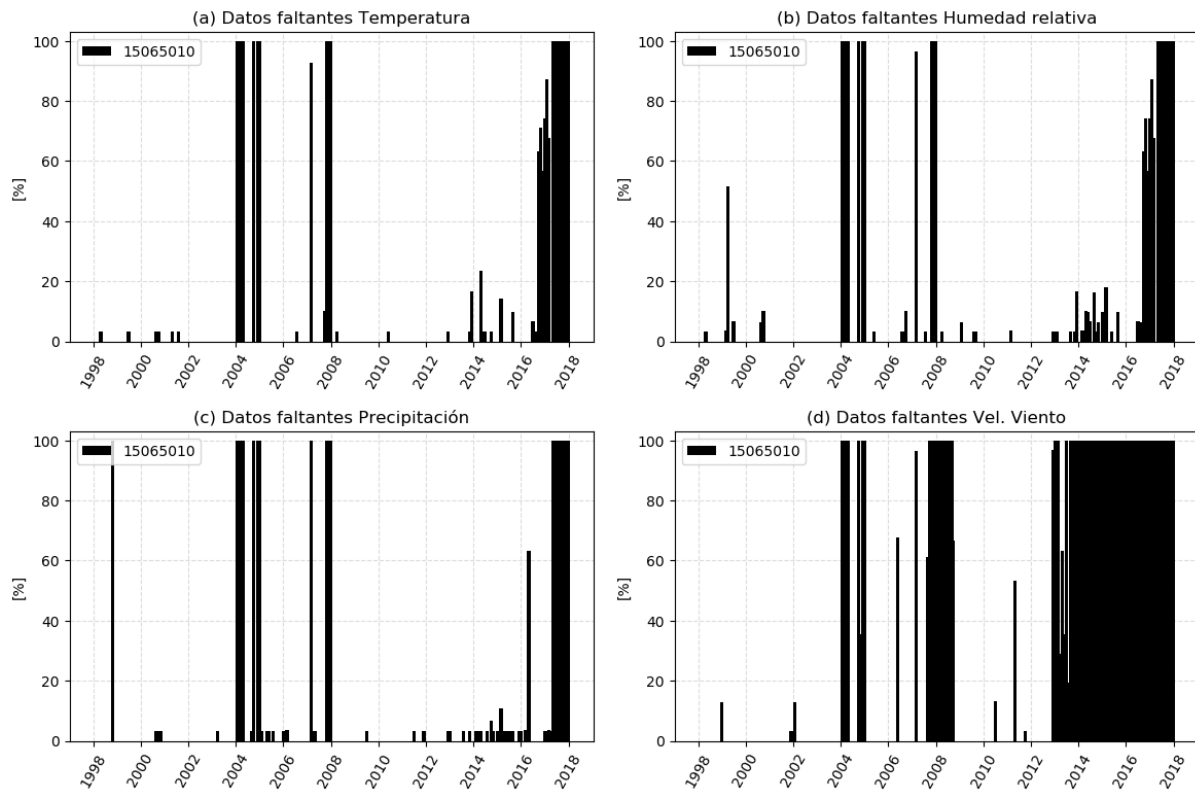


Figura 13. Cantidad de datos faltantes por mes para cada variable en la estación de Riohacha [15065010]. **(a):** Temperatura, **(b):** Humedad relativa, **(c):** Precipitación, **(d):** Vel. Viento.

Para la estación de Riohacha se encontró una cantidad notable de datos faltantes de temperatura, principalmente meses incompletos al 100% en los años 2004, 2007 y 2017. Algunos otros meses muestran un porcentaje de datos faltantes no mayor al 5% pero resaltan algunos del año 2016 donde algunos meses tienen hasta más del 60% de información perdida (Figura 13-a).

La humedad relativa medida en esta estación, y de manera consistente con las anteriores estaciones, muestra que los datos faltantes se encuentran en períodos de tiempo y magnitudes similares a los de temperatura. Destaca también un mes del año 1999, que llega a tener aproximadamente solo el 50% de datos (Figura 13-b).

En los datos de precipitación se encuentran faltantes al 100% en algunos meses de 1998, 2004, 2007 y 2017. Se aprecian también meses de porcentajes inferiores al 5% y uno en el 2016 que llega a sobrepasar el 60% de datos faltantes (Figura 13-c).

Finalmente, la velocidad del viento observada en esta estación cuenta con gran cantidad de datos faltantes, en especial entre 2013 y 2017 donde la información falta casi en su totalidad. En otros años como 2004, 2007 y 2008 también se encuentran picos que indican meses con el 100% de datos perdidos. Algunos meses en 2006, 2011 y 2013 no se encuentran vacíos, pero superan el 50% de datos faltantes (Figura 13-d).

3.1.3. Porcentaje total de datos perdidos.

A partir de las figuras anteriores donde se vio la cantidad de datos perdidos cada mes, contabilizados de manera separada para cada variable y cada estación de medición. En esta sección presentamos los consolidados de la cantidad total de valores perdidos (diarios) en el período de tiempo de estudio (1998-2017) y analizamos de manera general en cuál estación se presentan los mayores porcentajes totales y para cuáles variables. Los consolidados de datos perdidos para las estaciones y variables analizadas se presentan en la Tabla 2.

En la estación de Rionegro se encontraron porcentajes totales de datos perdidos, para toda la serie de tiempo, por debajo del 1% en las variables de temperatura, humedad relativa y precipitación. La velocidad del viento medida en la estación cuenta con una gran cantidad de datos faltantes: 39.55% para el período de interés 1998-2017.

Para la estación de Pasto, se pueden observar también porcentajes totales menores a 1% para temperatura, humedad relativa y precipitación. El caso de la precipitación en esta estación es interesante, pues es la estación que menos datos perdidos tiene para esta variable, con un porcentaje de tan solo 0.03% durante el período de interés. La cantidad de datos faltantes para velocidad del viento es superior a la de la estación de Rionegro y en general es el segundo valor más alto de datos perdidos (45.04%).

Por otra parte, para Quibdó, ya habíamos notado la gran fracción de datos perdidos a lo largo de los meses en el período 1998-2017 (ver Figura 10-d). El porcentaje total durante todo el período de tiempo ratifica esta estación como la que cuenta con la menor cantidad de datos (74.65% datos faltantes de velocidad del viento). Para las otras variables, los porcentajes totales tampoco superan el 1% de datos faltantes.

La estación de Montería es entre todas la que cuenta con mayor cantidad de datos faltantes de temperatura (10.59%) y la segunda en datos faltantes de humedad relativa y precipitación (11.10% y 8.05% respectivamente). En velocidad del viento se ubica en la tercera posición con mayor cantidad, con 41.9% de porcentaje total de datos faltantes para esta variable.

En la estación de Cartagena se encontraron porcentajes totales de datos perdidos entre 5% y 9% para todas las variables. Se resalta, nuevamente, que es la estación con menor cantidad de datos perdidos de velocidad del viento (solo 8.68%) pero a la vez es la que contiene los valores extremos considerados datos "espurios".

Finalmente, la estación de Riohacha se destaca como la estación con porcentajes más altos de datos perdidos para la humedad relativa y precipitación (11.16% y 9.56% respectivamente), la segunda en datos faltantes de temperatura (10.55%) y con una cantidad considerable de datos faltantes de velocidad del viento (34.41%) (Tabla 2).

Tabla 2. Porcentaje total de datos perdidos para cada variable en cada estación.

Porcentaje total de datos perdidos por variable				
Estación [código]	T	HR	P	Vel. V
Rionegro [23085200]	0.63	0.66	0.48	39.55
Pasto [52045020]	0.19	0.27	0.03	45.04
Quibdó [11045010]	0.52	0.62	0.45	74.65
Montería [13085040]	10.59	11.10	8.05	41.90
Cartagena [14015020]	5.97	5.83	6.57	8.68
Riohacha [15065010]	10.55	11.16	9.56	34.41

Es importante resaltar que todas las estaciones de estudio son estaciones meteorológicas principales (ubicadas en aeropuertos) que pueden ser contribuyentes de información para fuentes de datos de mayor escala, como el reanálisis ERA5, por lo que la mala medición o ausencia de medición también puede impactar negativamente la producción de conjuntos de reanálisis y reconstrucciones climáticas.

3.2. Comparación IDEAM y ERA5.

Se realizó una validación de los datos de ERA5 mediante una comparación sencilla y directa con las observaciones registradas por las estaciones del IDEAM estudiadas en las secciones anteriores. Para dicha comparación ambos conjuntos de datos fueron convertidos a escala mensual. El objetivo es explorar qué tan adecuado sería utilizar datos de ERA5 en estudios posteriores para las regiones de interés. La comparación realizada aquí se centra en valores promedio mensuales, el ciclo anual resultante, y las variaciones interanuales asociadas con tendencias de largo plazo y anomalías respecto al ciclo anual.

Para obtener los promedios mensuales de las series de tiempo provenientes de las estaciones del IDEAM se utilizó la función “resample” de la librería pandas de Python. Teniendo en cuenta la presencia de datos perdidos en las series de tiempo de todas las variables, se realizó el proceso de resampling únicamente para aquellos meses que cumplieran un umbral mínimo de 75% de datos diarios disponibles (o de otro modo, que la cantidad de datos diarios perdidos en el mes no fuera superior al 25%).

En general se esperaban valores similares para las variables ya que, como se mencionó en la sección anterior, las estaciones pueden aportar datos al reanálisis. Las diferencias que se noten entre ambas fuentes de información se pueden relacionar más a la representatividad de ERA5 al compararse con puntos fijos en el espacio y/o problemas asociados al modelo utilizado para el reanálisis. Por parte de las estaciones, estas disparidades también pueden ser influenciadas por los baches de datos faltantes que se encuentran en varias de las observaciones para una cantidad considerable de variables.

3.2.1. Temperatura.

Al comparar los promedios mensuales de temperatura de la estación de Rionegro con los datos de ERA5 para el pixel más cercano (Figura 14-a), se encontró que el reanálisis no dista mucho de las observaciones en cuanto a la magnitud de los datos y que también sigue la tendencia levemente ascendente que se observa en los promedios mensuales de temperatura para esta estación. En los datos de ERA5 se nota el efecto del fenómeno ENSO en su fase El Niño durante los años 1998 y 2015-16 y en la fase La Niña durante los años 2010-12, sobreestimando algunos valores durante El Niño 2015-16 y durante La Niña 2010-12 (aprox. 1°C más con respecto a las observaciones, ver Figura 14-a).

En la estación de Pasto es clara una subestimación en los promedios mensuales de temperatura por parte de ERA5. Los picos máximos y mínimos llegan a estar aproximadamente entre 1 y 1.5°C por debajo del valor observado. A pesar de esta subestimación, el reanálisis logra representar de buena manera las temperaturas durante la fase La Niña del ENSO en 2010-12 (Figura 14-c).

Para la estación de Quibdó se observa una subestimación por parte del reanálisis de valores un poco superiores a 1°C en gran parte de la serie (hasta 2010 aproximadamente). Los datos mensuales de ERA5 logran representar los fenómenos La Niña 2010-12 y El Niño 2015-16 observados por la estación. También se aprecia una tendencia creciente en los datos del reanálisis, implicando aumentos de temperatura promedio mensual con el pasar de los años (Figura 14-e).

Para la estación de Montería se puede ver cómo el reanálisis subestima a las observaciones durante el primer tramo de la serie (1998-2001) en valores que llegan a ser mayores a 1°C. En el tramo restante (2002-2017) resalta como ERA5 genera picos máximos de temperatura que se encuentran aproximadamente 1°C por encima de los datos medidos por la estación. Se representan las temperaturas observadas durante La Niña 2010-12 y El Niño 2015-16 relativamente bien, exceptuando los picos máximos mencionados anteriormente (Figura 14-g).

Los promedios mensuales de temperatura de ERA5 comparados con estación de Cartagena muestran una subestimación general de la variable con respecto a las observaciones (entre 1 y 1.3°C menos que las mediciones). A pesar de esta subestimación, el reanálisis no genera ni de manera aproximada el pico mínimo de temperatura observado en 2001. El reanálisis arroja temperaturas promedio mensual más bajas que las observadas durante las fases del fenómeno ENSO (El Niño 1998, La Niña 2010-12 y El Niño 2015-16) (Figura 14-i).

En cuanto a la estación de Rihacha, se puede ver como el reanálisis subestima las temperaturas observadas durante toda la serie, con 0.5-1°C menos que las observaciones en los mínimos y 1-1.5°C menos en los picos máximos. ERA5 representa con valores no muy distantes, excepto en los máximos, a las observaciones los eventos La Niña 2010-12 y El Niño 2015-16 (Figura 14-k).

Comparando el ciclo medio anual de temperatura para la estación de Rionegro con ERA5 (Figura 14-b), se observa que, aunque ERA5 sobreestima la magnitud de los valores los máximos de temperatura observados a mitad de año (mayo y agosto) y los valores mínimos a final y comienzo del año (noviembre a enero), estos coinciden en ubicación temporal (Figura 14-b).

En el ciclo medio anual para la estación de Pasto se nota una subestimación de aproximadamente 1°C por parte de ERA5 durante todo el ciclo. Dicho esto, el ciclo del reanálisis coincide correctamente con los valores máximos y mínimos de temperatura observados por la estación (agosto y noviembre, respectivamente) (Figura 14-d).

Para la estación de Quibdó se puede observar como el reanálisis subestima los datos observados y además no representa el máximo de temperatura ubicado en el mes de abril. El valor mínimo que ocurre en noviembre sí coincide entre ambas fuentes de información (Figura 14-f).

En la estación de Montería se nota como ERA5 sobreestima los valores del ciclo medio anual de temperatura en los primeros dos meses del año hasta en 0.5°C y luego pasa a subestimar la variable. Los máximos entre la estación y el reanálisis y la estación no coinciden temporalmente, pero los mínimos sí corresponden al mismo mes (noviembre) (Figura 14-h).

Para la estación de Cartagena, el ciclo medio anual generado a partir de los datos de ERA5 subestima en una magnitud relativamente amplia al generado por las observaciones, llegando a diferencias superiores a 1°C . El máximo de temperatura coincide levemente entre ambos ciclos (junio) mientras que en el mínimo no hay coincidencia alguna (enero para la estación y noviembre para ERA5) (Figura 14-j).

Finalmente, al comparar la estación de Riohacha con ERA5 se observa nuevamente una subestimación en el ciclo medio anual por parte del reanálisis que llega a ser hasta de 1.5°C . Dicho esto, tanto los valores máximos de mitad de año como los valores mínimos de comienzo de año coinciden temporalmente en ambas fuentes de información (Figura 14-l).

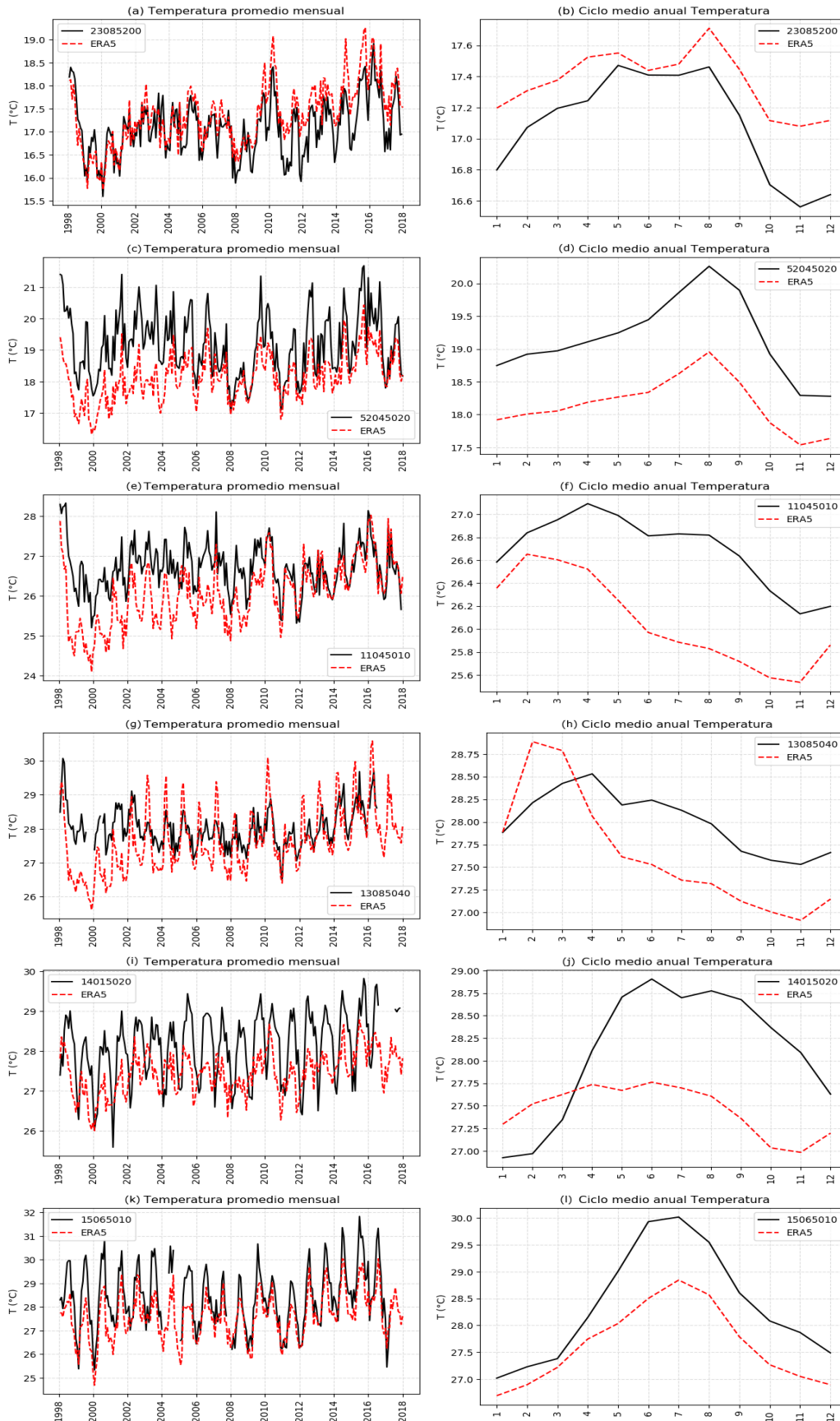


Figura 14. Valores promedio mensuales de temperatura (izq.) y ciclo medio anual de temperatura (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Rihacha [15065010].

Para comparar las variaciones de los valores mensuales graficamos además las anomalías respecto al ciclo anual y diagramas de dispersión de los valores mensuales de las observaciones vs ERA5 (Figura 15).

Para la estación de Rionegro se puede ver como los datos de observaciones y del reanálisis llevan una tendencia creciente en cuanto a esta variable. También se logra apreciar en ambos casos la presencia del fenómeno ENSO en 1998 (El Niño), 2010-12 (La Niña) y 2015-16 (El Niño) (Figura 15-a).

En el caso de la estación de Pasto se nota como el reanálisis presenta anomalías de menor magnitud en las épocas del ENSO (La Niña 2010-12 y El Niño 15-16) en comparación con las observaciones. En general ambas fuentes de información siguen una tendencia creciente (Figura 15-c).

ERA5 muestra anomalías positivas de temperatura durante la fase La Niña 2010-12 del ENSO en comparación con las anomalías negativas de la estación de Quibdó. El reanálisis muestra una tendencia creciente mientras se avanza en el tiempo (Figura 15-e).

En la estación de Montería se observan diferencias con respecto a las anomalías de ERA5 en especial en los años 1998-2001, donde el reanálisis arroja anomalías más negativas que las de las observaciones (hasta 1.5°C mayores). Frente a eventos ENSO, ambas fuentes representan manera similar las fases La Niña 2010-12 y El Niño 2015-16, además de mostrar una tendencia creciente (Figura 15-g).

Las anomalías del reanálisis coinciden en gran parte con las de la estación de Cartagena. Exceptuando algunas anomalías negativas donde ERA5 presenta anomalías negativas menores, las dos siguen un patrón de tendencia creciente y representan adecuadamente los eventos ENSO mencionados (Figura 15-i).

Por último, las anomalías de ERA5 coinciden en gran parte con las de la estación de Riohacha. Ambas fuentes de información muestran anomalías negativas y positivas para las fases La Niña y el Niño respectivamente además de seguir una tendencia creciente de temperatura (Figura 15-k).

En el análisis pareado IDEAM vs ERA5 se encontró para cada estación lo siguiente:

Para la estación de Rionegro se tiene una correlación relativamente muy buena entre los datos promedio mensuales de temperatura ($R = 0.845$) de ambas fuentes. Los valores de las observaciones aumentan a un ritmo de 0.75 veces la magnitud de los valores del reanálisis, de acuerdo con el coeficiente de pendiente de la regresión lineal (Figura 15-b).

La estación de Pasto cuenta con una correlación similar a la obtenida para la estación anterior con respecto al reanálisis ($R = 0.808$) pero en este caso los valores de la estación crecen en 1.09 veces los valores de ERA5 (Figura 15-d).

En la estación de Quibdó la correlación con ERA5 no es tan alta como en las anteriores ($R = 0.637$), por lo que la relación entre las observaciones y el reanálisis para temperatura se considera más débil. Se nota también como los valores del reanálisis crecen casi al doble de velocidad de las mediciones (coef. pendiente = 0.48) (Figura 15-f).

ERA5 muestra una correlación moderada con los datos de la estación de Montería ($R = 0.629$) a la vez que sus datos se mueven en rangos mucho más amplios que las observaciones (Figura 15-h).

Para la estación de Cartagena se encuentra la correlación más baja de temperatura con respecto a ERA5 de todas las estaciones ($R = 0.513$) y los valores observados aumentan más lentamente (coef. pendiente = 0.79) que los del reanálisis (Figura 15-j). Acá hay que notar que la figura de las anomalías (Figura 15-i) sugiere que las anomalías de las observaciones y ERA5 sí pueden estar mejor correlacionadas que lo que sugiere el valor de R . Tal vez este valor bajo de correlación se deba al efecto del ciclo anual (ver Figura 14-j), donde ERA5 no coincide con las observaciones.

Finalmente, los datos de temperatura de la estación de Riohacha cuentan con una correlación muy fuerte con respecto a los del reanálisis ($R = 0.902$) y aumentan a un ritmo no muy distinto (coef. pendiente = 1.18) de los de ERA5 (Figura 15-l).



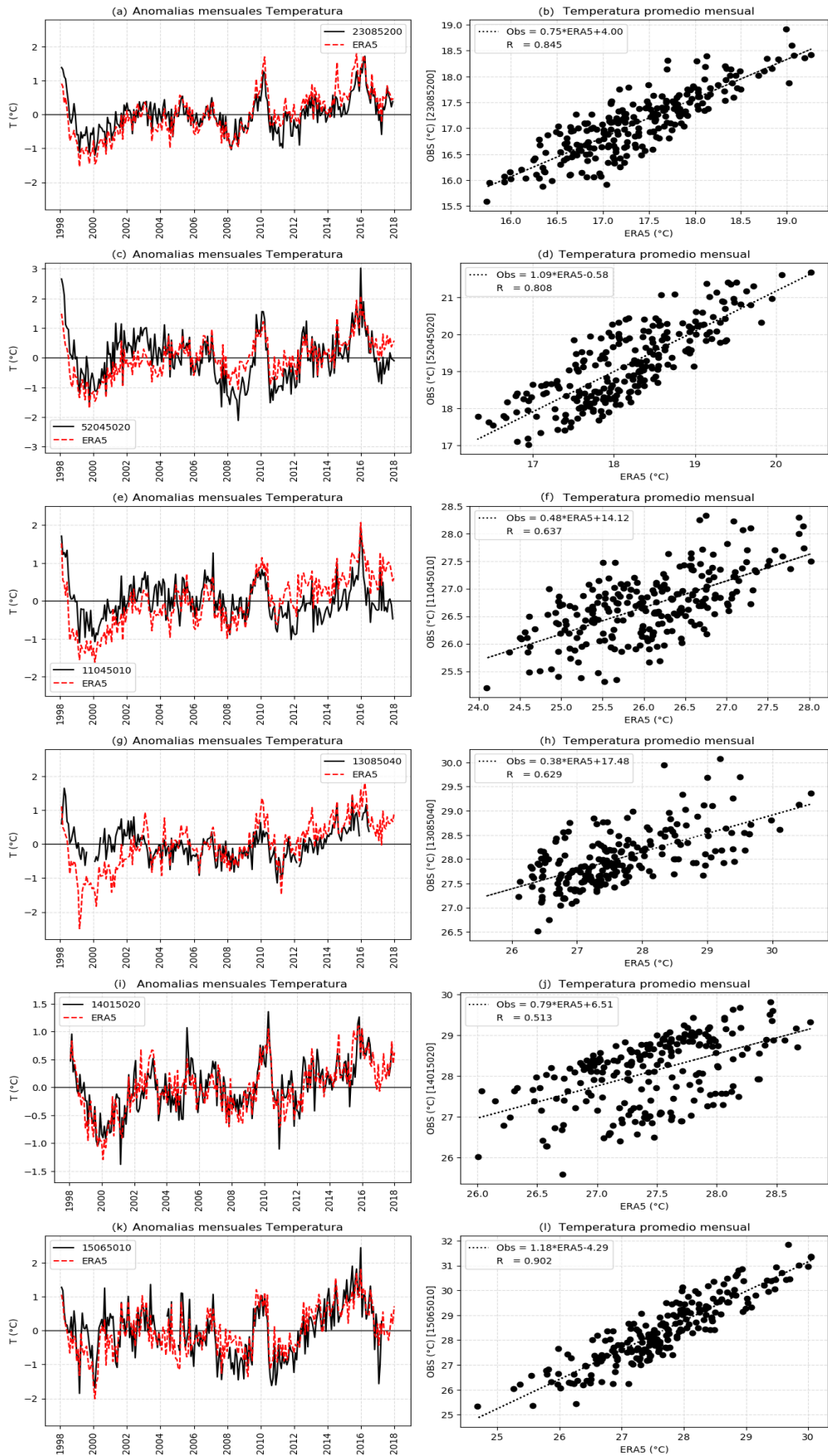


Figura 15. Anomalías mensuales de temperatura (izq.) y Scatterplot de temperatura OBS v ERA5 (der). (a)-(b): Rionegro [23085200], (c)-(d): Pasto [52045020], (e)-(f): Quibdó [11045010], (g)-(h): Montería [13085040], (i)-(j): Cartagena [14015020], (k)-(l): Riohacha [15065010].

3.2.2. Humedad relativa.

Los promedios mensuales de esta variable en la estación de Rionegro en comparación con ERA5 muestran una sobrestimación clara por parte del reanálisis. No se aprecia una tendencia clara en el comportamiento de la humedad relativa, pero se puede notar como los valores de las observaciones durante El Niño 2015-16 disminuyen al igual que para ERA5 (Figura 16-a).

En la estación de Pasto los valores del reanálisis son muy semejantes a las observaciones en promedio, aunque algunos picos mínimos de humedad observados en 2012, 2015 y 2016 son representados por ERA5 en menor magnitud (Figura 16-c).

En el caso de la estación de Quibdó se nota una amplia sobrestimación por parte de ERA5 con respecto a la humedad relativa del lugar, mostrando picos máximos incluso en tramos donde las observaciones registraron valores mínimos. Se observa de manera sutil el efecto de El Niño 2015-16 sobre esta variable (Figura 16-e).

Para la estación de Montería se puede ver como la amplitud de los datos de ERA5 entre máximos y mínimos es mayor a la de las observaciones. Para esta estación no se aprecia una tendencia clara en cuanto al comportamiento de la variable y solo se notan levemente los efectos de El Niño en 1998 y 2015-16 en la humedad relativa del lugar (Figura 16-g).

Los datos de ERA5 para comparar con la estación de Cartagena producen una gráfica similar a la que se vio para Montería, siendo en este caso más clara la presencia del fenómeno El Niño en 2015-16. Se notan discrepancias entre el reanálisis y las observaciones en algunos puntos particular donde la estación indica aumentos mientras ERA5 muestra mínimos de humedad relativa, como en 2000 y 2003 (Figura 16-i).

Finalmente, al comparar los promedios mensuales de ERA5 con la estación de Riohacha se aprecia una leve sobrestimación de la humedad relativa por parte del reanálisis, pero en general se sigue el mismo comportamiento. Se nota el aumento de humedad durante La Niña 2010-12 y la disminución en los valores de esta variable durante El Niño 2015-16 (Figura 16-k).

La comparativa del ciclo medio anual para la estación de Rionegro muestra que, si bien ERA5 sobrestima los valores de humedad relativa observados (2 a 4% más humedad), coinciden en las épocas donde la variable alcanza sus valores máximos (MAM y SON) y sus valores mínimos (JJA) (Figura 16-b).

Para la estación de Pasto la comparación del ciclo medio anual también se observa una sobrestimación por parte del reanálisis que varía entre 4-7.5% aproximadamente de más humedad. El ciclo para ERA5 coincide con las observaciones en cuanto al patrón estacional, ubicando los mínimos de humedad relativa de julio a septiembre y los máximos en noviembre-diciembre (Figura 16-d).

Para la estación de Quibdó se puede ver de manera inmediata una gran diferencia entre el ciclo medio anual de ERA5 y el de las observaciones. El reanálisis sobrestima la variable para este lugar y la diferencia entre las dos fuentes de datos llega a ser superior al 5%. En cuanto a la estacionalidad ERA5

discrepa totalmente con las observaciones a excepción de los primeros dos meses del año (Figura 16-f).

En la estación de Montería se observa como ERA5 subestima los mínimos de humedad que ocurren al comienzo del año y luego pasa a sobrestimar el resto de las observaciones. Temporalmente tanto los mínimos de enero-marzo como los máximos de septiembre-noviembre coinciden para los dos ciclos (Figura 16-h).

Similar al comportamiento en la estación de Montería, el reanálisis ERA5 subestima la variable durante el primer trimestre del año y los meses restantes sobrestima los valores de la misma con respecto a las observaciones de la estación de Cartagena. Los mínimos entre enero-marzo y máximos entre septiembre-noviembre coinciden para las dos fuentes de datos (Figura 16-j).

Por último, ERA5 sobrestima los valores de humedad con respecto a los de la estación de Rihacha hasta en más del 5% de humedad en algunas épocas. Ambos ciclos siguen el mismo patrón estacional, coincidiendo los picos de valores máximos en abril y octubre y también los picos mínimos en febrero y julio (Figura 16-l).



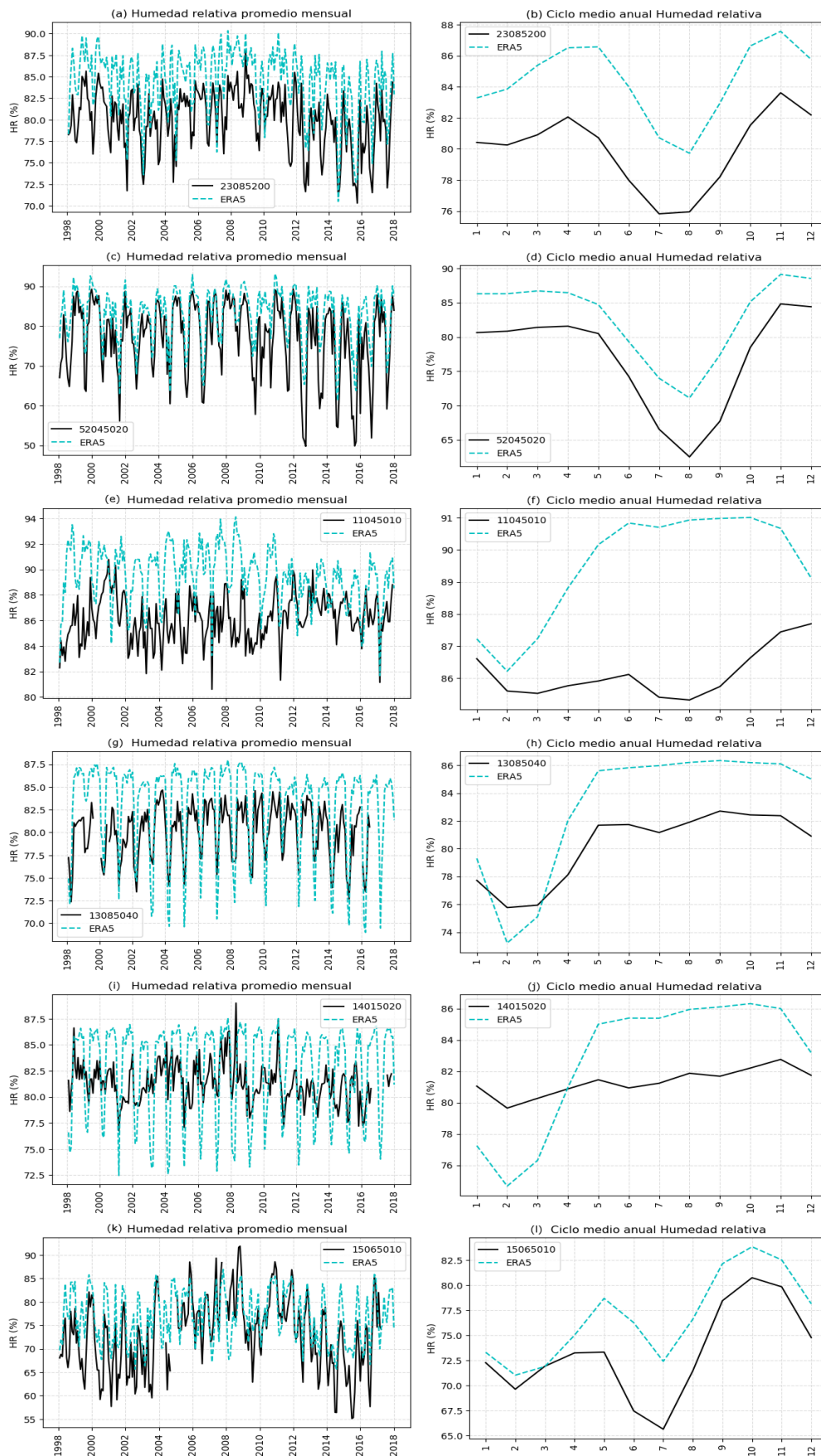


Figura 16. Valores promedio mensuales de humedad relativa (izq.) y ciclo medio anual de humedad relativa (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

Comparando las anomalías mensuales de humedad relativa para la estación de Rionegro con ERA5 se observan valores no muy distintos entre ambas fuentes de datos, siguiendo las dos un patrón levemente decreciente donde se aprecia el efecto de La Niña 2010-12 (anomalías más positivas) y de El Niño 2015-16 sobre la variable (con anomalías más negativas por parte de ERA5 durante este tramo) (Figura 17-a).

En la estación de Pasto se nota como el reanálisis cuenta con anomalías de menor amplitud a lo largo del período de tiempo. Algunos picos mínimos de las anomalías observadas (que indican hasta 20% menos de humedad) no son capturados por ERA5, como en 2010, 2012 y 2015. No se observa una tendencia clara de la variable (Figura 17-c).

Para la estación de Quibdó se ven discrepancias considerables con respecto a ERA5, puesto que en años como 2001, 2012, 2013 y 2015 las observaciones indican anomalías positivas mientras el reanálisis muestra anomalías negativas. No se nota de manera clara una tendencia y los efectos de las fases del ENSO se aprecian de manera moderada (Figura 17-e).

La comparativa en la estación de Montería muestra como ERA5 genera picos de anomalías positivas y negativas con mayor amplitud que lo observado por la estación. Se puede intuir una tendencia levemente decreciente a partir del año 2014 pero también es probable que estas anomalías negativas estén asociadas a El Niño 2015-16 (Figura 17-g).

Para Cartagena, las observaciones de la estación muestran anomalías positivas de mayor amplitud (hasta de 8% de humedad) mientras que ERA5 muestra anomalías más negativas (hasta -4.5% aproximadamente). Se pueden ver de manera moderada los aumentos y disminuciones en la variable debido a La Niña 2010-12 y a El Niño 2015-16, respectivamente (Figura 17-i).

Finalmente, las anomalías mensuales de la estación de Rihacha muestran una mayor amplitud entre sus valores positivos y negativos al compararla con ERA5. Si bien no se nota una tendencia clara, llaman la atención los dos "valles" de anomalías negativas formados entre 2000-2004 y 2014-2016, siendo el segundo tal vez consecuencia del fenómeno El Niño de 2015-16 (Figura 17-k).

Al realizar el análisis pareado entre las estaciones del IDEAM y los datos de ERA5 se encontró lo siguiente:

Para la estación de Rionegro existe una muy buena correlación ($R = 0.89$) entre las observaciones y los datos de ERA5 representativos del lugar. Los valores de las observaciones cambian más lentamente que los valores del reanálisis (coef. pendiente = 0.83) (Figura 17-b).

Para la estación de Pasto se observa una correlación muy buena con los datos provenientes del reanálisis ($R = 0.923$), además de que los valores para las observaciones incrementan más rápidamente (coef. pendiente = 1.28) que los datos de ERA5 (Figura 17-d).

En el caso de la estación de Quibdó, se encuentra una correlación pobre o mala entre las observaciones y ERA5 ($R = 0.195$), además de un crecimiento muy lento en los datos de la estación (coef. pendiente = 0.16 veces) con respecto a los del reanálisis (Figura 17-f).

Continuando con la estación de Montería, la correlación de los datos con ERA5 es buena ($R = 0.839$) y las observaciones aumentan de forma más lenta (coef. pendiente = 0.49) que los datos del reanálisis (Figura 17-h).

Para la estación de Cartagena se obtuvo una correlación débil con los datos de humedad de ERA5 ($R = 0.379$). También se notó que las observaciones crecen muy lento (coef. pendiente = 0.17) en comparación con ERA5 (Figura 17-j).

Por último, la correlación entre los datos de humedad relativa de la estación de Riohacha y los datos de ERA5 se considera como buena ($R = 0.765$) y además los datos de ambos aumentan en magnitudes casi similares (coef. pendiente = 1.06) (Figura 17-l).



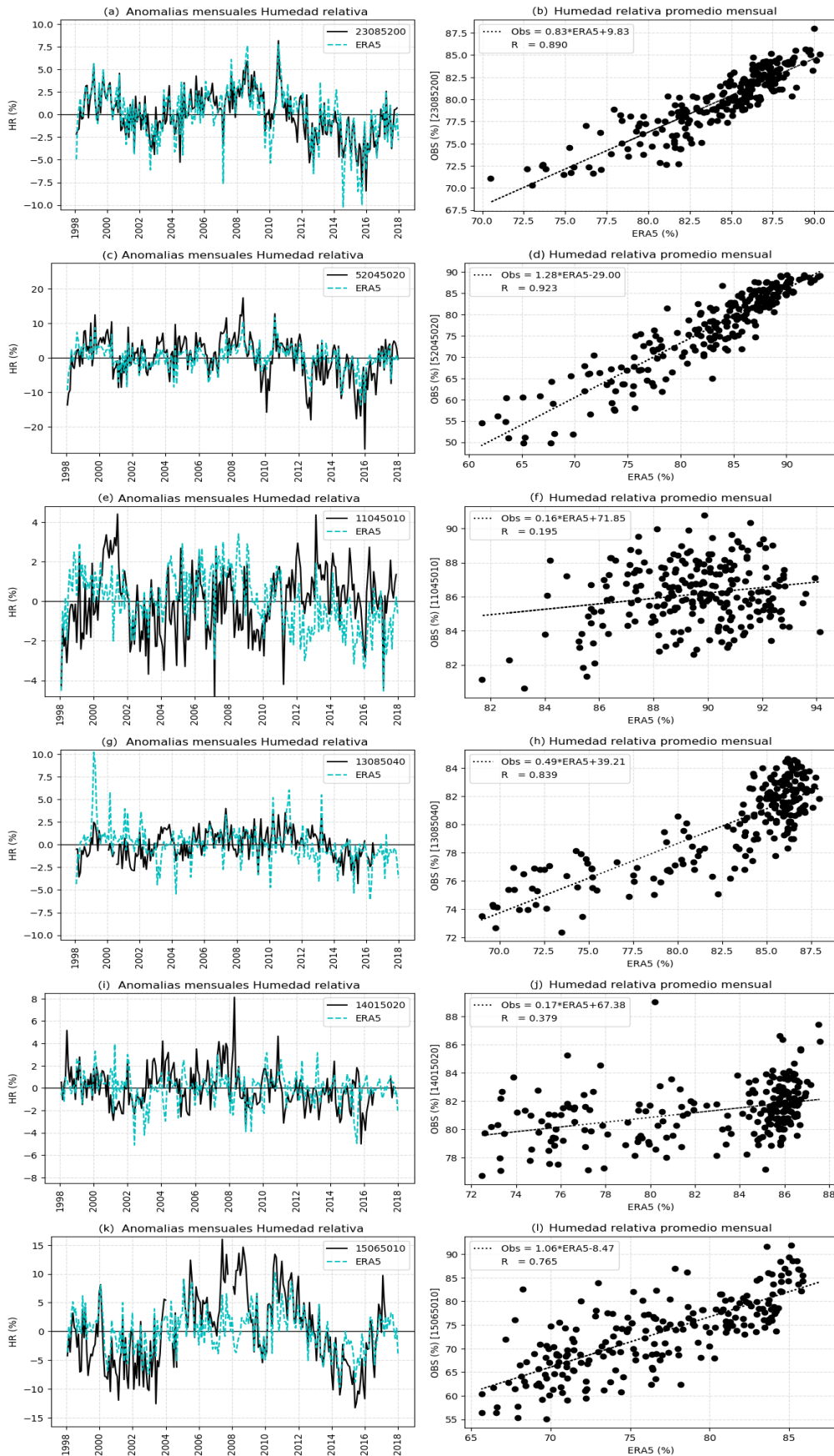


Figura 17. Anomalías mensuales de humedad relativa (izq.) y Scatterplot de humedad relativa OBS v ERA5 (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

3.2.3. Precipitación.

Comparando con los promedios mensuales de lluvias de la estación de Rionegro (Figura 18-a), se nota como ERA5 sobrestima los valores a lo largo de la serie de tiempo en especial en algunos picos máximos. Resalta el pico máximo de las observaciones en el año 2011, posiblemente asociado al fenómeno La Niña, el cual ERA5 representa con valores mucho menores.

En la estación de Pasto se nota una sobrestimación por parte de ERA5 que alcanza a ser tan amplia como 400mm de lluvia de diferencia entre ambas fuentes de datos (por ejemplo, entre 2010 y 2012). Se notan ligeramente los picos más altos durante La Niña 2012-12 y como decrece la lluvia durante El Niño 2015-16 (Figura 18-c).

Para la estación de Quibdó los valores de ERA5 se notan muy similares a las observaciones. Se logran apreciar los efectos del ENSO en 2010-12 (La Niña) y la disminución de la lluvia en 2015-16 (El Niño), además destacan un pico máximo y otro mínimo de gran amplitud en el año 2017 pertenecientes a las observaciones que ERA5 no muestra (Figura 18-e).

Comparando ERA5 con la estación de Montería se puede ver como el reanálisis sobrestima la lluvia en este lugar incluso en tramos donde la lluvia observada eran 0 mm (por ejemplo, entre 2016 y 2017). Para esta estación no es tan claro el efecto del fenómeno ENSO, pero se alcanza a notar la disminución de las lluvias durante la fase El Niño 2015-16 (Figura 18-g).

ERA5 sobrestima los valores mensuales de lluvia en comparación con lo observado en la estación de Cartagena en gran parte de los datos, incluyendo las precipitaciones ocurridas durante El Niño 2015-16. Contrario a esto, algunos picos máximos de lluvia observados por la estación como en el año 2007 o los años 2010-12 (asociados a La Niña) son subestimados por el reanálisis, mostrando valores menores a lo observado (Figura 18-i).

En cuanto a la estación de Riohacha, se aprecia una sobrestimación general por parte de ERA5. Resaltan especialmente los valores por encima de las observaciones durante los 2014 a 2017 entre los cuales ocurre una fase del ENSO (Figura 18-k).

Es importante mencionar que a escala mensual ni las observaciones ni el reanálisis muestran tendencias como las que se notan en los registros de temperatura. Esto no significa que no existan tendencias en las lluvias en absoluto. Algunos estudios sugieren que por una parte puede aumentar la intensidad de la lluvia durante períodos cortos (minutos/horas) al tiempo que se hace menos frecuente la lluvia. Este balance puede enmascarar las posibles tendencias cuando el análisis se hace con valores mensuales.

El ciclo medio anual de precipitación para la estación de Rionegro en comparación con ERA5 muestra que, aunque el reanálisis sobrestima las magnitudes de la lluvia, genera adecuadamente la estacionalidad del lugar. Se puede ver el régimen bimodal de lluvias correspondiente a la región y también coinciden los máximos (en mayo y octubre) y los mínimos (en enero y julio) (Figura 18-b).

En la estación de Pasto el reanálisis acierta en la forma del régimen bimodal de lluvias, aunque lo hace sobrestimando los valores. Además, coincide los dos mínimos (febrero y agosto) y en uno de los máximos (abril). El otro valor máximo, que ocurre en noviembre, lo ubica un mes después (Figura 18-d).

Para la estación de Quibdó ERA5 muestra un ciclo medio anual muy distinto al de las observaciones. Logra coincidir en el valor mínimo de lluvias del mes de febrero (subestimando la magnitud) pero ubica su valor máximo en octubre, mes donde las observaciones muestran un valle mínimo o de disminución en la precipitación (Figura 18-f).

ERA5 genera un régimen monomodal adecuado para la estación de Montería, si bien sobrestima la cantidad de lluvia con respecto a las observaciones. Ambos ciclos anuales coinciden en el valor mínimo (enero) y en el máximo (mayo), pero en las observaciones la precipitación comienza a decrecer en el mes de septiembre mientras que el reanálisis ubica esta disminución en el mes de octubre (Figura 18-h).

Para la estación de Cartagena el reanálisis genera un ciclo anual que sobrestima lo observado pero que sigue la forma del régimen bimodal de la región. Ambos ciclos coinciden en los valores máximos (mayo y octubre) y en los valores mínimos (enero-febrero y junio) de precipitación durante el año (Figura 18-j).

Finalmente, ERA5 produce un ciclo medio anual para la estación de Riohacha que sigue correctamente el régimen bimodal correspondiente a la región, aunque sobrestimando los valores de lluvia. La ubicación de los valores máximos (mayo y octubre) y mínimos (febrero y julio) coincide para los dos ciclos (Figura 18-l).

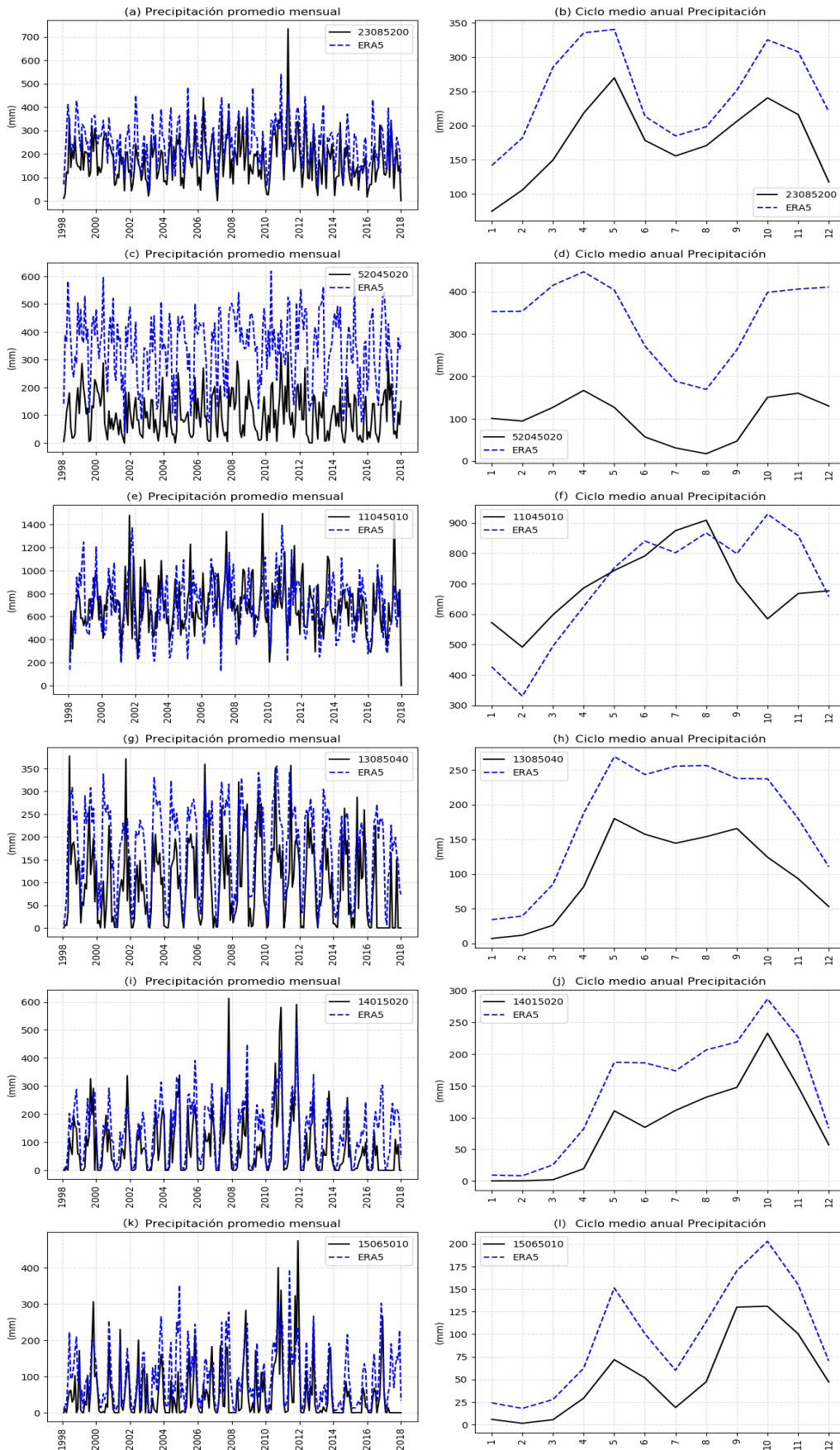


Figura 18. Valores mensuales totales de precipitación (izq.) y ciclo medio anual de precipitación (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

Comparando las anomalías mensuales de precipitación se encontró lo siguiente:

Para la estación de Rionegro las anomalías de ERA5 no son muy distintas a las de las observaciones y varían en general entre -200 y 200 mm de lluvia. Se nota una anomalía máxima de precipitación en las observaciones durante el evento La Niña 2010-12 que el reanálisis no detecta (Figura 19-a).

En la estación de Pasto se puede ver como las anomalías generadas por ERA5 sobrestiman a las observadas en la estación. Particularmente, aparecen anomalías mucho más negativas en los años 1998, 2007 y 2015 (este último asociado a El Niño 2015-16) (Figura 19-c).

El reanálisis muestra anomalías con menor amplitud para la estación de Quibdó, donde las observaciones alcanzan anomalías positivas y negativas no detectadas por ERA5 (por ejemplo, en 2007 y 2017) (Figura 19-e).

En la estación de Montería se muestra de manera clara el efecto del fenómeno La Niña 2010-12 en las anomalías durante ese período tanto para observaciones como para ERA5. Resalta el hecho que durante parte de 2016 el reanálisis muestra anomalías positivas o negativas no muy grandes, mientras que las observaciones indican reducciones superiores a 150 mm de lluvia (posiblemente debido a El Niño 2015-16) (Figura 19-g).

En la estación de Cartagena las anomalías de ERA5 se asemejan a las de los datos observados, pero el reanálisis tiene dificultad a la hora de representar valores de mayor magnitud como los picos máximos durante La Niña 2010-12 y el mínimo de 2016 asociado a El Niño 2015-16. En ambos casos el reanálisis subestima los cambios en la precipitación (Figura 19-i).

Por último, al comparar las anomalías de la estación de Riohacha con ERA5 se aprecia una variación similar para ambas fuentes de datos. Resalta como el reanálisis produce anomalías positivas que las observaciones muestran como negativas en los años 2004-2005 y que no logra representar los mayores incrementos de lluvia durante el tramo 2010-2012 (Figura 19-k).

Al realizar el análisis pareado IDEAM vs ERA5 se observó, para la estación de Rionegro, una correlación relativamente buena moderada ($R = 0.722$) entre las observaciones y el reanálisis además que los datos observados muestran cambios con amplitudes algo menores que los datos de ERA5 (coef. pendiente = 0.75) (Figura 19-b).

Para la estación de Pasto, la gráfica muestra mayor dispersión en los valores comparados ($R = 0.672$). Los datos observados crecen a una tasa mucho más lenta (coef. pendiente = 0.41) que los de ERA5 (Figura 19-d).

Para la estación de Quibdó la correlación entre observaciones y ERA5 es moderada/débil ($R = 0.436$) y las observaciones crecen lentamente (coef. pendiente = 0.40) con respecto a los datos del reanálisis. Se aprecian algunos datos muy dispersos que se alejan del ajuste de la regresión lineal (Figura 19-f).

Para la estación de Montería la correlación entre observaciones y ERA5 es relativamente moderada ($R = 0.711$) y los datos del reanálisis aumentan más rápidamente en comparación con las observaciones (coef. pendiente = 0.68) (Figura 19-h).

Para la estación de Cartagena se observa una correlación con ERA5 relativamente buena y un poco mejor en comparación con las demás estaciones ($R = 0.796$). Los datos de la estación crecen de manera más lenta que los de ERA5 (coef. pendiente = 0.82) y se notan algunos datos dispersos que pueden modificar la correlación entre las dos fuentes de datos (Figura 19-j).

Finalmente, la correlación entre los datos de la estación de Riohacha y los datos de ERA5 se puede considerar como moderada ($R = 0.678$). En esta estación los datos observados también crecen con menor rapidez que los del reanálisis (coef. pendiente = 0.67) (Figura 19-l).



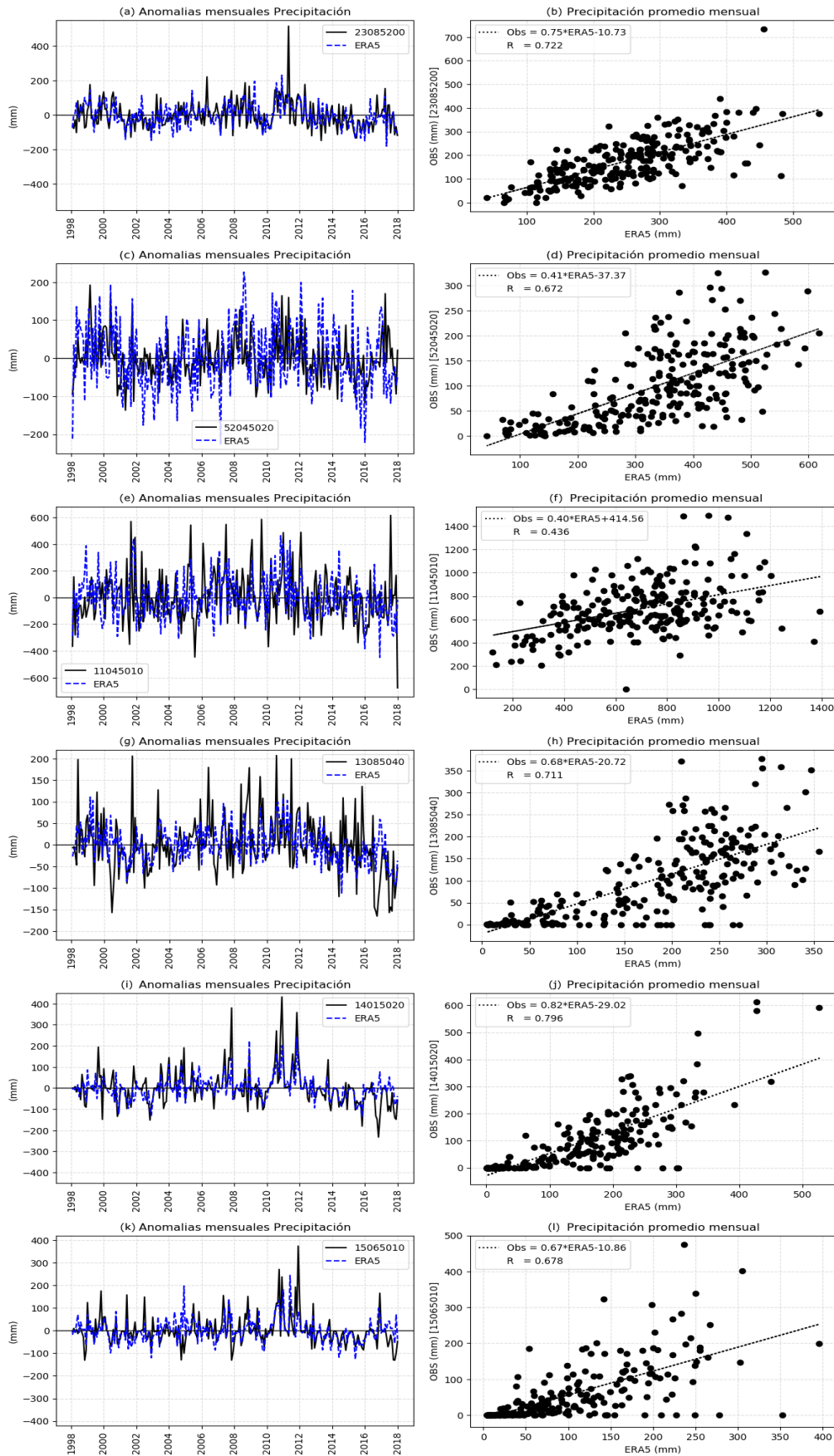


Figura 19. Anomalías mensuales de precipitación (izq.) y Scatterplot de precipitación OBS v ERA5 (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

3.2.4. Velocidad del viento

Debido a la gran cantidad de datos faltantes en general para esta variable, no es posible realizar un análisis muy detallado sobre su comportamiento y sobre si la representación de ERA5 es adecuada o no.

Para la estación de Rionegro ERA5 subestima los valores promedio mensuales de velocidad del viento con respecto a las observaciones. En algunos años esta subestimación puede superar los 1.5 m/s (Figura 20-a).

Para la estación de Pasto el reanálisis subestima los vientos en cantidades considerables (hasta más de 2 m/s) (Figura 20-c).

En el caso de la estación de Quibdó, los valores disponibles de vientos no distan tanto de los de ERA5 (se observa una ligera subestimación), pero debido a la gran cantidad de datos faltantes no se puede asegurar qué tan buena o mala es la estimación de la variable (Figura 20-e).

En la estación de Montería el reanálisis ERA5 sobrestima los vientos e incluso en algunos tramos parece ir en dirección contraria a las observaciones (ERA5 aumenta cuando las observaciones disminuyen). La gran cantidad de datos faltantes no permite analizar profundamente la comparación (Figura 20-g).

Para la estación de Cartagena ERA5 sobrestima los valores las velocidades del viento alrededor de 2.5 m/s en gran parte de la serie, especialmente los picos. Las observaciones presentan algunos datos irreales (superan los 25 m/s) para la variable que condicionan el análisis de vientos para esta estación (Figura 20-i).

En la estación de Riohacha se nota la sobrestimación por parte de ERA5 en los vientos, aunque el reanálisis trata de representar las disminuciones y aumentos de la variable. La cantidad de datos faltantes no permite realizar un análisis más concreto (Figura 20-k).

En cuanto a la comparación del ciclo medio anual, se pueden realizar unas cortas observaciones:

En la estación de Rionegro ERA5 muestra una gran subestimación con respecto a los datos observados (valores hasta 300% menores que los de la estación). Aun así, el reanálisis logra coincidir con las observaciones en los valores mínimos (abril y noviembre) y en los valores máximos (julio-agosto) (Figura 20-b).

Para la estación de Pasto ocurre algo similar a la de Rionegro; ERA5 subestima de gran manera las velocidades del viento, pero logra coincidir en el máximo del mes de agosto (Figura 20-d).

La estación de Quibdó cuenta con un ciclo medio anual para vientos, pero debido a la gran cantidad de datos faltantes no es posible afirmar o negar que ERA5 represente la estacionalidad de buena manera. En cuanto a magnitud, el reanálisis subestima los vientos en aproximadamente 0.5 m/s (Figura 20-f).

Para la estación de Montería los datos de ERA5 sobrestiman la magnitud de los vientos y no coinciden los máximos ni los mínimos con respecto a las observaciones (Figura 20-h).

ERA5 muestra un ciclo medio anual para la estación de Cartagena donde se subestiman los valores en un tramo mientras en otro se sobrestiman. El ciclo medio anual de las observaciones se nota “desplazado” con respecto al de ERA5 y la amplitud similar entre ambos ciclos puede deberse a los valores atípicos “espurios” que afectan el ciclo de las observaciones. No coinciden en máximos o mínimos (Figura 20-j).

Si bien la estación de Rihacha cuenta con bastantes datos faltantes, los datos de ERA5 logran coincidir en algunos meses como la disminución en mayo y el máximo de julio. En cuanto a magnitud el reanálisis sobrestima la magnitud de los vientos hasta en un 169% en ciertas épocas del año. Esto es importante y se debe tener en cuenta a la hora de utilizar información de ERA5 en investigaciones sobre energía eólica, aprovechamiento de gran interés en la región del norte del país (Figura 20-l).



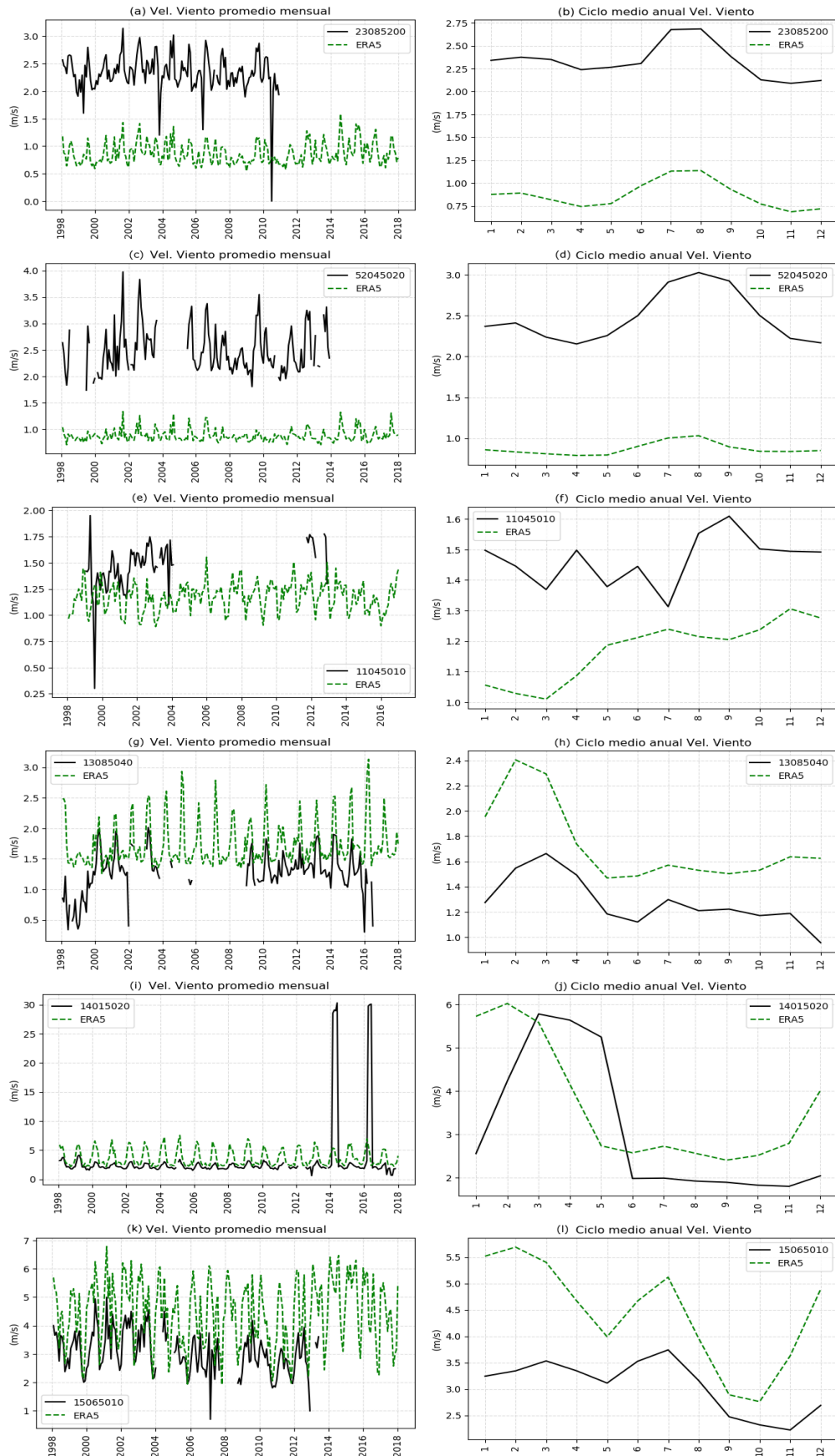


Figura 20. Valores promedio mensuales de Vel. viento (izq.) y ciclo medio anual de Vel. viento (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Rihacha [15065010].

Al comparar las anomalías mensuales de velocidad del viento se pudo observar lo siguiente:

En la estación Rionegro tanto las observaciones como ERA5 varían en magnitudes similares, a excepción de una anomalía negativa de observaciones ubicada en el año 2010 (Figura 21-a).

Para la estación de Pasto las observaciones muestran mayor amplitud en las oscilaciones comparadas con ERA5. Algunas anomalías negativas (1999-2000 y 2008-2010) y positivas (2001-2003) no son representadas por el reanálisis (Figura 21-c).

Para la estación de Quibdó los pocos datos disponibles no concuerdan con las anomalías de ERA5 (Figura 21-e).

Para la estación de Montería también hay una gran cantidad de datos faltantes. Resalta como ERA5 muestra anomalías positivas al inicio del 2016 mientras las observaciones indican anomalías negativas (Figura 21-g).

Para la estación de Cartagena se aprecian valles correspondientes a anomalías negativas para las observaciones, mientras las anomalías de ERA5 se mantienen cercanas al cero. Los valores atípicos de 2014 y 2016 medidos por la estación sobresalen en la gráfica y condicionan la visibilidad de otros tramos (Figura 21-i).

Finalmente, ERA5 representa relativamente bien las anomalías observadas en la estación de Riohacha, exceptuando algunas anomalías negativas en el año 2006-2007. Se nota también una cantidad considerable de datos faltantes (Figura 21-k).

En los análisis pareados en general se nota la influencia de los datos faltantes o de los datos "contaminados" a la hora de relacionar las dos fuentes de datos. Para cada estación se encontró lo siguiente:

A pesar de los datos faltantes, los datos de velocidad del viento de la estación de Rionegro presentan una correlación relativamente moderada con respecto a los de ERA5 ($R = 0.679$) y las observaciones crecen más rápido que los datos de ERA5 (coef. pendiente = 1.25) (Figura 21-b).

Para la estación de Pasto existe una moderada correlación con los datos de ERA5 ($R = 0.704$) y también se nota que las observaciones crecen mucho más rápido que los datos del reanálisis (coef. pendiente = 2.82) (Figura 21-d).

Los pocos datos disponibles para la estación de Quibdó tienen una correlación muy pobre con respecto a los de ERA5 ($R = 0.074$) y la regresión lineal también muestra que las observaciones cambian muy lentamente con respecto al reanálisis (coef. pendiente = 0.12) (Figura 21-f).

Para la estación de Montería se observa una correlación moderada con los datos de ERA5 ($R = 0.513$). Se pueden notar valores muy dispersos respecto a la regresión lineal, y el coeficiente de la pendiente (0.52) indica que los datos observados crecen lentamente con respecto a los del reanálisis (Figura 21-h).

Los datos "contaminados" observados en la estación de Cartagena condicionan por completo su correlación con los datos de ERA5. Si bien en la gráfica se observan muchos valores cercanos al ajuste lineal, estos valores atípicos provocan que la correlación se vuelva pobre entre las dos fuentes de datos ($R = 0.153$). Además de esto, los valores observados crecen

aproximadamente a la mitad de rapidez con lo que lo hacen los valores de ERA5 (coef. pendiente = 0.51) (Figura 21-j).

Finalmente, los datos disponibles de velocidad del viento para la estación de Riohacha muestran una correlación relativamente moderada con respecto a los datos de ERA5 ($R = 0.739$). Los datos observados crecen más lentamente que los datos del reanálisis (coef. pendiente = 0.49) (Figura 21-l).



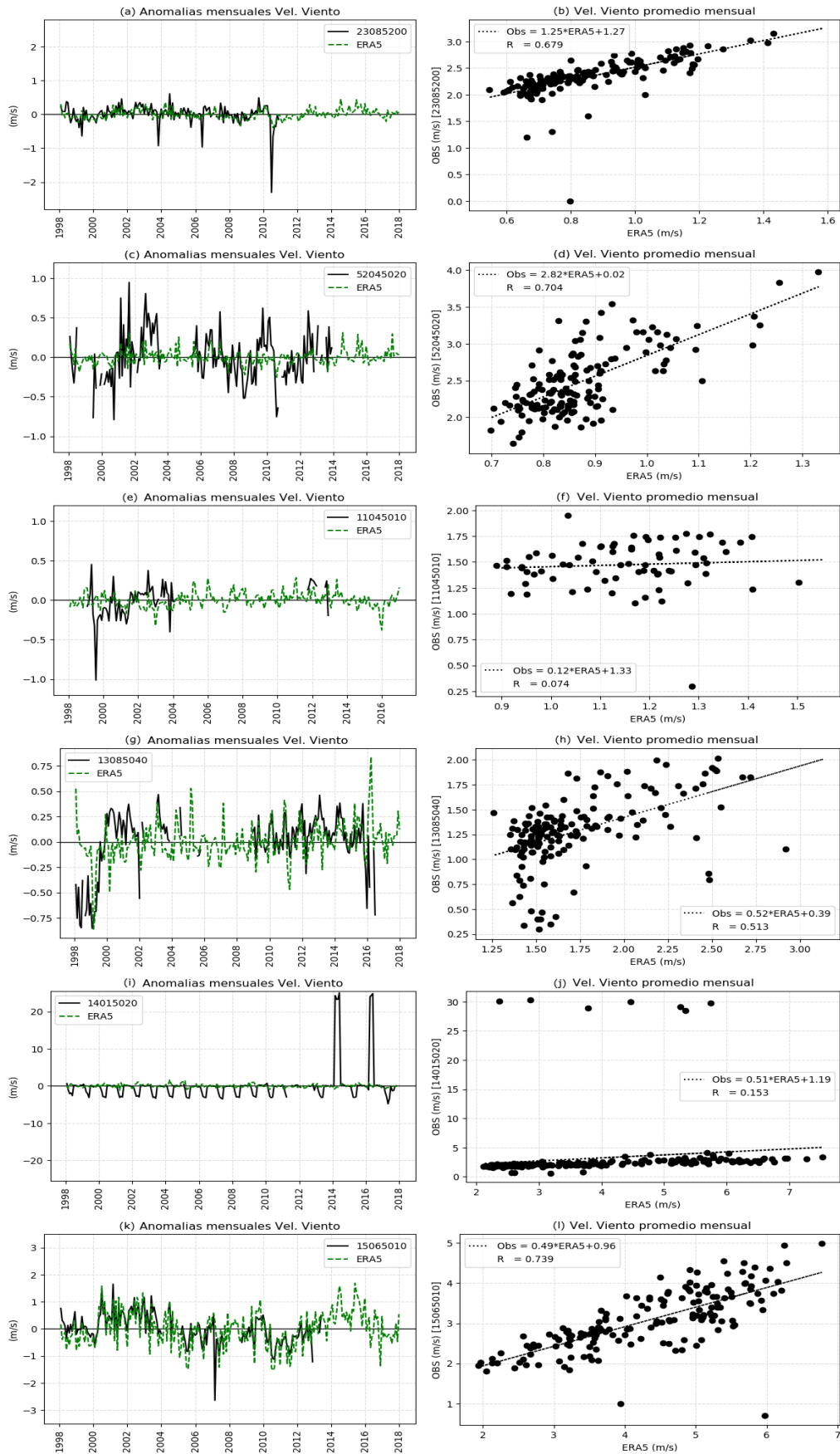


Figura 21. Anomalías mensuales de Vel. viento (izq.) y Scatterplot de Vel. viento OBS v ERA5 (der). **(a)-(b):** Rionegro [23085200], **(c)-(d):** Pasto [52045020], **(e)-(f):** Quibdó [11045010], **(g)-(h):** Montería [13085040], **(i)-(j):** Cartagena [14015020], **(k)-(l):** Riohacha [15065010].

Capítulo 4. Conclusiones

1. El lenguaje de programación Python, en conjunto con sus librerías especializadas con herramientas útiles para distintas áreas científicas (NumPy, pandas, SciPy, entre otros), provee funciones y módulos que permiten realizar un procesamiento ágil y detallado de datos climatológicos. Sus librerías para graficar datos también son flexibles e intuitivas.

El hecho de ser un lenguaje libre y ampliamente utilizado hace que Python y sus componentes estén en constante desarrollo y mejora por parte de sus mismos usuarios, además de que su uso global permite encontrar fácilmente documentación y guías para determinados propósitos que no son exclusivos de las ciencias atmosféricas (como el análisis de datos y de series de tiempo, por ejemplo).

2. Los datos faltantes en las estaciones de observación pueden provocar sesgos en los resultados posteriores, provocando que los análisis que se realicen no coincidan correctamente con el comportamiento climatológico real de los lugares de estudio. Siendo la velocidad del viento la variable con mayor cantidad de datos faltantes para todas las estaciones, la descripción correcta de su comportamiento en las distintas estaciones se vuelve compleja. En algunos casos, como en la estación de Rionegro, es posible reducir el período de interés de tal manera que las fechas con datos faltantes no se tengan en cuenta y tal vez se logre una mejor representación de la variable. En otros, esto no es posible debido a que los baches de datos perdidos se encuentran distribuidos a lo largo del período de tiempo.

Además de los datos faltantes, es importante realizar en estudios posteriores una revisión de la calidad de los datos observados, ya que como se vio en la estación de Cartagena con los vientos, pueden existir datos “contaminados” o irreales que son nocivos para el resto de la serie de tiempo y modifican por completo el detalle, la certeza y la calidad de las descripciones posteriores sobre la climatología del lugar.

3. La comparación entre los datos del IDEAM y los del reanálisis ERA5 permitió conocer las sub o sobrestimaciones por parte de ERA5 con respecto a las observaciones de las estaciones de medición. Se pudo notar al comparar los promedios mensuales cómo el reanálisis presenta sesgos de mayor o menor amplitud con respecto a las observaciones. En primer lugar estos sesgos entre ambos conjuntos de datos se pueden atribuir a la capacidad de representación por parte del modelo utilizado para ERA5. Pero se debe tener presente que las observaciones de las estaciones de medición también aportan información a este producto global (más probable aún en el caso de las estaciones elegidas, siendo estaciones principales de aeropuertos) y que la ausencia de mediciones o presencia de malas mediciones también pueden terminar modificando las estimaciones que realiza el reanálisis.

4. Al comparar la estacionalidad y variabilidad de las observaciones con las estimadas por ERA5 se pudo observar que, si bien el reanálisis no suele acertar en la magnitud de las variables, sí puede ser útil a la hora de estimar comportamientos estacionales y la variabilidad climática asociada a fenómenos como el ENSO, incluso para representar algunas tendencias crecientes/decrecientes de las variables con el paso del tiempo.

En lo adecuado de los estimados por ERA5 entra en juego el factor de representatividad por parte del reanálisis con respecto a puntos fijos como son las estaciones del IDEAM. Teniendo presente que ERA5 no cuenta con información para puntos únicos, sino que estima promedios espaciales para pequeños píxeles que dependen de la resolución utilizada (que en este caso es de cerca de 31km), tal vez se podrían obtener mejores resultados o por lo menos sesgos más pequeños realizando una espacialización de las variables para las observaciones del IDEAM, de modo que ambos conjuntos de datos representen valores promedio de áreas semejantes.

5. Evaluando la pertinencia de uso de los datos de ERA5 para representar variables climatológicas en Colombia, se concluye que, aunque para algunas variables y en algunos lugares del país sigue el mismo comportamiento que las observaciones, no es recomendable utilizar el reanálisis como única fuente de datos. Menos aún si no se plantea realizar correcciones en cuanto a las magnitudes de las variables.

Para algunas variables como temperatura o humedad relativa y en algunos lugares del país como Rionegro, Pasto o Riohacha, puede ser útil usar los datos de ERA5 para análisis climatológico siempre y cuando se plantee una metodología para corregir la magnitud de los valores y acercar las estimaciones del reanálisis lo más que se pueda a los datos reales, puesto que en cuanto a estacionalidad y variabilidad climática ERA5 representa adecuadamente el comportamiento de dichas variables en estas regiones particulares. Este no es necesariamente el caso para otros lugares, donde hay diferencias entre ERA5 y observaciones no solo en la magnitud (i.e. valores medios) sino además en la variabilidad (i.e. baja correlación), como es el caso de la humedad relativa para Quibdó o la temperatura para Cartagena.

Referencias Bibliográficas

Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology*, 200, 233-248.

World Meteorological Organization. (2011). *Guide to climatological practices*. Second edition (WMO-No. 100), Geneva.

Lin, J. W. B. (2012). Why Python is the next wave in earth sciences computing. *Bulletin of the American Meteorological Society*, 93(12), 1823-1824.

Hersbach, H., & Dee, D. (2016). ERA5 reanalysis is in production. *ECMWF newsletter*, 147(7).

World Meteorological Organization. (2017). *Guidelines on the Calculation of Climate Normals*. (WMO-No. 1203), Geneva.

Lin, J. W. B. (2012). *A Hands-On Introduction to Using Python in the Atmospheric and Oceanic Sciences*. (Self-published). Disponible en: <http://www.johnny-lin.com/pyintro/>

Parker, W. S. (2016). Reanalyses and observations: What's the difference? *Bulletin of the American Meteorological Society*, 97(9), 1565-1572.

Anexos

Anexo 1. Reporte: Revisión general de las variables atmosféricas obtenidas por la estación meteorológica del ITM - Campus Fraternidad.

Descripción

El propósito de este reporte es otorgar una visión general sobre el estado y comportamiento de las variables atmosféricas capturadas por la estación meteorológica del observatorio astronómico ITM – Campus Fraternidad, ubicado en el barrio Boston, zona oriental de la ciudad de Medellín.

La base de datos creada a partir de las mediciones realizadas por la estación meteorológica cuenta con mediciones aproximadamente cada cinco minutos, en un rango de fechas que inicia en 2016-03-10 18:00:51 y termina en 2018-08-29 15:25:09, para un total de 148688 datos disponibles para cada variable de interés. Es importante resaltar que casi la totalidad de los datos faltantes para la serie de tiempo de los datos recopilados no se encuentran explícitamente denotados en las misma, ya que al no aparecer las marcas de tiempo (índices de fecha y hora) las celdas que deberían representar valores perdidos (e.g NaN o -999) se omiten y se continúa hasta la próxima medición que contenga valores numéricos válidos.

Metodología

Luego de elegir las variables de interés (Temperatura, Precipitación, Presión, Humedad relativa y Velocidad de los vientos), se quería realizar un análisis general sobre el estado de los datos obtenidos por la estación meteorológica. Para dicho análisis se utilizó el lenguaje de programación Python (Spyder IDE) y algunas de sus librerías gratuitas. Inicialmente se realizó un remuestreo o *resampling* de los datos hacia escalas temporales particulares para analizar el comportamiento de las variables atmosféricas; estos remuestreos consisten en llevar las mediciones originales a valores promedio (acumulados en el caso de la lluvia) de las mismas en escala horaria, diaria y mensual.

Los resultados de cada *resampling* permiten conocer de manera detallada la cantidad de datos faltantes y su valor promedio/acumulado en dichas escalas de tiempo durante todo el rango delimitado por las mediciones en la base de datos.

Resultados y análisis

En general, la conversión de escala para todas las variables muestra una gran cantidad de datos faltantes sea en escala horaria, diaria o mensual. El porcentaje de datos faltantes con respecto a los datos totales que debería de tener cada variable medida oscila entre el 42% y el 44%, indicando que en la estación meteorológica se pierden (o no se están midiendo) una cantidad considerable de datos, lo que puede dificultar la realización de una descripción detallada sobre el comportamiento de la atmósfera en la zona. Revisando las variables una por una, se observa que algunos valores no son acordes a la climatología de Medellín. Particularmente, los valores promedios de presión atmosférica (994,857 hPa) y los valores acumulados de precipitación (3755,65 mm/año) se encuentran muy por encima de los valores

normalmente medidos en la ciudad (aprox. 850 hPa y 1750 mm/año respectivamente). (Tabla A1)

Esto puede indicar problemas en la calibración de dichas variables en la estación meteorológica, por lo que es recomendable revisar guías o manuales de la misma donde se indique cómo realizar una medición adecuada.

Tabla A1. Estado general de los datos de la estación meteorológica ITM - Campus Fraternidad

Variable	Resampling	#Total de datos	# Datos faltantes	% Datos faltantes	Promedio/Acumulado
Temperatura externa (°C)	Horario	21646	9289	42,913	21,356
	Diario	903	402	44,518	21,352
	Mensual	30	13	43,333	21,322
Precipitación (mm)	Horario	21646	9288	42,909	5,299
	Diario	903	402	44,518	130,473
	Mensual	30	13	43,333	3755,65
Presión (hPa)	Horario	21646	9289	42,913	994,857
	Diario	903	402	44,518	994,87
	Mensual	30	13	43,333	994,82
Humedad relativa externa (%)	Horario	21646	9289	42,913	72,47
	Diario	903	402	44,518	72,5
	Mensual	30	13	43,333	72,62
Velocidad del viento (km/h)	Horario	21646	9289	42,913	2,695
	Diario	903	402	44,518	2,696
	Mensual	30	13	43,333	2,773

Para la revisión también se realizaron gráficas de las variables de interés en el período dentro del rango de tiempo que cuenta con la mayor cantidad de datos (septiembre de 2017 a agosto de 2018) con el fin de examinar si su comportamiento es acorde a la climatología de Medellín. Particularmente, se graficó el ciclo diario de cada variable en cada uno de estos 12 meses y se observó de forma somera que los ciclos graficados son consistentes con la realidad climatológica de la ciudad y con las interacciones entre las variables, más allá de que en algunos casos los valores medidos se consideren erróneos. Se puede ver como la temperatura alcanza sus valores más altos en las tardes, acorde al lag o retraso que sufre normalmente la variable, mientras que la

humedad relativa y la presión decrecen en este mismo período de tiempo debido a su relación inversa con la temperatura. (Figura A1).

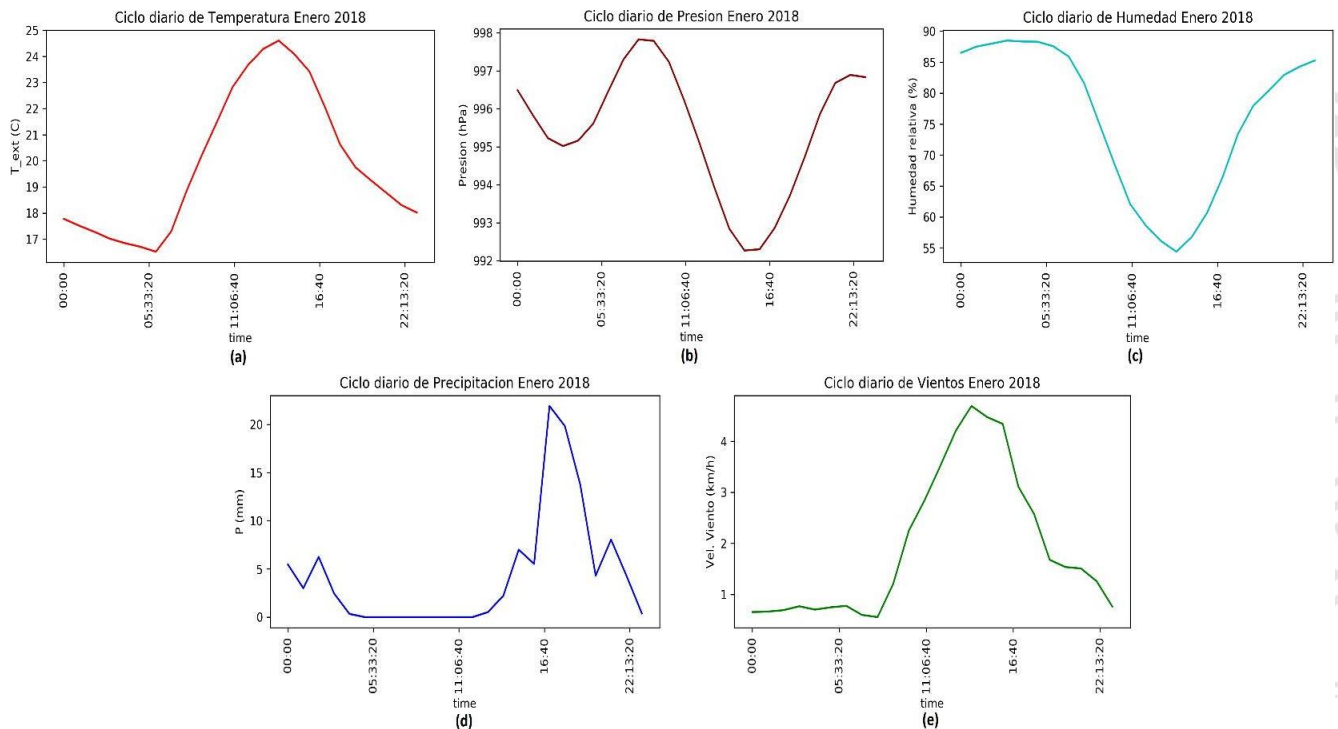


Figura A1. Ciclos diarios de enero 2018 para las variables de interés: (a) Temperatura, (b) Presión, (c) Humedad relativa, (d) Precipitación y (e) Vel. Viento.

En las siguientes páginas se encuentran gráficas con ejemplos del comportamiento de las variables a diferentes escalas de tiempo, desde los datos tomados cada 5 minutos, pasando por los promedios horarios y diarios.

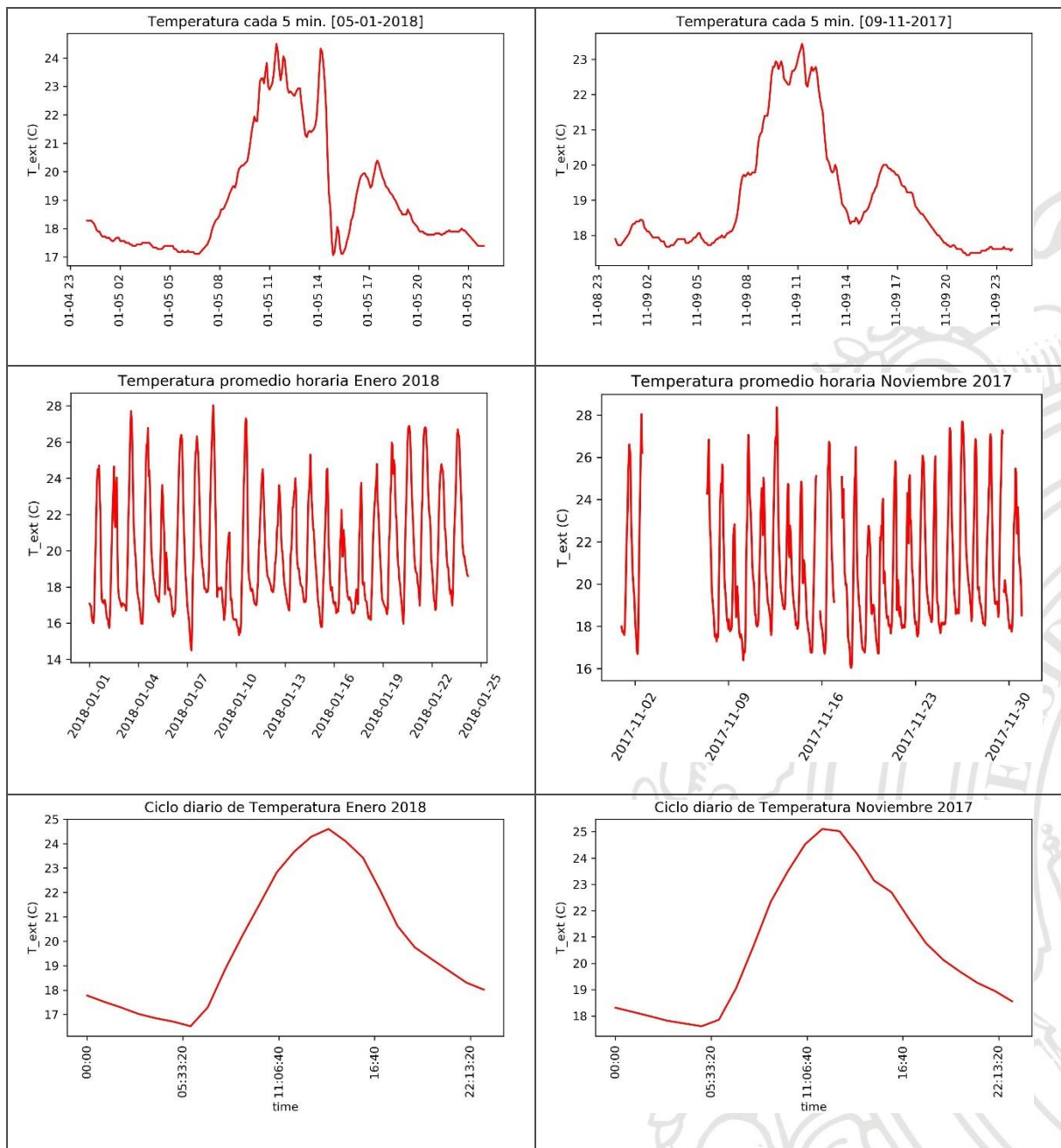


Figura A2. Series de tiempo de **Temperatura** para **enero-2018 (izquierda)** y **noviembre-2017 (derecha)**. La **primera fila** muestra los datos tomados cada 5 minutos por la estación para un día de cada uno de los meses. La **segunda fila** muestra los promedios horarios para todos los días del mes correspondiente; en noviembre pueden notarse baches, correspondiendo a tiempos con valores perdidos. La **tercera fila** muestra el ciclo diario promedio para cada mes (i.e. El resultado de promediar las temperaturas de las 3 pm de todos los días de enero, etc.)

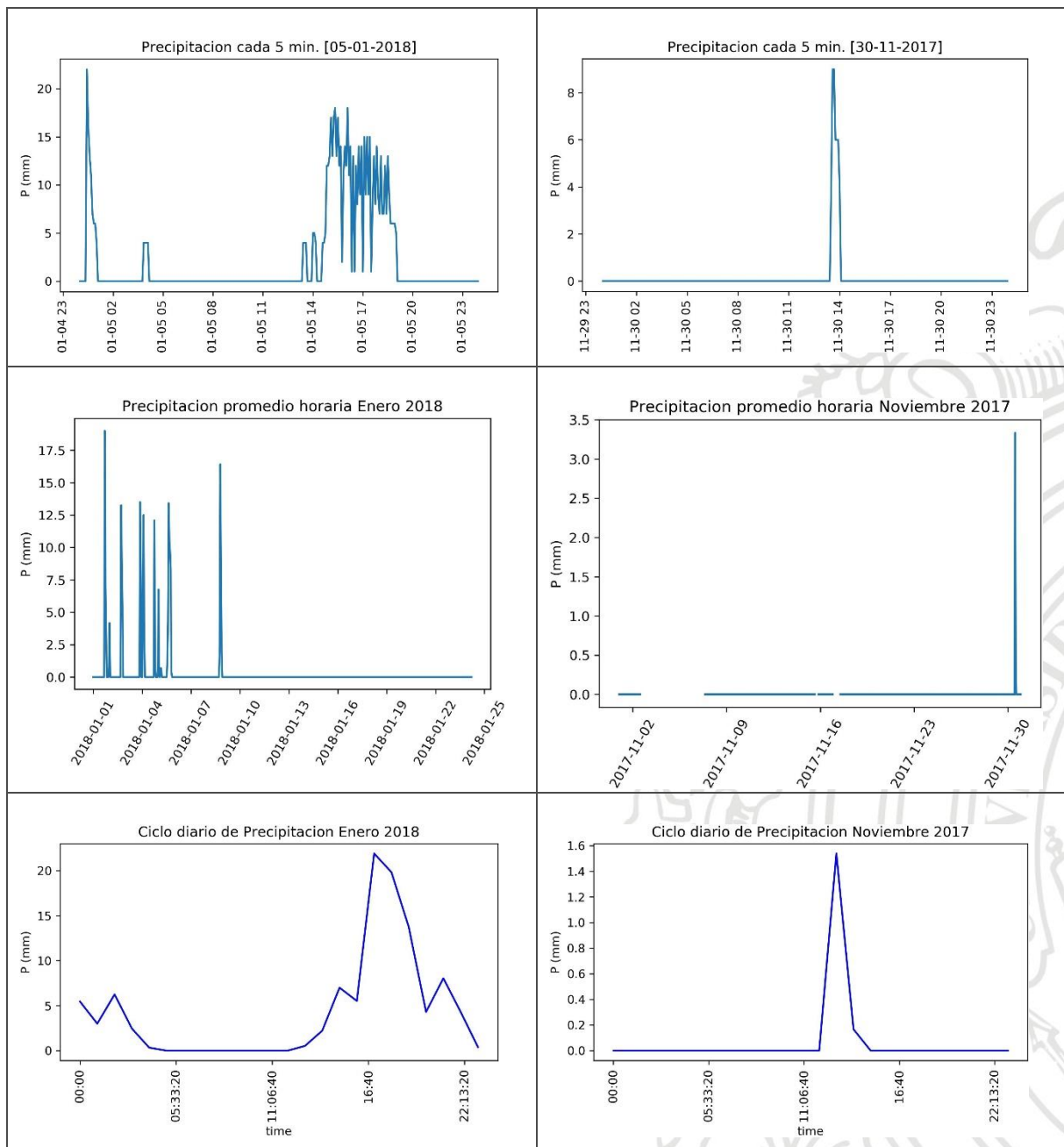


Figura A3. Series de tiempo de **Precipitación** para **enero-2018 (izquierda)** y **noviembre- 2017 (derecha)**. La **primera fila** muestra los datos tomados cada 5 minutos por la estación para un día de cada uno de los meses. La **segunda fila** muestra los promedios horarios para todos los días del mes correspondiente; en noviembre pueden notarse baches, correspondiendo a tiempos con valores perdidos. La **tercera fila** muestra el ciclo diario promedio para cada mes (i.e. El resultado de promediar las temperaturas de las 3 pm de todos los días de enero, etc.). En enero-2018 la estación registró varios días lluviosos, mientras que en noviembre-2017 solo registró uno. Note que en ambos meses las lluvias tienen un pico en la tarde.

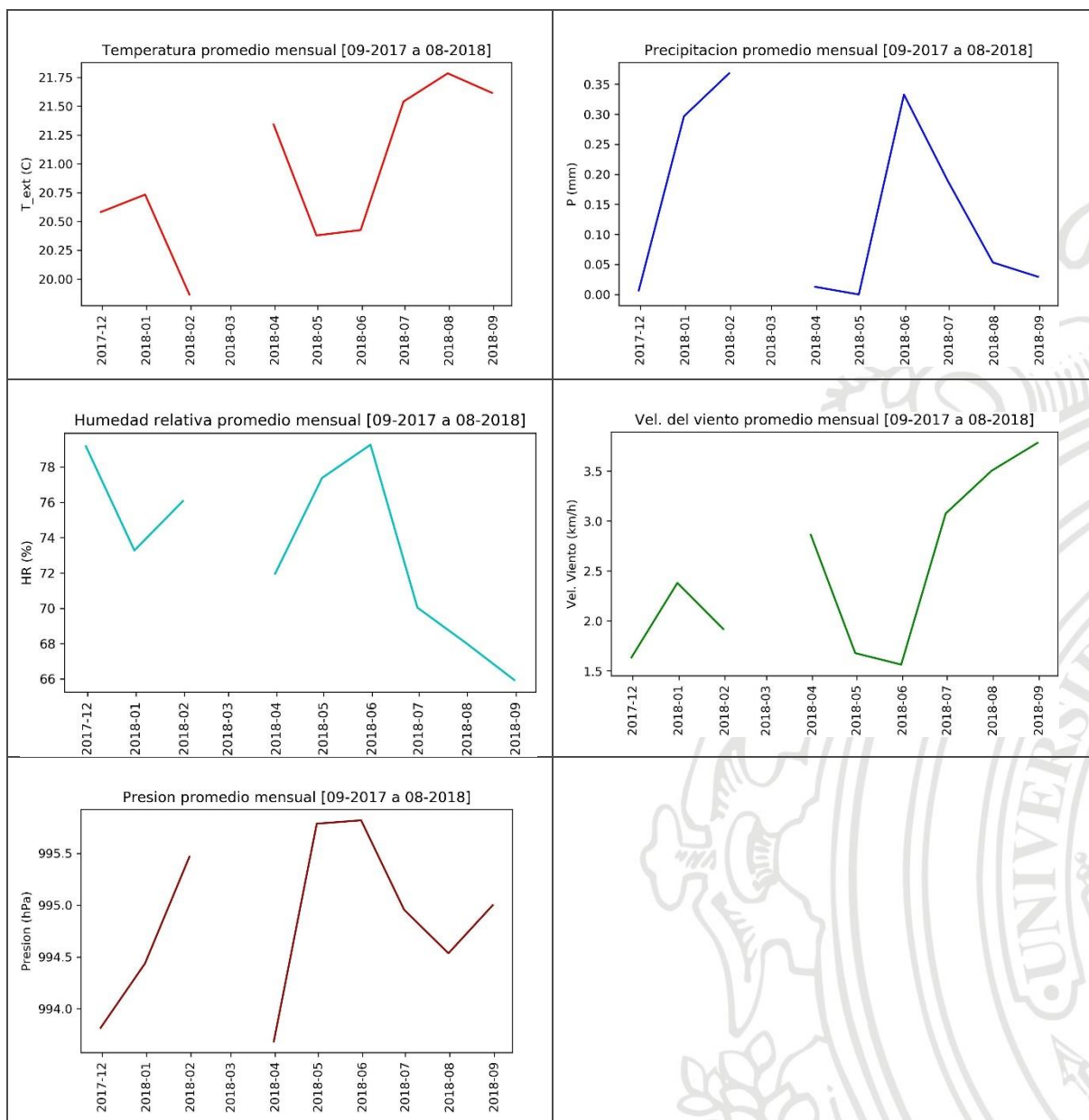


Figura A4. Promedios mensuales de las diferentes variables para el periodo 2017-12 a 2018-09. En este intervalo los valores perdidos en 2018-03 fueron tantos que no se calcularon valores promedio para dicho mes. Los valores de precipitación corresponden a promedios horarios. Los valores de presión deben ser revisados ya que el valor medio reportado acá no corresponde con los valores típicos para Medellín.