



UNIVERSIDAD  
DE GRANADA



---

Facultad de Ciencias

## MÁSTER EN GEOFÍSICA Y METEOROLOGÍA

TRABAJO FIN DE MÁSTER

# Análisis del sesgo muestral en medidas micrometeorológicas

Presentado por:

Paloma Rodríguez Martín

Profesores:

Dr. Andrew S. Kowalski

Dra. Matilde M del Valle García -Valdecasas Ojeda

Curso académico 2023/2024

## ABSTRACT

This research is based on Charles' law (for gases at constant pressure), which shows that the sampling of data is not random but depends on weighting factors. Therefore, we look for the error (bias) that occurs when calculations are performed arithmetically without taking into account the weighting factors.

In this case, the samples are micrometeorological data collected during the month of July from Eddy Covariance towers located in four different ecosystems. Statistical moments are calculated with and without weighting factors on these approximately one billion data points in order to provide a comprehensive analysis of the results.

It is shown that there is indeed an overestimation of the calculations when performed arithmetically.

Esta investigación se basa en la Ley de Charles (ley de los gases a presión constante), donde se confirma, que el muestreo de los datos no es aleatorio, depende de factores de ponderación. Por ello, se busca el error (sesgo) que se produce al realizar cálculos de forma aritmética sin tener en cuenta los factores de ponderación.

En este caso, las muestras son datos micrometeorológicos recogidos en el mes de julio mediante torres Eddy Covariance situadas en cuatro ecosistemas diferentes. Con estos, aproximadamente mil millones de datos, se realizan estadísticos de momento con y sin factores de ponderación con el fin de realizar un análisis exhaustivo de los resultados.

Se muestra, que realmente, hay una sobreestimación de los cálculos al realizarse de forma aritmética.

## LISTA DE ACRÓNIMOS Y SÍMBOLOS

Símbolo	Significado
$\mu$	Media
$\mu_s$	Media sesgada (media aritmética)
$\mu_{s,T}$	Media sesgada de la temperatura (media aritmética de la temperatura)
$\mu_{s,u}$	Media sesgada de la velocidad longitudinal (media aritmética de la velocidad longitudinal)
$\mu_{s,v}$	Media sesgada de la velocidad latitudinal (media aritmética de la velocidad latitudinal)
$\mu_{s,w}$	Media sesgada de la velocidad vertical (media aritmética de la velocidad vertical)
$\mu_{s,q}$	Media sesgada del vapor de agua (media aritmética del vapor de agua)
$\mu_{s,CO_2}$	Media sesgada del $CO_2$ (media aritmética del $CO_2$ )
$\mu_c$	Media corregida (media ponderada)
$\mu_{c,T}$	Media corregida de la temperatura (media ponderada de la temperatura)
$\mu_{c,u}$	Media corregida de la velocidad longitudinal (media ponderada de la velocidad longitudinal)
$\mu_{c,v}$	Media corregida de la velocidad latitudinal (media ponderada de la velocidad latitudinal)
$\mu_{c,w}$	Media corregida de la velocidad vertical (media ponderada de la velocidad vertical)
$\mu_{c,q}$	Media corregida del vapor de agua (media ponderada del vapor de agua)
$\mu_{c,CO_2}$	Media corregida del $CO_2$ (media ponderada del $CO_2$ )
$\sigma^2$	Varianza
$\sigma_s^2$	Varianza sesgada (varianza aritmética)
$\sigma_c^2$	Varianza corregida (varianza ponderada)
$\sigma$	Desviación estándar
$\sigma_s$	Desviación estándar sesgada (desviación estándar aritmética)
$\sigma_{s,T}$	Desviación estándar sesgada de la temperatura (desviación estándar aritmética de la temperatura)
$\sigma_{s,u}$	Desviación estándar sesgada de la velocidad longitudinal (desviación estándar aritmética de la velocidad longitudinal)
$\sigma_{s,v}$	Desviación estándar sesgada de la velocidad latitudinal (desviación estándar aritmética de la velocidad latitudinal)
$\sigma_{s,w}$	Desviación estándar sesgada de la velocidad vertical (desviación estándar aritmética de la velocidad vertical)
$\sigma_{s,q}$	Desviación estándar sesgada del vapor de agua (desviación estándar aritmética del vapor de agua)
$\sigma_{s,CO_2}$	Desviación estándar sesgada del $CO_2$ (desviación estándar aritmética del $CO_2$ )
$\sigma_c$	Desviación estándar corregida (desviación estándar ponderada)
$\sigma_{c,T}$	Desviación estándar corregida de la temperatura (desviación estándar ponderada de la temperatura)
$\sigma_{c,u}$	Desviación estándar corregida de la velocidad longitudinal (desviación estándar ponderada de la velocidad longitudinal)
$\sigma_{c,v}$	Desviación estándar corregida de la velocidad latitudinal (desviación estándar ponderada de la velocidad latitudinal)
$\sigma_{c,w}$	Desviación estándar corregida de la velocidad vertical (desviación estándar ponderada de la velocidad vertical)
$\sigma_{c,q}$	Desviación estándar corregida del vapor de agua (desviación estándar ponderada del vapor de agua)
$\sigma_{c,CO_2}$	Desviación estándar corregida del $CO_2$ (desviación estándar ponderada del $CO_2$ )
$E$	Asimetría o skewness
$E_s$	Skewness sesgado (skewness aritmético)
$E_{s,T}$	Skewness sesgado de la temperatura (skewness aritmético de la temperatura)
$E_{s,u}$	Skewness sesgado de la velocidad longitudinal (skewness aritmético de la velocidad longitudinal)
$E_{s,v}$	Skewness sesgado de la velocidad latitudinal (skewness aritmético de la velocidad latitudinal)

<b>Símbolo</b>	<b>Significado</b>
$E_{s,w}$	Skewness sesgado de la velocidad vertical (skewness aritmético de la velocidad vertical)
$E_{s,q}$	Skewness sesgado del vapor de agua (skewness aritmético del vapor de agua)
$E_{s,CO_2}$	Skewness sesgado del $CO_2$ (skewness aritmético del $CO_2$ )
$E_c$	Skewness corregido (skewness ponderado)
$E_{c,T}$	Skewness corregido de la temperatura (skewness ponderado de la temperatura)
$E_{c,u}$	Skewness corregido de la velocidad longitudinal (skewness ponderado de la velocidad longitudinal)
$E_{c,v}$	Skewness corregido de la velocidad latitudinal (skewness ponderado de la velocidad latitudinal)
$E_{c,w}$	Skewness corregido de la velocidad vertical (skewness ponderado de la velocidad vertical)
$E_{c,q}$	Skewness corregido del vapor de agua (skewness ponderado del vapor de agua)
$E_{c,CO_2}$	Skewness corregido del $CO_2$ (skewness ponderado del $CO_2$ )
K	Apuntamiento, curtosis o kurtosis
$K_s$	Kurtosis sesgado (kurtosis aritmético)
$K_{s,T}$	Kurtosis sesgado de la temperatura (kurtosis aritmético de la temperatura)
$K_{s,u}$	Kurtosis sesgado de la velocidad longitudinal (kurtosis aritmético de la velocidad longitudinal)
$K_{s,v}$	Kurtosis sesgado de la velocidad latitudinal (kurtosis aritmético de la velocidad latitudinal)
$K_{s,w}$	Kurtosis sesgado de la velocidad vertical (kurtosis aritmético de la velocidad vertical)
$K_{s,q}$	Kurtosis sesgado del vapor de agua (kurtosis aritmético del vapor de agua)
$K_{s,CO_2}$	Kurtosis sesgado del $CO_2$ (kurtosis aritmético del $CO_2$ )
$K_c$	Kurtosis corregido (kurtosis ponderado)
$K_{c,T}$	Kurtosis corregido de la temperatura (kurtosis ponderado de la temperatura)
$K_{c,u}$	Kurtosis corregido de la velocidad longitudinal (kurtosis ponderado de la velocidad longitudinal)
$K_{c,v}$	Kurtosis corregido de la velocidad latitudinal (kurtosis ponderado de la velocidad latitudinal)
$K_{c,w}$	Kurtosis corregido de la velocidad vertical (kurtosis ponderado de la velocidad vertical)
$K_{c,q}$	Kurtosis corregido del vapor de agua (kurtosis ponderado del vapor de agua)
$K_{c,CO_2}$	Kurtosis corregido del $CO_2$ (kurtosis ponderado del $CO_2$ )
$x_i$	Datos micrometeorológicos tomados por torres Eddy covariance ( $x_1, x_2, x_3 \dots x_N$ )
$T_i$	Datos concretos de temperatura tomados por torres Eddy covariance ( $T_1, T_2, T_3 \dots T_N$ )
$u_i$	Datos concretos de velocidad longitudinal tomados por torres Eddy covariance ( $u_1, u_2, u_3 \dots u_N$ )
$v_i$	Datos concretos de velocidad latitudinal tomados por torres Eddy covariance ( $v_1, v_2, v_3 \dots v_N$ )
$w_i$	Datos concretos de velocidad vertical tomados por torres Eddy covariance ( $w_1, w_2, w_3 \dots w_N$ )
$c_i$	Datos concretos de densidad de $CO_2$ tomados por torres Eddy covariance ( $c_1, c_2, c_3 \dots c_N$ )
$q_i$	Datos concretos de densidad de $H_2O$ tomados por torres Eddy covariance ( $q_1, q_2, q_3 \dots q_N$ )
$\tilde{T}$	Temperatura ponderada
$\tilde{u}$	Velocidad del viento longitudinal ponderado
$\tilde{v}$	Velocidad del viento latitudinal ponderado
$\tilde{w}$	Velocidad del viento vertical ponderado
$\tilde{c}$	Densidad del $CO_2$ ponderado
$\tilde{q}$	Densidad del $H_2O$ ponderado
$C_{p,i}$	Calor específico
$\rho_i$	Densidad del aire seco
N	Número de datos
EC	Eddy Covariance

## ÍNDICE

<b>1. INTRODUCCIÓN Y ANTECEDENTES.....</b>	<b>páginas 1-10</b>
1.1 <i>Introducción.....</i>	<i>páginas 1-2</i>
1.2 <i>Datos micrometeorológicos.....</i>	<i>página 2</i>
1.3 <i>Estadísticos de momento .....</i>	<i>páginas 3-6</i>
1.3.1 <i>Estadísticos de momento sesgados (aritméticos).....</i>	<i>páginas 4-6</i>
1.3.1.1 <i>Media (primer momento).....</i>	<i>página 4</i>
1.3.1.2 <i>Desviación típica (segundo momento).....</i>	<i>páginas 4-5</i>
1.3.1.3 <i>Asimetría (tercer momento).....</i>	<i>página 5</i>
1.3.1.4 <i>Curtosis (cuarto momento).....</i>	<i>página 6</i>
1.3.2 <i>Estadísticos de momento corregidos (ponderados)....</i>	<i>páginas 7-8</i>
1.3.2.1 <i>Media (primer momento).....</i>	<i>página 7</i>
1.3.2.2 <i>Desviación típica (segundo momento).....</i>	<i>página 7</i>
1.3.2.3 <i>Asimetría (tercer momento).....</i>	<i>página 7</i>
1.3.2.4 <i>Curtosis (cuarto momento).....</i>	<i>página 8</i>
1.4 <i>Revisión bibliográfica de los estadísticos de momento .....</i>	<i>páginas 8-10</i>
<b>2. JUSTIFICACIÓN DEL ESTUDIO y OBJETIVO.....</b>	<b>página 10</b>
<b>3. METODOLOGÍA.....</b>	<b>página 11-14</b>
3.1 <i>Lugar de estudio.....</i>	<i>páginas 11-12</i>
3.1.1 <i>Padul.....</i>	<i>página 11</i>
3.1.2 <i>Balsa Blanca.....</i>	<i>página 11</i>
3.1.3 <i>Gádor.....</i>	<i>página 12</i>
3.1.4 <i>Olivar de Úbeda.....</i>	<i>página 12</i>
3.2 <i>Toma de datos.....</i>	<i>página 13</i>
3.3 <i>Diseño experimental.....</i>	<i>páginas 13-14</i>
<b>4. RESULTADOS Y ANÁLISIS.....</b>	<b>páginas 15-25</b>
4.1 <i>Sesgo de la media.....</i>	<i>páginas 15-18</i>
4.2 <i>Sesgo de la desviación típica.....</i>	<i>páginas 18-21</i>
4.3 <i>Sesgo en la asimetría .....</i>	<i>páginas 21-24</i>
4.4 <i>Sesgo de la curtosis .....</i>	<i>páginas 24-25</i>
<b>5. CONCLUSIONES.....</b>	<b>página 26</b>
<b>6. AGRADECIMIENTOS.....</b>	<b>página 27</b>
<b>7. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>páginas 28-30</b>
<b>ANEXO I. FORMULARIO</b>	
<b>ANEXO II. UBICACION DE LAS ZONAS DE ESTUDIO</b>	
<b>ANEXO III. EJEMPLO DE DESARROLLO DE SCRIPT DE MATLAB</b>	

# 1. INTRODUCCIÓN Y ANTECEDENTES

## 1.1 Introducción

La micrometeorología es la rama de la meteorología que estudia la capa límite para poder comprender los procesos de formación de los fenómenos atmosféricos. En la toma de datos se produce, sin embargo, un problema común conocido como sesgo de muestreo que se descubre al analizar datos en este campo. El sesgo de muestreo es el término utilizado para describir el error que puede aparecer en las mediciones, como resultado de una mala selección de la muestra.

El método de cálculo, la técnica de muestreo, la variabilidad temporal y espacial son solo algunos de los factores que pueden contribuir a este error. Para obtener resultados confiables al estudiar datos micrometeorológicos, es necesario abordar y comprender este sesgo de muestreo. Para evitar distorsiones en los datos y llegar a una comprensión más precisa de la dinámica atmosférica, es necesario contar con un método adecuado para seleccionar la muestra y controlar estos factores (Hollinger y Richardson, 2005).

Es un desafío extremadamente difícil ya que, en un sistema termodinámico aunque esté compuesto por paquetes de aire seco, a igual masa y presión y térmicamente diferentes (ejemplo  $-20^{\circ}\text{C}$  y  $+20^{\circ}\text{C}$ ), el volumen cálido agregado es mayor según la Ley de Charles. La ley de Charles (ley de los gases a presión constante o la ley de Charles y Gay-Lussac) establece que, a presión constante, el volumen de una cantidad fija de gas es directamente proporcional a su temperatura absoluta. Es decir, hay una mayor probabilidad de muestrear aire seco caliente que aire seco frío. Esto conlleva a una sobreestimación en este caso de la temperatura y además, asimilar que las velocidades opuestas de los paquetes cálidos y fríos suman cero en sus momentos (asimilando velocidad nula del conjunto), lo que implica un sesgo de muestro ya que nos daría una velocidad media ascendente errónea (Kowalski y García-Valdecasas Ojeda, 2022).

En el caso del aire seco, al ser considerado como un modelo de gas ideal, las moléculas se consideran puntos que no tienen ni tamaño ni fuerza de atracción entre ellas. Por lo que, en este caso, se puede considerar que el volumen depende únicamente de la temperatura ya que, no hay más componentes que afecten al comportamiento.

Por el contrario, el aire húmedo (mezcla de aire seco y agua, por tanto más número de moléculas) ocupa un volumen ligeramente menor a la misma temperatura y presión. Esta deducción procede de la ley de los gases ideales, donde se establece que el producto de la presión (P) y el volumen (V) de un gas es proporcional a la cantidad de sustancia (n) multiplicada por la constante de los gases ideales (R) y la temperatura absoluta (T) (Río e Iza, 2007).

$$PV=nRT \quad [\text{Ec.1}]$$

Hay una mayor probabilidad de muestrear aire caliente que aire frío, por tanto, el muestreo de datos realmente no es aleatorio.

Este análisis se enfoca en analizar el sesgo de los cálculos estadísticos que se producen en los datos micrometeorológicos obtenidos a través de torres Eddy covariance al no tener en cuenta la diferencia de propiedades del aire con respecto a la temperatura, vapor de agua,  $CO_2$ , velocidad longitudinal, latitudinal y vertical.

Al incorporar factores de ponderación de acuerdo con las leyes de conservación física, se propone un nuevo conjunto de reglas de promedio que calcula con precisión las medias de los parámetros mencionados del aire. El formato de desarrollo de este estudio comenzará con los métodos estadísticos aritméticos y a continuación con sus ponderaciones para poder realizar una comparativa más idónea entre ellos y el sesgo cometido.

## 1.2 Datos micrometeorológicos

En esta investigación se buscará si existe un sesgo debido a la falta de ponderación en seis parámetros micrometeorológicos:

- Temperatura (T): conocida como la energía cinética de las partículas que contienen esa sustancia u objeto. En micrometeorología se enfoca en mediciones a pequeña escala, por tanto, estas medidas son tomadas a menudo a nivel del suelo o como en el caso de este estudio, en las capas más bajas de la atmósfera. Su medida más habitual es en grado Celsius ( $^{\circ}C$ ), en el que será expresado (Tipler y Mosca, 2021).
- Direcciones del aire (u, v, w): definidas como las velocidades (rapidez de movimiento) que toma el viento en diferentes direcciones de la atmósfera.
  - La velocidad longitudinal (u) es la que toma el aire en dirección este-oeste, se relaciona con los vientos zonales, paralelos a las líneas de longitud en la Tierra.
  - La velocidad latitudinal (v) hace referencia a los vientos meridionales que se mueven en las líneas de latitud de la Tierra (dirección norte-sur).
  - Por último, la velocidad vertical (w), permite entender la distribución vertical de humedad y temperatura, así como los procesos de convección y los flujos ya que hace referencia al movimiento del aire en vertical (ascendente o descendente con respecto a la atmósfera).

Habitualmente, como en este estudio, las velocidad se expresan en metros por segundo (m/s) (Vinnichenko, 2013)(Yeh, 1949).

- Densidad del dióxido de carbono ( $CO_2$ ): es la masa de dióxido de carbono tomado por la torre Eddy covariance en un volumen medido, en este caso, se expresa como miligramos de  $CO_2$  en un metro cúbico de aire ( $mg/m^3$ ).
- Densidad de vapor de agua (q): es la concentración de vapor de agua en el aire. Es decir, es la masa vapor de agua que tiene un volumen de aire. La densidad de vapor está altamente relacionada con la temperatura y presión atmosférica, ya que al aumentar la temperatura, el aire puede contener mayor vapor de agua resultando en una mayor densidad. Es de gran importancia en el meteorología ya que influye en grandes procesos como la formación de nubes y precipitación entre otros. Se medirá en gramos de vapor de agua por metro cúbico de aire ( $g/m^3$ ).

### 1.3 Estadísticos de momento.

Los métodos estadísticos se presentan como una ciencia de adecuado análisis e interpretación a partir de datos experimentales. Además, estos métodos se utilizan a menudo para diseñar los experimentos más relevantes con un propósito específico (Checa-García, 2017). La estadística se define como la ciencia y el arte de recopilar, resumir y analizar datos basados en fluctuaciones aleatorias. Se compone de dos amplias categorías de resumen y análisis de datos conocidas como estadísticas descriptivas e inferenciales (Nick, 2007). En primera instancia, el investigador tiene como objetivo obtener un resumen conciso y comprensible de los elementos esenciales en términos de práctica de una colección de datos (habitualmente una muestra, pero ocasionalmente datos de poblaciones completas).

En la estadística inferencial se usa una colección de técnicas más sofisticadas diseñadas para permitir poder deducir y proponer evidencias científicas de una población (origen) a través de una muestra concreta (Pérez, 2016). Este estudio se centra en cuatro parámetros estadísticos:

- Medidas de tendencia central o de centralización es un valor que caracteriza a todo el conjunto de datos como una sola medida. Las tres medidas básicas de tendencia central son la media, la mediana y la moda. La media es la suma de los valores de un conjunto de datos dividida por el número total de observaciones. La mediana es el valor central de la distribución cuando los datos se ordenan de mayor a menor (o viceversa). La moda es el valor más común en el conjunto de datos.
- Medidas de posición: que nos indica la posición que ocupa un valor en el conjunto de los datos. Las tres formas principales son: percentiles, cuartiles y deciles.
- Medidas de dispersión/variación: son medidas de tendencia central que proporcionan información importante para describir los datos, pero no puede capturar la variabilidad del conjunto de datos. Una medida de dispersión/varianza describe el grado en que las variables son similares o diferentes. La varianza y la desviación estándar son medidas de dispersión que muestran qué tan cerca está cada valor observado de la media de todo el conjunto de datos (Kaur et al., 2018).
- Medidas de forma: nos dan información sobre la distribución que forman los datos y cuanto se parece a una distribución normal. Las medidas de forma son la simetría y la curtosis (Franzese y Juliano, 2019).

Este estudio se centrará en los *estadísticos de momento* ya que utilizan todos los datos de la distribución y aportan información de la variabilidad, permitiendo minimizar el sesgo producido en los cálculos. Además, reducen su complicación al tratar grandes cantidades de datos al basarse en sumas, sumas de cuadrados, sumas de cubos y sumas a la cuarta potencia. Aunque, la facilidad del cálculo se ve contrarrestada con que son índices matemáticos de difícil interpretación práctica.

Estos *estadísticos de momento* son: momento de orden 1 (media), momento de orden 2 respecto a la media (desviación típica), momento de orden 3 (asimetría), momento de orden 4 (apuntamiento o curtosis) (Pérez, 2016).



### 1.3.1 Estadísticos de momento sesgados (aritméticos)

Los cálculos aquí denominados aritméticos son las operaciones matemáticas tal cual las conocemos que se calculan sin tener en cuenta factores de ponderación. Son las operaciones que comúnmente se utilizan en todos los resultados para simplificarlos y poder realizar una fácil interpretación. En este apartado se establecen cada uno de los estadísticos de momento, enfocándolos en el cálculo aritmético de las medidas meteorológicas que serán analizadas, para hacer más sencilla la comprensión a la hora de analizar los resultados.

#### 1.3.1.1 Media (primer momento)

El tratamiento estadístico y matemático del promedio es fundamental para la ciencia atmosférica y subyace en muchas de las leyes fundamentales de la meteorología. Incluso a la escala más pequeña, las variables meteorológicas y de flujo suelen implicar alguna forma de promediación, como la realizada sobre volúmenes instrumentados o celdas de modelos de cuadrícula (Kowalski, 2012). Con mucho, es la medida de tendencia central más utilizada en datos continuos.

La media aritmética es la suma de todos los valores de la variable, dividida por N, el número de observaciones. La media se denomina informalmente promedio, pero este término puede ser ambiguo y debería ser evitado (Johnson, 2014). Por tanto la media aritmética ( $\mu_s$ ) se obtiene como:

$$\mu_s = \frac{1}{N} \sum_{i=1}^N x_i \quad [\text{Ec.2}]$$

Donde  $x_i$  son los datos de cualquier medida, en concreto para esta investigación son datos micrometeorológicos tomados por torres Eddy covariance ( $x_1, x_2, x_3 \dots x_N$ ); temperatura ( $T_i$ ), velocidad longitudinal ( $u_i$ ), velocidad latitudinal ( $v_i$ ), velocidad vertical ( $w_i$ ), vapor de agua ( $q_i$ ) y  $CO_2$  ( $c_i$ ).

La media aritmética se usa en prácticamente todos los análisis estadísticos. En general, en cualquier tipo de estudio, sea del ámbito que sea, sirve para simplificar los resultados.

#### 1.3.1.2 Desviación estándar (segundo momento)

La varianza es una medida de variabilidad (o dispersión) que en meteorología y en otras áreas permite evaluar la dispersión de los datos y determinar la cantidad de variabilidad. En este estudio, se expresa como  $\sigma_s^2$  para hacer referencia a la varianza calculada aritméticamente de los datos micrometeorológicos. Se define como la desviación cuadrática media de la media, es decir, la media de los cuadrados de las diferencias entre los valores individuales de  $x_i$  y el valor medio  $\mu_s$  (calculado en el apartado anterior):

$$\sigma_s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_s)^2 \quad [\text{Ec.3}]$$

Para el cálculo de la desviación estándar de forma aritmética ( $\sigma_s$ ) simplemente se haría la raíz cuadrada de la varianza:

$$\sigma_s = \sqrt{\sigma_s^2} \quad [\text{Ec.4}]$$

La desviación nos permite saber cuánto de alejados se encuentran los valores con respecto a la media, y por tanto, nos da una estimación de cómo es esa muestra. Se suele representar como en la *figura 1*, para poder visualizar de forma sencilla la caracterización de los datos y poder saber si se encuentran cercanos a la media o tienen una distribución más alejada y por tanto, con un rango de máximo y mínimos más amplio (American Meteorological Society, 2015).

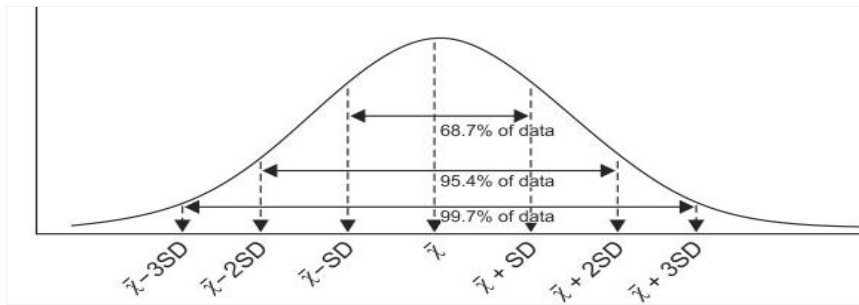


Figura 1. Representación de la desviación típica. Fuente: Lee et al., (2015).

### 1.3.1.3 Asimetría (tercer momento)

El tercer momento corresponde a la asimetría (del inglés skewness), permite saber el grado de distribución de los datos conforme a un centro, en este caso de la media. Como se observa en la *figura 2*, si sabemos dónde se encuentra la media, mediana y moda en la distribución de los datos, se puede saber la asimetría de estos. Una distribución es simétrica si media, moda y mediana coinciden en un mismo punto. La Ec.5 muestra una forma de calcular la asimetría (de las muchas existentes) llamada coeficiente de Fisher:

$$E_s = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_s)^3}{(\sigma_s)^3} \quad [\text{Ec.5}]$$

Donde  $E_s$  es el coeficiente de Fisher y su valor nos da una idea de hacia dónde se localizan las colas de la distribución. Si el resultado da mayor a cero ( $E_s > 0$ ) es una distribución asimétrica positiva, si es menor que cero ( $E_s < 0$ ) es una distribución asimétrica negativa y por último, si es igual a cero ( $E_s = 0$ ), es simétrica (Doane y Seward, 2011).

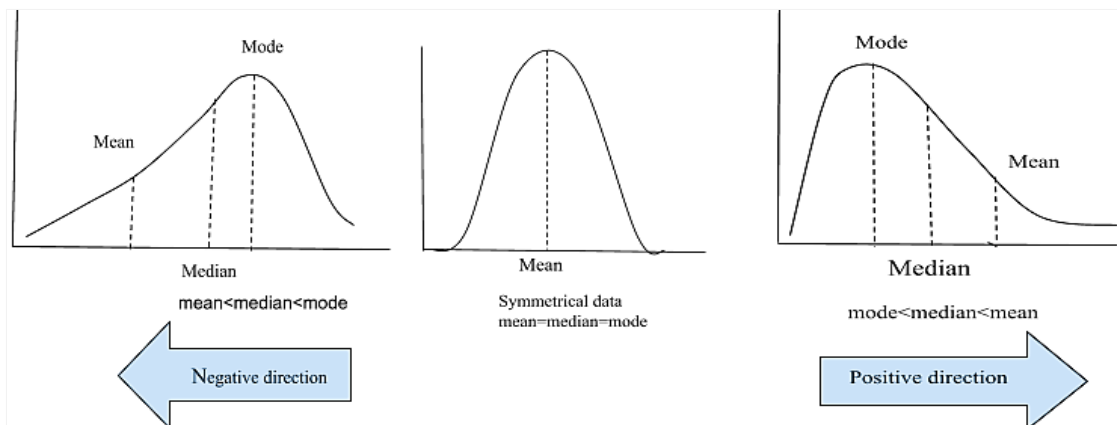


Figura 2. Posición general de la media, la mediana y la moda al representar la asimetría. Fuente: Arnold y Groeneveld, 1995.

### 1.3.1.4 Curtosis (cuarto momento)

Una medida estadística que describe la forma de una distribución de datos en un conjunto de datos es la curtosis (en inglés kurtosis). Explica la prominencia de la distribución en relación con la distribución normal y sus colas. La curtosis se emplea con frecuencia como una métrica para determinar cuánto se desvía una distribución de la normalidad. Para comparar la distribución con una distribución normal, que tiene una curtosis de 0, se utiliza este método estadístico. El cuarto momento estandarizado, que se calcula de la siguiente manera [Ec.6], es una fórmula utilizada con frecuencia para la curtosis a pesar de que existen muchas otras definiciones y fórmulas para el fenómeno:

$$K_S = \frac{\frac{1}{N} \sum_1^N (x_i - \mu_S)^4}{(\sigma_S)^4} \quad [\text{Ec.6}]$$

Donde  $K_S$  es el coeficiente de curtosis de Pearson. Según el resultado de este coeficiente las distribuciones de los datos (como se observa en la *figura 3*) pueden ser:

- Las distribuciones leptocúrticas o leptokúrticas ( $K_S > 3$ ), que difieren de las distribuciones normales por tener colas más ligeras y una protuberancia más pronunciada, se indican mediante valores de curtosis positivos. Esto sugiere que existe una mayor probabilidad de valores extremos en el conjunto de datos.
- Los valores de curtosis negativos significan una distribución platicúrtica o platikúrtica ( $K_S < 3$ ), que es diferente de una distribución normal en que tiene una protuberancia más plana y colas más pronunciadas. La probabilidad de valores extremos en el conjunto de datos se reduce.
- Una distribución mesocúrticas o mesokúrtica, definida como una con un valor de curtosis de 3, es aquella con una cola que es igualmente pesada y prominente como la cola de la distribución normal (Royston, 1992).

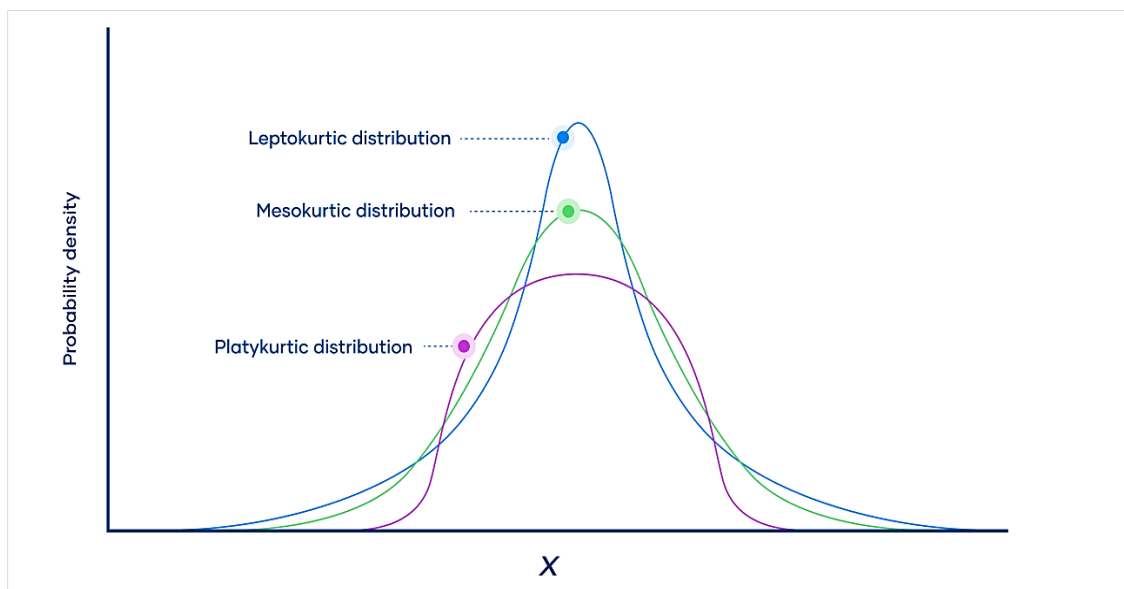


Figura 3. Representación de los diferentes tipos de distribuciones según el coeficiente de curtosis.  
Fuente: Turney, (2022).

### 1.3.2. Estadísticos de momento corregidos (ponderados)

Los momentos de una distribución de datos se pueden calcular utilizando momentos estadísticos ponderados, que tienen en cuenta los pesos asignados a cada observación. Al contrario que los momentos no ponderados, los momentos estadísticos ponderados se calculan multiplicando los datos por los pesos apropiados antes de realizar los cálculos.

#### 1.3.2.1 Media (primer momento)

En el caso de la media de forma ponderada, se suman todos los datos multiplicados por el peso que tenga ese dato y se divide entre la suma de los pesos. Para ello, se utiliza la media utilizada en el Ec.2 pero teniendo en cuenta los pesos. Por tanto, el cálculo de la media ponderada para los diferentes datos micrometeorológicos ( $\mu_c$ ) sería:

$$\mu_c = \frac{\frac{1}{N} \sum_{i=1}^N x_i \cdot a_i}{\frac{1}{N} \sum_{i=1}^N a_i} \quad [\text{Ec.7}]$$

Donde  $a_i$  es el peso asignado para cada dato  $x_i$ . El resto de los parámetros son los mismos que para la media aritmética (Kowalski, 2012).

#### 1.3.2.2 Desviación estándar (segundo momento)

La desviación estándar de manera ponderada es una medida estadística que permite combinar la variabilidad que tiene el conjunto de datos [E.4], con los pesos que tiene cada uno de estos datos. Igualmente, se utiliza la ecuación de la varianza aritmética [Ec.3], pero ya realizando la raíz cuadrada y ponderando cada uno de los valores medidos, permitiendo obtener así, la desviación estándar ponderada ( $\sigma_s$ ) como:

$$\sigma_c = \sqrt{\frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_c)^2 \cdot a_i}{\frac{1}{N} \sum_{i=1}^N a_i}} \quad [\text{Ec.8}]$$

Siendo  $\mu_c$  la media ponderada de los valores calculada anteriormente,  $a_i$  es el peso asignado para cada dato aportado por las torres Eddy ( $x_i$ ) (American Meteorological Society, 2015).

#### 1.3.2.3 Asimetría (tercer momento)

La asimetría ponderada, es una medida de asimetría de un conjunto de datos que tiene en cuenta los pesos asignados a cada uno de ellos. Una medida estadística conocida como asimetría identifica la falta de simetría de una distribución de datos. Teniendo en cuenta la [Ec.5] al añadirle los pesos quedaría como:

$$E_c = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_c)^3 \cdot a_i}{\frac{1}{N} \sum_{i=1}^N (\sigma_c)^3 \cdot a_i} \quad [\text{Ec.9}]$$

Donde  $E_c$  es la asimetría para los datos analizados calculado de forma ponderada y  $\sigma_c$  es la desviación estándar corregida calculada en el apartado anterior. El resto de los parámetros son los definidos anteriormente.

#### 1.3.2.4 Curtosis (cuarto momento)

La curtosis ponderada se utiliza para cuantificar el grado de desviación de los datos con respecto a una distribución normal pero teniendo en cuenta los pesos asignados para cada dato. La curtosis es una medida estadística que cuantifica cuánto aumenta o disminuye una distribución de datos en comparación con la distribución normal. Teniendo en cuenta la [Ec.6] donde se cuantifica la curtosis de forma aritmética, al añadirle la ponderación de los pesos de cada dato, obtendríamos la curtosis ponderada o coeficiente de Pearson ponderado ( $K_c$ ):

$$K_c = \frac{\frac{1}{N} \sum_1^N (x_i - \mu_c)^4 \cdot a_i}{\frac{1}{N} \sum_1^N (\sigma_c)^4 \cdot a_i} \quad [\text{Ec.10}]$$

#### 1.4 Revisión bibliográfica de los estadísticos de momento.

Tras una larga búsqueda de artículos, no se han encontrado investigaciones en las que se haya calculado los distintos momentos estadísticos teniendo en cuenta que el muestreo de los datos no es aleatorio (como se ha explicado anteriormente en el punto 1.1). Por tanto, la revisión bibliográfica se ha realizado con el enfoque de saber como se usan los distintos momentos estadísticos en la investigaciones (ya que, son muchas las formas de utilidad de estos) y sobre todo, como se interpretan estos resultados.

Hay un gran número de estudios que utilizan el primer y segundo momento estadístico para ver la concordancia de los datos y determinar la validez de la muestra. Un ejemplo es en el estudio de Rai et al.,(2017). En este estudio, se utiliza el modelo WRF-LES (investigación y pronóstico del tiempo con simulación de grandes remolinos) con un espaciado de cuadrícula de 30 metros para simular varias variables meteorológicas de importancia para aplicaciones de flujos de entrada turbulento en la capa límite planetaria (PBL) convectiva sobre un área de terreno complejo. El objetivo de utilizar un espaciado de cuadrícula de 30 m en la simulación LES, es resolver tantos remolinos que contienen energía como sea posible, lo que lleva a mejores estimaciones del transporte vertical de impulso, calor y humedad dentro de la capa límite. Los resultados de la simulación se compararon con las observaciones y se encontró que concuerdan bien en lo que respecta a los momentos de primer y segundo orden, así como a los espectros de turbulencia (Rai et al., 2017).

Además, existen estudios como el de Lenschow et al., (2000), donde se utilizan los momentos estadísticos para eliminar datos que no concuerden con las mediciones. En este caso, se crea una técnica para eliminar el ruido no correlacionado de las mediciones lidar de vapor de agua y velocidad vertical del aire en un PBL convectivo con el fin de recuperar estimaciones insesgadas de momentos hasta el cuarto orden. Se demostró que, tanto los lidars con detección de constituyentes como con sensores de velocidad pueden medir momentos de orden superior en el PBL con suficiente precisión para que las comparaciones con, por ejemplo, simulaciones numéricas de grandes remolinos sean factibles. También concluyeron que es posible obtener resultados aún mejores con mejoras en la instrumentación y las técnicas de medición. Es el caso de que los datos DIAL (sistema lidar de absorción diferencial) que deberían registrarse con una resolución temporal más alta, del orden de 1s para permitir un muestreo más fino del subrango inercial. Además, la SNR de HRDL (relación señal-ruido del Lidar Doppler de alta resolución) debería aumentarse en la zona de arrastre, por ejemplo, aumentando la energía del pulso o la potencia promedio (Lenschow et al., 2000).

En el caso de la investigación de Mortarini et al., (2009) se acerca más a poder comparar algunos resultados con esta investigación. Se tomaron medidas continuas tanto de viento como de turbulencia durante un año en tres alturas; 5, 9 y 25 metros en una torre ubicada a las afueras de la ciudad de Turín en Italia. De estos datos, se analizaron los de abril del año 2007, donde incluyen los resultados de cada nivel de la media, desviación estándar, asimetría y curtosis de las componentes tridimensionales del viento y la temperatura sónica. En sus resultados se observa que los datos de la componente vertical concluyen bastante bien, pero por el contrario, las componentes horizontales se desvían de los valores prescritos, lo cual se justifica por los efectos de los edificios y otros obstáculos y por el alto porcentaje de viento bajo (Mortarini et al., 2009).

En el estudio de Alberghi et al., (2002) se observa como aún sigue en continuo estudio los estadísticos de momento y sus relaciones. En este artículo, se analizó la velocidad vertical en la capa límite convectiva atmosférica a partir de datos de SODAR en dos estaciones de Roma en volúmenes centrados en altura con una muestreo de 621 Hz de frecuencia. En este conjunto de datos se encontró que la asimetría y la curtosis estaban relacionados. Por ello, se crea un modelo cuadrático que permite cuantificar esta relación. Es decir, permite calcular la curtosis a partir de la asimetría con un grado alto de confianza. Proponen la ecuación:

$$K = \alpha (S^2 + 1)$$

Donde K es la curtosis, S es la asimetría y  $\alpha$  es un parámetro libre que se propone como modelo para la relación entre la asimetría y la curtosis en la capa límite convectiva fuertemente dominada por la convección. Se obtiene al ajustar los datos de la capa límite convectiva (Figura 4) obteniendo un valor aproximado de  $2,4 \pm 0,3$  (Alberghi et al., 2002).

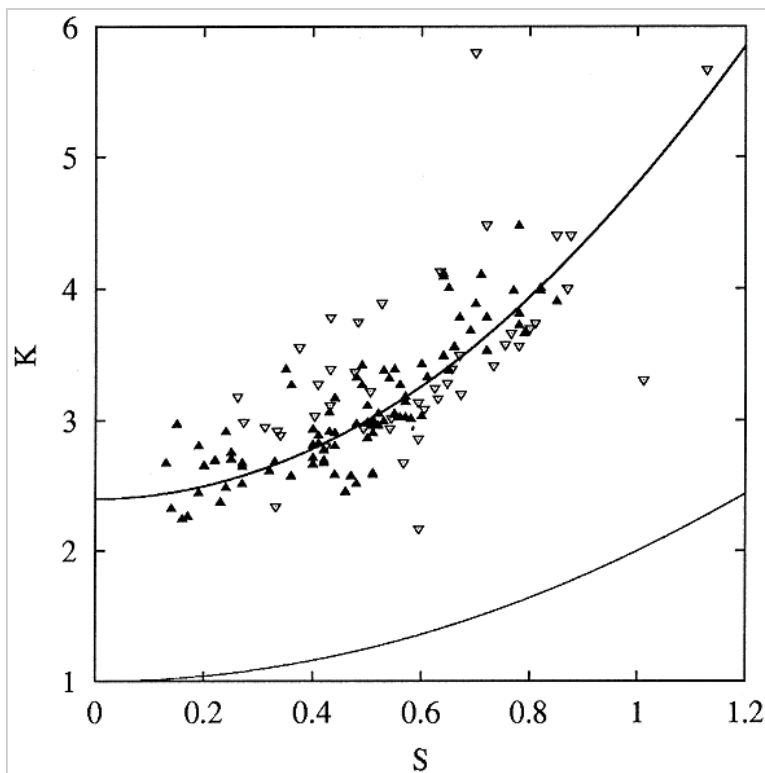


Figura 4: Relación entre asimetría y curtosis (línea gruesa continua, línea superior) límite estadístico (línea delgada continua, línea inferior), conjunto de datos actual(▲), datos recogidos en la literatura (▽).

Fuente: Alberghi et al., 2002

Igualmente se observa en la investigación de Ostashev y Wilson, (2000) que aún siguen en continuo estudio los estadísticos de ponderación en donde se analizan las fluctuaciones de temperatura en los estadísticos de alto orden (asimetría y curtosis) al ser de gran importancia tanto para los esquemas de parametrización como para la teoría básica de turbulencia atmosférica. En esta investigación, se usan datos tomados en un prado llano y uniforme. Se observó, que siguen prácticamente la distribución de la teoría de similitud (Teoría de Similitud de Monin-Obukhov), aunque con una ligera dispersión. Esto se concluyó que tenía el motivo de que, cuando la estratificación térmica estaba cercana a la neutralidad, la temperatura seguía una distribución prácticamente gaussiana, en cambio, cuando la inestabilidad aumentaba, la asimetría y la curtosis también aumentaban y se acercaban a valores prácticamente constantes (1,5 para la asimetría y 5,3 para la curtosis) en condiciones muy inestables. Debido a estos resultados, se propusieron fórmulas para la definición de unas condiciones inestables:

$$S \text{ (asimetría)} = \frac{-7.1\zeta}{(1 - 4.8\zeta)}$$

$$K \text{ (curtosis)} = \frac{(3 + 9.5\zeta^2)}{(1 + 1.8\zeta^2)}$$

Siendo  $\zeta$ , el parámetro de estabilidad adimensional, calculado como la altura de la capa superficial entre la longitud de Obukhov (Ostashev y Wilson, 2000).

En el artículo de Chen et al., (2017) se discute el impacto de la ponderación de modelos climáticos para estudio hidrometeorológicos de cambio climático. Se evalúa como la ponderación en los modelos climáticos globales tiene un impacto limitado sobre las variables hidrológicas. Se utilizan 5 métodos diferentes de ponderaciones en 28 modelos climáticos. Este estudio se realizó en una única cuenca dominada por nieve y utilizando un único modelo hidrológico. Dado que los resultados y los sesgos de los modelos climáticos varían tanto espacial como temporalmente, se necesita trabajo adicional para generalizar las conclusiones extraídas de este estudio. Se concluyó que, tiene sentido dar más peso a los modelos de mejor rendimiento. Sin embargo, en la práctica, es una tarea compleja ordenar modelos según el rendimiento, y asignar ponderaciones basadas en el rendimiento añade otra capa de subjetividad a todo el proceso. Las métricas de ponderación comunes que se centran principalmente en los sesgos de la media y la varianza solo reflejan un aspecto del desempeño de un modelo climático. Estas métricas son importantes, pero para los estudios de impacto del cambio climático (por ejemplo, hidrología y agricultura), es posible que se necesiten otras métricas (Chen et al., 2017).

## 2. JUSTIFICACIÓN DEL ESTUDIO Y OBJETIVO

La justificación de este estudio radica en la necesidad de comprender y corregir el sesgo que resulta al no ser un muestreo aleatorio el de los datos micrometeorológicos tomados, según la Ley de Charles. Estos cálculos imprecisos, pueden dar lugar a interpretaciones y conclusiones incorrectas sobre el comportamiento del clima y los fenómenos atmosféricos. Además, la falta de estos estudios sobre los sesgos estadísticos ligados a las muestras de datos hace aún más importante su investigación.

El objetivo es; mediante la inclusión de factores de ponderación que tengan en cuenta adecuadamente estas variaciones, conseguir datos no sesgados.

### 3. METODOLOGÍA

#### 3.1 Lugar de estudio

Este estudio se realiza en cuatro regiones de medida con el fin de analizar el impacto del uso de métricas ponderadas vs. aritméticas a la hora de caracterizar el comportamiento micrometeorológico en cuatro ecosistemas muy diferentes. Estos son: Padul, Balsa Blanca, Gádor y en una zona de olivar en Úbeda. En el Anexo II se muestra la ubicación geográfica para poder facilitar su comprensión.

##### 3.1.1. Padul

Está ubicado en la provincia de Granada, España. El municipio se compone según los datos del Instituto de Estadística y cartografía de Andalucía en 2019 de una extensión de 89,09 Km<sup>2</sup>. Dentro de este municipio, la zona estudiada es la zona de humedal y turbera del Padul (37°0'42.26"N, 3°36'20.65"W). Este lugar está incluido en la Convención Ramsar para los Humedales (sitio número 1674, "Convención sobre Humedales de Importancia Internacional Especialmente como Hábitat de corrientes de agua") (Acosta et al., 2019). El humedal tiene unos 3,3 km<sup>2</sup> y está a una altura de 744 m sobre el nivel del mar. La temperatura media del aire es de unos 16°C y precipitación anual de 450 mm. Predominan los vientos del noroeste (Serrano-Ortiz et al., 2020).

Los juncos, concretamente *Phragmites australis* [Cav.], Trin. ex. Steud. es la vegetación predominante en el paisaje estudiado. El junco común es una planta alta, polinizada por el viento, con brotes anuales que crecen hasta 5 m de altura a partir de un extenso sistema de raíces y tallos (Acosta et al., 2019).

##### 3.1.2. Balsa Blanca

Esta área se encuentra en el Parque Natural de Cabo de Gata-Níjar en la provincia de Almería, al sureste de España (Sánchez-Cañete et al., 2013). La zona marcada es un pastizal semiárido ubicado en la coordenadas (N36°56'26.0", W2°01'58.8"), a una altitud de 208 m sobre el nivel del mar y situado a 6,3 km del mar Mediterráneo.

Con respecto al clima, es semiárido subtropical. Su temperatura media anual es de 18°C y su precipitación media anual de aproximadamente 220 mm. Se caracteriza por tener las mayores precipitaciones en otoño (López-Ballesteros et al., 2016) e invierno y también, se caracteriza por una larga estación seca, que comienza en junio y termina abruptamente en septiembre-octubre con eventos de pulsos de lluvia (Sánchez-Cañete et al., 2013).

El ecosistema tipo corresponde al *espartal*, un pastizal mediterráneo semiárido donde la especie dominante es *Machrochloa tenacissima*. Este tipo de ecosistema está muy extendido en la región del Mediterráneo occidental (López-Ballesteros et al., 2018).

##### 3.1.3. Gádor

El área de estudio se encuentra en la Sierra de Gádor, Almería, España, concretamente, en el Llano de los Juanes, uno de los macizos montañosos de esta Sierra (Sánchez-Cañete, 2009). Sus coordenadas son 36°55'41.7"N, 2°45'1.7"W. Es una meseta subalpina con matorral mediterráneo que se ubica a 1600 m sobre el nivel del mar y a 25 km de la costa (Serrano-Ortiz et al., 2014).



Con respecto al clima, tiene una temperatura media anual de 12°C con máximas en verano de 31°C y mínimas en invierno de -6°C y precipitación de 475mm, siendo el verano extremadamente seco y otoño e invierno las estaciones lluviosas. Destaca por ser un terreno kárstico (Sánchez-Cañete, 2009). La cubierta vegetal está dominada por tres especies perennes; *Festuca scariosa* (19%), *Hormathophilla spinosa* (7%) y *Genista pumila* (6 %) (Serrano-Ortiz et al., 2014).

#### 3.1.4. Olivar de Úbeda

Esta zona de estudio se ha llevado a cabo en “Cortijo Guadiana” (37°54'39.30"N, 3°13'42.40"W), en Úbeda (Jaén, España). Pertenece a la empresa aceitera Castillo de Canena, SL (Aranda-Barranco et al, 2023).

Está situado a 370 m sobre el nivel de mar. Se caracteriza por un clima mediterráneo con temperaturas media anuales de 16°C y una precipitación anual de 495 mm. La evapotranspiración potencial anual en esta zona de cultivo (calculada mediante la ecuación de Penman-Monteith) es de 1220 mm. El terreno de cultivo tiene una extensión de 1500 ha. La especie de cultivo es huerta de olivo de regadio (*Olea europaea* L. “Arbequina”) (Chamizo et al., 2017).

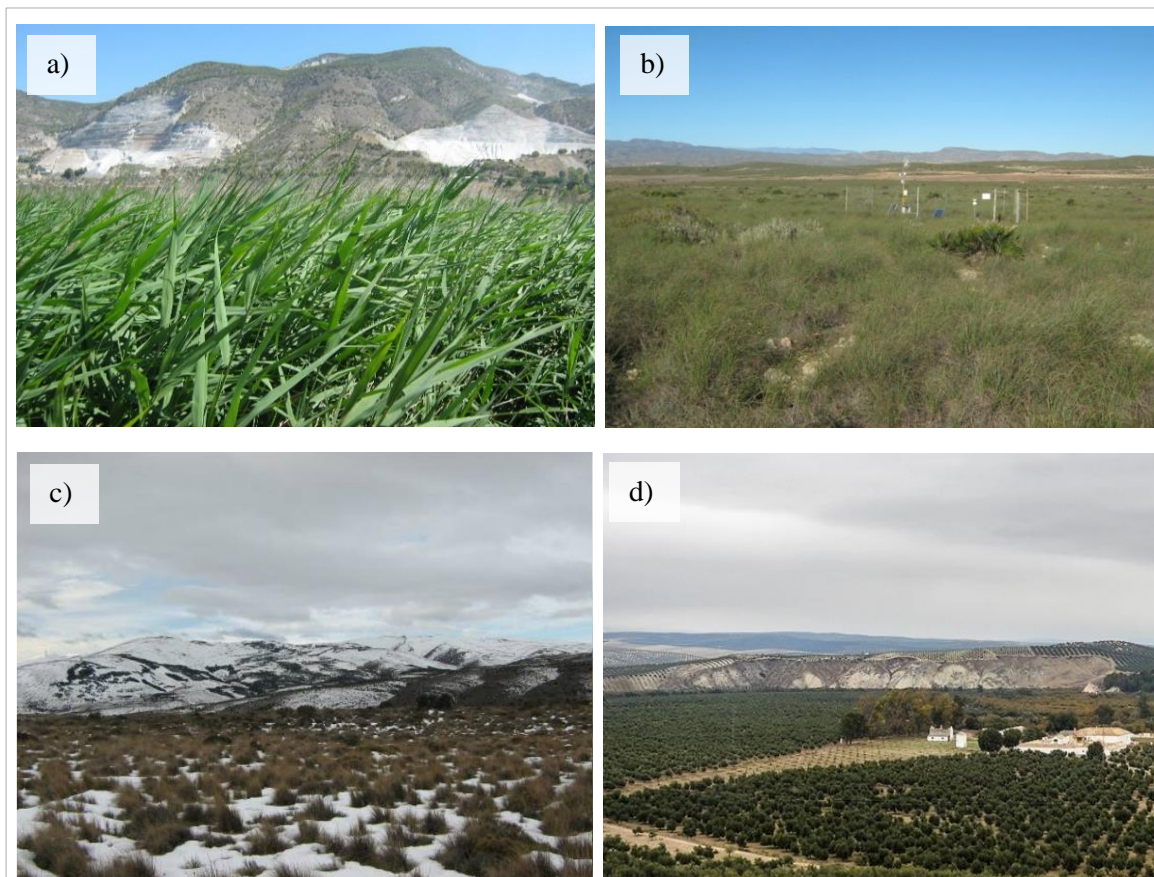


Figura 5. Zonas de estudio: a) Padul. Fuente: Serrano-Ortiz, 2020. b) Balsa Blanca. Fuente: López Ballesteros, 2017. c) Sierra de Gádor. Fuente: Sánchez-Cañete, 2009. d) Olivar de Úbeda. Fuente: Arjona. L. (s.f.).

### 3.2 Toma de datos

Los datos fueron proporcionados por el Departamento de Física Aplicada y el Departamento de Ecología de la Universidad de Granada. Las mediciones se tomaron a través de torres Eddy Covariance (EC) que son estructuras diseñadas para aportar datos de flujos de energía, temperatura, agua, gases... etc. Concretamente, todas las torres de los ecosistemas tratados tenían una respuesta rápida (10Hz).

En esta investigación se toma solamente los datos del mes de julio, al haber mayor radiación en los ecosistemas estudiados y por tanto, un posible sesgo mayor. De las medidas proporcionadas por la torre Eddy se seleccionó la hora de la toma del dato, la temperatura sónica ( $^{\circ}\text{C}$ ), velocidad longitudinal (m/s), velocidad latitudinal (m/s), velocidad vertical (m/s), la cantidad de humedad ( $\text{mg}/\text{m}^3$ ) y el Infra-red Gas Analyzer (IRGA). Se realizó una revisión mensual de las condiciones de los instrumentos (principalmente la limpieza de las lentes) y la configuración de la torre.

El anemómetro sónico se utilizó para medir las componentes del vector del viento y la temperatura sónica, en todas las ubicaciones es 'CSAT-3, Campbell Scientific, Logan, UT, EE. UU.'. Con respecto al analizador de densidad de  $\text{H}_2\text{O}$  y  $[\text{CO}]_2$  es 'LI-7500; LI-COR Inc., Lincoln, NE, EE. UU.' (close-path) en las torres ubicadas en Padul, Balsa Blanca y Gádor. En cambio, para la torre ubicada en el olivar de Úbeda es el 'IRGA, Li-Cor 7200; Lincoln, NE, EE. UU.' (open-path). En el primer caso se necesitará inclusión

### 3.3 Diseño experimental

La torres Eddy proporcionan 10 datos cada segundo, es decir, al tener todos los datos del mes de julio da como resultado 26.784.000 de filas de datos para un solo ecosistema, lo que quiere decir 107.136.000 filas de datos en el total de los lugares de estudio. Al ser nueve parámetros los proporcionados por la instrumentación que nos interesan; hora de la toma del dato, la temperatura sónica ( $^{\circ}\text{C}$ ), la cantidad de humedad ( $\text{mg}/\text{m}^3$ ), IRGA (divididos en `diag_csat` y `diag_irga_raw`), densidad del dióxido de carbono ( $\text{mg}/\text{m}^3$ ), velocidad longitudinal (m/s), velocidad latitudinal (m/s) y velocidad vertical (m/s), nos daría un total de 964.224.000 datos (prácticamente mil millones).

Al ser una gran cantidad de datos, estos, se agrupan en archivos de datos de media hora (es decir, 18000 filas por cada parámetro) para facilitar su manipulación. Por tanto, para el mes de julio tendría que haber un conjunto de 1488 paquetes de datos de medias horas (48 medias horas de un día multiplicado por los 31 días del mes de julio) por cada ecosistema.

Es prácticamente imposible que los datos de un mes completo salgan todos válidos ya que hay otros factores externos no controlables que influyen en la toma de datos como son el viento, las hojas, lluvia, insectos, personas, fallos de la instrumentación...etc que hacen que esos valores se tengan que filtrar y en su caso, eliminar.

Por tanto, de los 1488 paquetes iniciales primero se hizo un filtrado según; la franja de verificación del IRGA entre 246-248 y los datos repetidos, solapados o fuera de la franja de tiempo, quedándose en; Padul con 1450, Balsa Blanca con 322 (por que además, a partir del 15 de julio los paquetes de datos solo tienen entre 3.000-5.000 datos en vez de 18000, lo que altera los cálculos), Gádor con 835 (por que además, solo se aportó datos hasta el 19 de julio) y el olivar de Úbeda con 1263 datos.

En segundo lugar, se calcula la ponderación de cada uno de los datos, para posteriormente ser usado en los métodos estadísticos. Para esta ponderación se usó el artículo de Kowalski, 2012.

La temperatura ponderada ( $\tilde{T}$ ) se calculó mediante [Ec.11] donde se utiliza como factores de ponderación; la densidad del aire seco ( $\rho_i$ ) y el calor específico ( $C_{p,i}$ ).

$$\tilde{T} \equiv \frac{\frac{1}{N} \sum_{i=1}^N \rho_i C_{p,i} T_i}{\frac{1}{N} \sum_{i=1}^N \rho_i C_{p,i}} \quad [\text{Ec.11}] \quad (\text{Kowalski, 2012})$$

Para la velocidades ponderadas ( $\tilde{u}, \tilde{v}, \tilde{w}$ ) se usó [Ec.12] para cada una de ellas que tiene en cuenta que la velocidad ponderada se calcula a través de los datos de velocidad obtenidos ( $u_i, v_i, w_i$ ) y multiplicados por la densidad del aire seco ( $\rho_i$ ).

$$\tilde{u}, \tilde{v}, \tilde{w} \equiv \frac{\frac{1}{N} \sum_{i=1}^N \rho_i (u,v,w)_i}{\frac{1}{N} \sum_{i=1}^N \rho_i} \quad [\text{Ec.12}] \quad (\text{Kowalski, 2012})$$

En el caso de las ponderaciones en las densidades de  $CO_2$  ( $\tilde{c}$ ) y  $H_2O$  ( $\tilde{q}$ ) se usó la [Ec.13] que al igual que en el caso anterior, se multiplica cada uno de los datos ( $c_i, q_i$ ) por la densidad del aire seco. Se necesita matizar que en el caso de los ecosistemas que tienen una instrumentación ‘open-path’ para la toma de estos datos (antes comentado), ya llevan incorporado el factor de ponderación al tomar los datos y por tanto, en estos casos no será posible el análisis del sesgo. El único lugar que puede ser, por tanto, analizado para las densidades es el Olivar de Úbeda (Guadiana).

$$\tilde{c}, \tilde{q} \equiv \frac{\frac{1}{N} \sum_{i=1}^N \rho_i (c,q)_i}{\frac{1}{N} \sum_{i=1}^N \rho_i} \quad [\text{Ec.13}] \quad (\text{Kowalski, 2012})$$

Por último, una vez obtenido para cada uno de los datos, su ponderación, ya se pueden realizar los estadísticos antes mencionados y analizarlos. Además, una vez obtenido los sesgos para todos los ecosistemas, también se realizó un análisis de la diferencia del sesgo producido tanto por el día como el producido por la noche (S.día-noche). Las horas tomadas como ‘día’ son desde las 8 am hasta las 8 pm y las de ‘noche’ las restantes. Este análisis se realizó por que, durante el día, el forzamiento solar produce variaciones de temperatura y convección, influyendo estas fluctuaciones en la temperatura haciendo que las variaciones en el volumen de una parcela (Ley Charles) sean más importantes.

Al tratarse de casi mil millones de datos, para su menor complicación, se utilizó el programa MATLAB2022b que permitió crea bucles que realizan la misma operación para cada uno de los datos seleccionados. Para la representación de los gráficos también se hizo uso de este programa. En el Anexo III. EJEMPLO DE DESARROLLO DE SCRIPT DE MATLAB se detalla como se obtuvieron los datos para la temperatura, para mostrar un ejemplo de los cálculos que se realizaron y los comandos más comunes.

## 4. RESULTADOS Y ANÁLISIS

### 4.1 Sesgo de la media

Los resultados estadísticos de la media de las distintas magnitudes micrometeorológicas para los cuatro ecosistemas se encuentran representados en la *Tabla 1*.

RESUMEN DEL ESTADÍSTICO DE PRIMER MOMENTO RESPECTO AL ORIGEN (MEDIA)					
		PADUL	BALSA BLANCA	GÁDOR	OLIVAR ÚBEDA
Datos micrometeorológicos	Año (julio)	2013	2013	2016	2018
Temperatura (°C)	$\mu_{s,T}$	23,5337	22,9463	25,8604	32,1607
	$\mu_{c,T}$	23,5313	22,9119	25,85497	32,1597
	Sesgo	0,0024	0,0316	0,00543	0,0040736
	<b>Error (%)</b>	<b>0,0102</b>	<b>0,1377</b>	<b>0,0210</b>	<b>0,0127</b>
	S. día-noche	0,00152	0,0363	0,0040736	0,0009390
Velocidad longitudinal (m/s)	$\mu_{s,u}$	0,8161	1,9299	0,2075	0,17407
	$\mu_{c,u}$	0,8148	1,8912	0,24123	0,17317
	Sesgo	0,00132	0,038752	0,003377	0,000902
	<b>Error (%)</b>	<b>0,162</b>	<b>2,008</b>	<b>1,627</b>	<b>0,518</b>
	S. día-noche	0,00078	0,045572	-0,02949	-0,004080
Velocidad latitudinal (m/s)	$\mu_{s,v}$	0,81082	1,3178	-0,89789	-0,70965
	$\mu_{c,v}$	0,81028	1,3131	-0,89327	-0,70797
	Sesgo	0,00054	0,004647	-0,00462	-0,00168
	<b>Error (%)</b>	<b>0,067</b>	<b>0,353</b>	<b>0,515</b>	<b>0,237</b>
	S. día-noche	0,00131	-0,023992	-0,03542	0,01213
Velocidad vertical (m/s)	$\mu_{s,w}$	0,01018	-0,068651	-0,04187	-0,00739
	$\mu_{c,w}$	0,01006	-0,069015	-0,04243	-0,00736
	Sesgo	0,000118	-0,000364	-0,000554	0,0000234
	<b>Error (%)</b>	<b>1,159</b>	<b>0,530</b>	<b>1,323</b>	<b>0,406</b>
	S. día-noche	0,000313	0,0030973	0,000458	0,0005261
Densidad de $CO_2$ ( $mg/m^3$ )	$\mu_{s,CO_2}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			368,5801
	$\mu_{c,CO_2}$				368,5747
	Sesgo				0,00533
	<b>Error (%)</b>				<b>0,0014</b>
	S. día-noche				0,0038
Densidad de $H_2O$ ( $g/m^3$ )	$\mu_{s,q}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			19,2837
	$\mu_{c,q}$				19,28177
	<b>Error (%)</b>				<b>0,01</b>
	Sesgo				0,00193
	S. día-noche				0,0026415

Tabla 1. RESUMEN DEL ESTADÍSTICO DE PRIMER MOMENTO RESPECTO AL ORIGEN (MEDIA). En esta tabla, se representa tanto los resultados de las medias aritméticas ( $\mu_s$ ) como ponderadas ( $\mu_c$ ) de los distintos datos micrometeorológicos. La fila de ‘Sesgo’ corresponde a la diferencia de la media aritmética menos la media ponderada de cada uno de los parámetros. La línea coloreada con fondo azul determina el error porcentual con respecto a la de la medida aritmética (sesgo dividido entre medida aritmética y multiplicado por 100). En color rojo se recalcan los errores más grandes. La franja denominada como ‘S. día-noche’ es la diferencia que hay entre el sesgo de día menos el sesgo producido por la noche.

Estos datos nos sugieren que sí existe en principio un sesgo producido por la sobreestimación de los datos de temperatura al tratar la media de forma aritmética y por tanto, que es más adecuado calcularla de forma ponderada. La realidad es que el error que se muestra es muy pequeño, siendo el mayor el producido en Balsa Blanca, pero, aun así, siendo poco relevante.

Existe, de forma general, un mayor sesgo durante el día que durante la noche, dando lugar a un tipo de gráfica de dientes de sierra, ya que, durante el día, se muestran grandes picos positivos y durante la noche el sesgo llega a rozar el 0 durante muchos momentos. Esto, puede observarse en la *figura 6*, donde se representan los datos para Padul, pero, ocurre de forma igualitaria en los cuatro ecosistemas. Una justificación de esta diferencia se produce debido a que durante el día, el forzamiento solar produce variaciones de temperatura y convección, influyendo estas fluctuaciones en la temperatura haciendo que las variaciones en el volumen de una parcela (Ley Charles) sean más importantes.

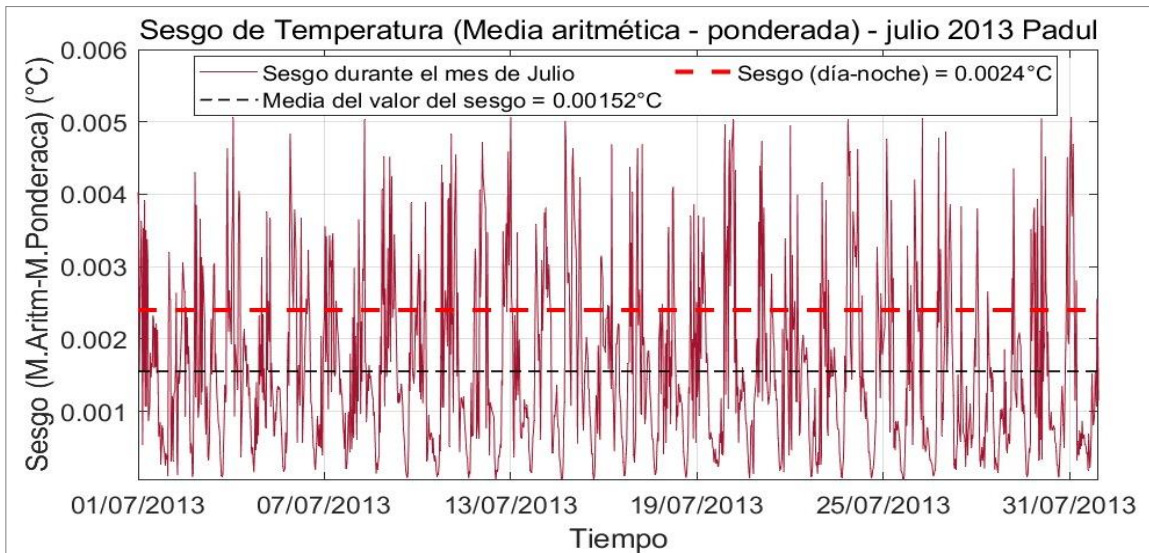


Figura 6. Sesgo producido en el cálculo de la media de temperatura en julio de 2013 en Padul. Se muestra el sesgo producido por la sobrestimación de la media aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color rojo oscuro que va variando sus grados (eje y) frente al tiempo del mes del julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal roja claro la media de este sesgo en el mes de julio que es de 0,0024 °C. Por último, también se observa una línea discontinua negra que representa la diferencia entre el sesgo producido durante el día menos el producido durante la noche que toma un valor de 0,00152 °C.

Si hacemos una comparativa con los resultados obtenidos para la velocidad longitudinal se observa que realmente este sesgo es algo mayor que en el caso de la temperatura. Si tenemos en cuenta los mayores sesgos producidos en la temperatura y en la velocidad longitudinal (ambos en el Balsa Blanca) en el primer caso nos daría un sesgo (error) de un 0,14% con respecto al cálculo de la media aritmética, en el caso de la velocidad longitudinal sería de un 2%. Es decir, el error es aproximadamente catorce veces mayor en la velocidad longitudinal que en la temperatura. Realmente sigue siendo pequeño el error, aunque es notorio. Desde un punto de vista matemático y teniendo en cuenta que cuanto más idóneos sean los resultados y el muestreo, mejor comprensión tendremos en la atmósfera, en principio, debería de tenerse en cuenta para los cálculos. En la *figura 7*, se esquematiza el ejemplo de Balsa Blanca, para hacer más visual este sesgo producido.

En el caso de la velocidad longitudinal (al igual que en la velocidad latitudinal y vertical), no se observó una diferencia de sesgo entre el día y la noche, lo cual es lógico, ya que no depende del ciclo día-noche, si no de las fuerzas que actúan sobre el aparato medidor de la Torre Eddy como son la resistencia del aire, aceleración... En el caso de velocidad latitudinal los sesgos son muy pequeños y prácticamente insignificantes, aun así, como en todos los parámetros, hay una sobrestimación de la muestra calculada aritméticamente y sin tener en cuenta factores de ponderación.

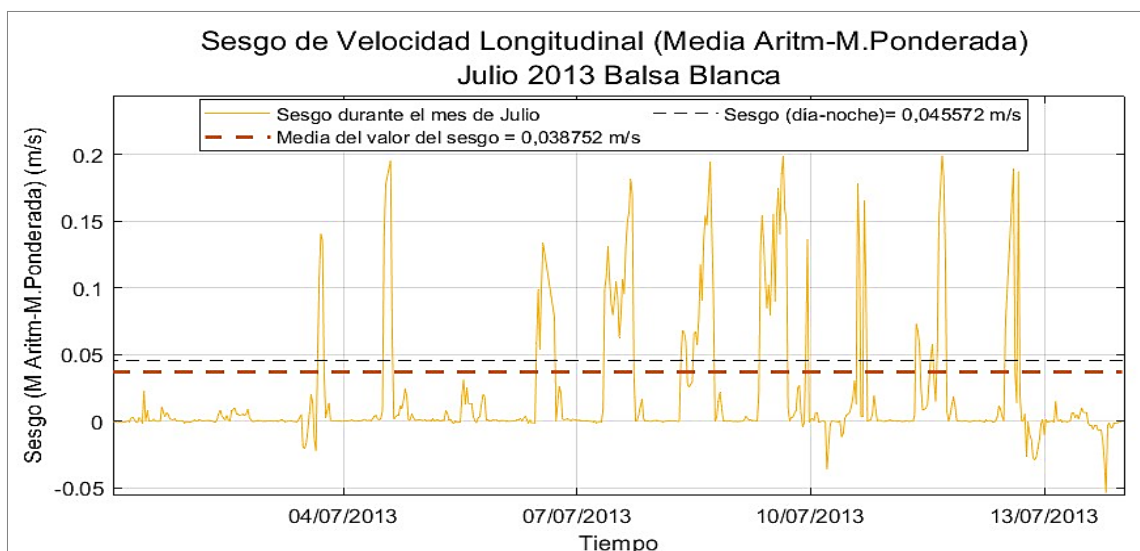


Figura 7. Sesgo producido en la media de la velocidad longitudinal en julio de 2013 en Balsa Blanca. Se muestra el sesgo producido por la sobrestimación de la media aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color amarillo claro que va variando su cantidad (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal naranja la media de este sesgo en el mes de julio que es de 0,038752 m/s. Por último, también se observa una línea discontinua negra que representa la diferencia entre el sesgo producido durante el día menos el producido durante la noche que toma una valor de 0,045572 m/s.

En cambio, sí que se muestra que la velocidad vertical tanto en Padul como en Gádor son de aproximadamente 1,1-1,3%. Realmente son errores algo más relevantes ya que dependiendo de la exactitud del estudio tratado puede llegar a ser relevante, aunque, debido a la complejidad de la obtención de los resultados mediante factores de ponderación, no llega a ser muy relevante. Es cierto, que al ser tan pequeñas las medidas de la velocidad vertical, aunque el sesgo sea muy pequeño, el error porcentual siempre va a tender a ser más grande. En la *figura 8*, se muestra el sesgo de la velocidad vertical en Padul como ejemplo para ver su distribución.

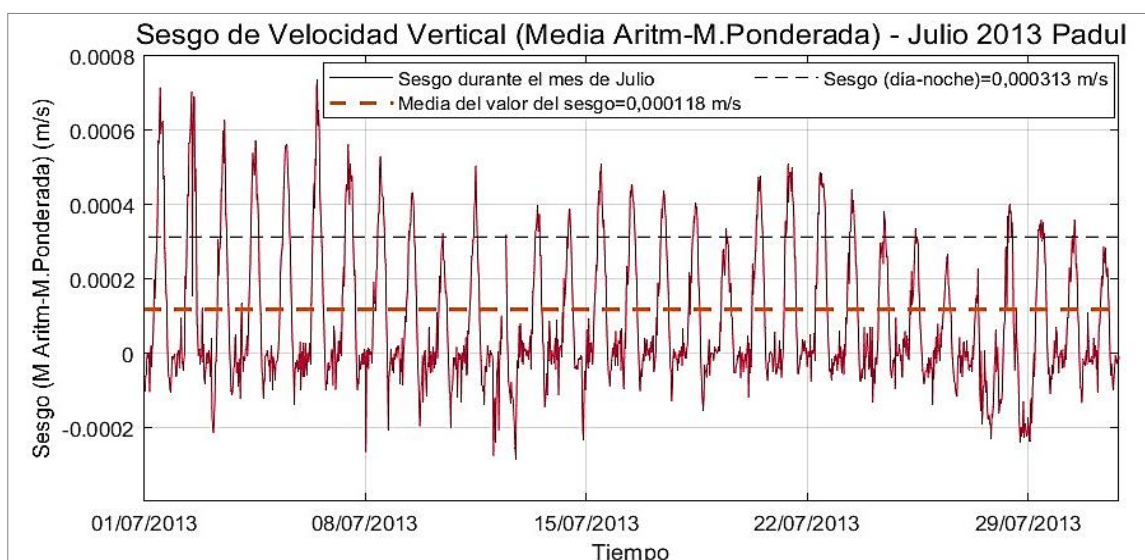


Figura 8. Sesgo producido en la media de la velocidad vertical en julio de 2013 en Padul. Se muestra el sesgo producido por la sobrestimación de la media aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color granate que va variando su cantidad (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal naranja la media de este sesgo en el mes de julio que es de 0,000118 m/s. Por último, también se observa una línea discontinua negra que representa la diferencia entre el sesgo producido durante el día menos el producido durante la noche que toma una valor de 0,000313 m/s.

En el caso de las densidades del  $CO_2$  y  $H_2O$  no se encontraron sesgo de gran significación. Estos cálculos solo se realizaron en el Olivar de Úbeda debido a la instrumentación, como se comentó anteriormente. La densidad del dióxido de carbono que al calcularla de forma aritmética da una media de  $368,5801 \text{ mg/m}^3$  y el sesgo  $0,00533 \text{ mg/m}^3$ , lo que supone un error de un 0,014% en los datos. Es un error insignificante, en la figura 9 se muestra la distribución de este. En el caso de la densidad del vapor de agua ocurre un caso parecido ya que el error es de una sobrestimación de un 0,01%.

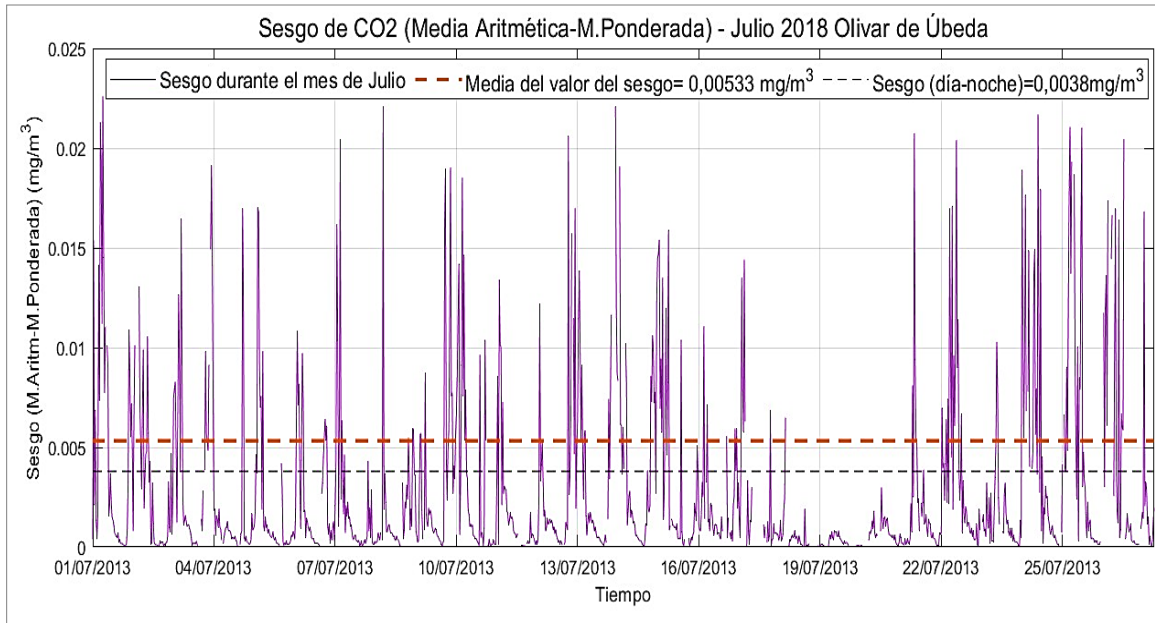


Figura 9. Sesgo producido en la media del  $CO_2$  en julio de 2018 en olivar de Úbeda. Se muestra el sesgo producido por la sobrestimación de la media aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color morado que va variando su cantidad (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal naranja la media de este sesgo en el mes de julio que es de  $0,00533 \text{ mg/m}^3$ . Por último, también se observa una línea discontinua negra que representa la diferencia entre el sesgo producido durante el día menos el producido durante la noche que toma un valor de  $0,0038 \text{ mg/m}^3$ .

Además, en ambas densidades si se muestra una diferencia del sesgo entre el día y la noche, aunque no es muy notoria, aun así, tiene una justificación clara la variación de densidad ya que depende de la temperatura (a más temperatura, la densidad del aire disminuye), humedad relativa (siendo mayor la densidad de vapor mayor si la humedad relativa es mayor), presión atmosférica (los cambios de presión afectan a la densidad del aire) y por supuesto, la actividad antropogénica como la quema de combustibles fósiles que afectan de manera significativa a la cantidad de  $CO_2$  durante el día, al ser mayor la actividad durante el día.

#### 4.2 Sesgo de la desviación típica

Los resultados estadísticos de la desviación típica de las distintas magnitudes micrometeorológicas para los cuatro ecosistemas se encuentran representados en la Tabla 2. Como se comentó en el punto 1.3.1.2 la desviación nos da información de la dispersión con respecto a la media.. En este caso, se puede decir que de forma generalizada los datos son cercanos a la media y con una baja desviación estándar, lo que indicia una distribución óptima y homogénea.

RESUMEN DEL ESTADÍSTICO DE SEGUNDO ORDEN RESPECTO A LA MEDIA (DESVIACIÓN TÍPICA)					
		PADUL	BALSA BLANCA	GÁDOR	OLIVAR ÚBEDA
Datos micrometeorológicos	Año (julio)	2013	2013	2016	2018
Temperatura (°C)	$\sigma_{s,T}$	0,67624	0,43611	0,97602	0,516383
	$\sigma_{c,T}$	0,67614	0,43488	0,97489	0,516256
	Sesgo	0,0000928	0,001227	0,001134	0,0001264
	<b>Error (%)</b>	<b>0,014</b>	<b>0,281</b>	<b>0,116</b>	<b>0,024</b>
	S. día-noche	0,0000903	0,001413	0,000944	0,000115
Velocidad longitudinal (m/s)	$\sigma_{s,u}$	0,65856	0,99265	0,92324	0,8244
	$\sigma_{c,u}$	0,65828	0,98687	0,92848	0,8211
	Sesgo	0,00028	0,005777	0,001476	0,00324
	<b>Error (%)</b>	<b>0,043</b>	<b>0,582</b>	<b>0,160</b>	<b>0,393</b>
	S. día-noche	0,00108	-0,001511	0,001189	-0,002069
Velocidad latitudinal (m/s)	$\sigma_{s,v}$	0,7157	1,0015	0,91855	0,8809
	$\sigma_{c,v}$	0,7152	0,9954	0,91479	0,8785
	Sesgo	0,00049	0,0079312	0,00376	0,00246
	<b>Error (%)</b>	<b>0,068</b>	<b>0,792</b>	<b>0,409</b>	<b>0,279</b>
	S. día-noche	-0,00092	0,004725	0,00766	-0,00483
Velocidad vertical (m/s)	$\sigma_{s,w}$	0,2534	0,45457	0,31938	0,38745
	$\sigma_{c,w}$	0,2533	0,45141	0,31870	0,38681
	Sesgo	0,0000131	0,0031632	0,00068	0,000636
	<b>Error (%)</b>	<b>0,005</b>	<b>0,696</b>	<b>0,212</b>	<b>0,164</b>
	S. día-noche	-0,0003889	-0,0095796	0,004918	0,0005269
Densidad de CO <sub>2</sub> (mg/m <sup>3</sup> )	$\sigma_{s,CO_2}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			0,2853
	$\sigma_{c,CO_2}$				0,2846
	Sesgo				0,0007
	<b>Error (%)</b>				<b>0,245</b>
	S. día-noche				0,006447
Densidad de H <sub>2</sub> O (g/m <sup>3</sup> )	$\sigma_{s,q}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			0,45236
	$\sigma_{c,q}$				0,45065
	Sesgo				0,00171
	<b>Error (%)</b>				<b>0,378</b>
	S. día-noche				-0,000487

Tabla 2. RESUMEN DEL ESTADÍSTICO DE SEGUNDO ORDEN RESPECTO A LA MEDIA (DESVIACIÓN TÍPICA). En esta tabla, se representa tanto los resultados de las desviaciones estándar aritméticas ( $\sigma_s$ ) como ponderadas ( $\sigma_c$ ) de los distintos datos micrometeorológicos. La línea de ‘Sesgo’ corresponde a la resta de los de la desviación estándar aritmética menos la desviación estándar ponderada de cada uno de los parámetros medidos individualmente. La línea coloreada con fondo azul determina el error porcentual con respecto a la sobrestimación de la medida aritmética. En color rojo se recalcan los errores más grandes. La franja denominada como ‘S.día-noche’ es la diferencia que hay entre el sesgo de día menos el sesgo producido por la noche.

En la *Figura 10* se muestra un ejemplo del sesgo de la desviación típica en este caso de la temperatura del ecosistema de Padul. En él, se puede observar que sigue existiendo la diferenciación del sesgo en el día y la noche, pero al analizar los datos, se observa que tanto el valor del sesgo como la diferencia del sesgo del día menos la noche son muy mínimos.



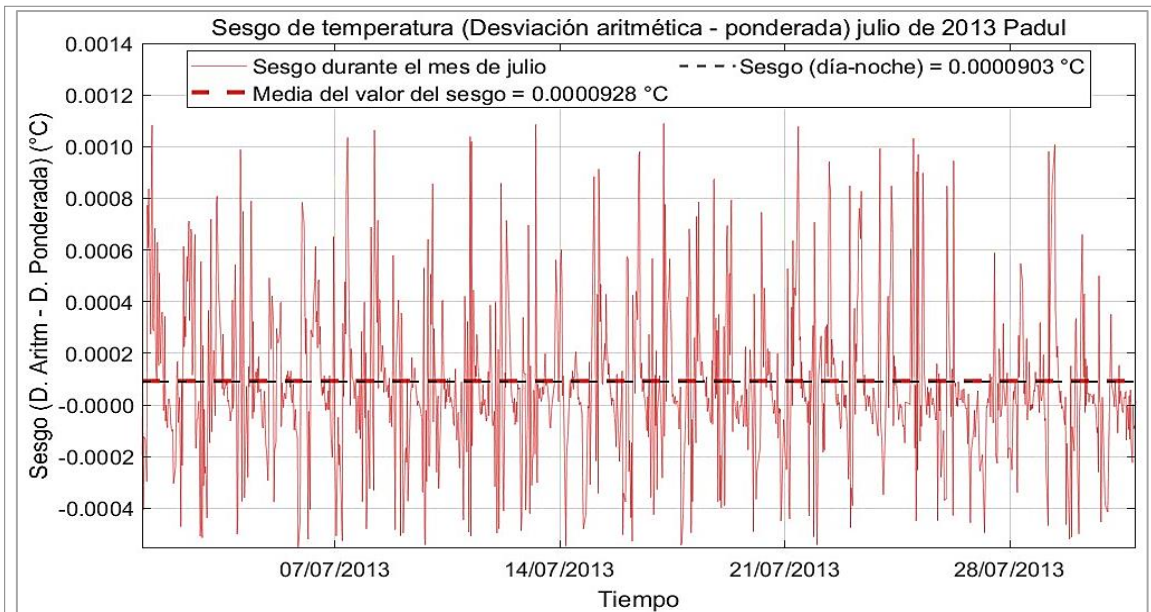


Figura 10. Sesgo producido en la desviación típica de la temperatura en julio de 2013 en Padul. Se muestra el sesgo producido por la sobrestimación de la desviación típica calculada de forma aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color rojo claro que va variando sus grados (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal roja la media de este sesgo en el mes de julio que es de 0,0000928 °C que es prácticamente la misma que la línea discontinua negra que representa la diferencia entre el sesgo producido durante el día menos el producido en la noche que toma un valor de 0,0000903°C.

Realmente todos los ecosistemas tienen un comportamiento similar en cuanto a la diferencia entre desviaciones lo cual también ocurre en todas las variables, menos en el caso de Balsa Blanca. Es el ecosistema que en todos los parámetros obtiene el mayor sesgo. En la figura 11 se muestra el sesgo de la velocidad longitudinal para este ecosistema.

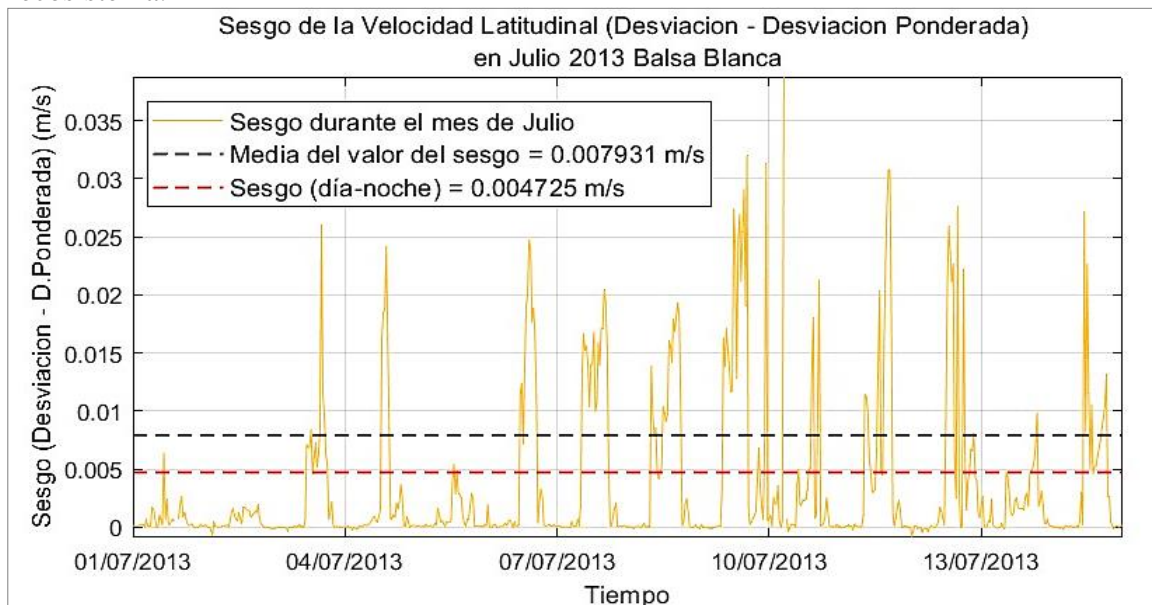


Figura 11. Sesgo producido en la desviación típica de la velocidad latitudinal en julio de 2013 en Balsa Blanca. Se muestra el sesgo producido por la sobrestimación de la desviación típica calculada de forma aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color amarillo que va variando en m/s (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal negra la media de este sesgo en el mes de julio que es de 0,00791 m/s y con una línea discontinua roja que representa la diferencia entre el sesgo producido durante el día menos el producido en la noche que toma un valor de 0,004725 m/s.

Este mayor sesgo en todos los parámetros en este ecosistema puede venir justificado a que al realizar la filtración de los datos, al final, se acaba con un número más pequeño de datos (de hecho casi una cuarta parte), con respecto al resto de muestras de los demás ecosistemas, lo que puede sugerir, que a un número más pequeño de muestra, la probabilidad de sesgo o error es algo mayor. Ya que, cuando una muestra es más pequeña, es más probable que los valores individuales impacten más en el resultado de la desviación estándar aritmética o ponderada. Aun así, la desviación sigue resultando insignificante, menor del 0,8% en todos los casos.

También es de comentar, que realmente en la gráfica parece que los datos no son muy homogéneos y que varían bastante. En los próximos estadísticos se analizará más profundamente si estos datos realmente son óptimos.

### *4.3 Sesgo de la asimetría*

Los resultados de la asimetría se encuentran en la Tabla 3

Una distribución sería completamente simétrica si el resultado fuera 0, es decir, cuanto más cercano esté de este, se puede confirmar que los datos son óptimos. Si la asimetría es positiva o negativa simplemente indica que los resultados están distribuidos más hacia un lado u otro de la curva. En general, los resultados muestran una asimetría positiva es decir que su distribución de datos se extiende más hacia la derecha pero, también existen algunos datos como en la velocidad longitudinal y latitudinal del olivar de Úbeda, entre otros, que muestran asimetría negativa es decir, que su distribución de datos está más representada en la parte negativa. Tanto uno como otro ya sean negativos o positivos se acercan mucho a 0 y por tanto, siguen siendo igual de idóneos.

Al tratar con números tan pequeños y cercanos a 0, aunque el sesgo sea muy pequeño en comparación con el dato aritmético, el porcentaje de error será muy grande. Esto ocurre expresamente en los datos de velocidades del olivar de Úbeda donde los sesgos que se producen en las velocidades son muy pequeños, pero, al ser los datos en sí de la asimetría aritmética y ponderada ya pequeños, al calcular el porcentaje de error da bastante alto. En este caso, da en torno a un 3% de error, pero al tratarse de números tan sumamente pequeños, podría llegarse a la conclusión de que es un error despreciable. De hecho, los sesgos son de milésimas decimales.

Al igual que en el caso de la desviación, lo que sí es altamente relevante son los errores producidos en Balsa Blanca para todos los parámetros. Haciendo una comparativa con los demás ecosistemas se puede llegar a concluir que realmente ha podido ser un error de la muestra inicial. Ya que, desde una primera instancia se tuvo que eliminar los días después del 17 de Julio de este mismo mes al estar mal medidos y además, ser con diferencia el ecosistema que más se han eliminado datos en el filtrado. Además, en el estadístico de la media (figura 7) ya se mostraba una distribución muy poco uniforme, siendo este, el ecosistema en el que más sesgo se producía. Igualmente, ocurría en la Tabla 2 de la desviación estándar, donde se demostraba que los mayores sesgos se volvían a producir en el mismo ecosistema. Y ahora, al tratarse del estadístico que nos informa de la distribución de los datos, el error se ve agravado. En la Figura 12 se muestra la distribución del estadístico de asimetría para Balsa Blanca y en la Figura 13 para Gádor, ambos del parámetro de velocidad latitudinal.

RESUMEN DEL ESTADÍSTICO DE TERCER ORDEN RESPECTO A LA MEDIA (ASIMETRÍA)					
		PADUL	BALSA BLANCA	GÁDOR	OLIVAR ÚBEDA
Datos micrometeorológicos	Año (julio)	2013	2013	2016	2018
Temperatura (°C)	$E_{s,T}$	0,000492	0,24341	0,24697	0,31979
	$E_{c,T}$	0,000491	0,10431	0,24620	0,31973
	Sesgo	0,000001	0,13910	0,00077	0,000054
	<b>Error (%)</b>	<b>0,203</b>	<b>57,146</b>	<b>0,312</b>	<b>0,017</b>
	S. día-noche	0,00043	0,15608	0,00078	0,000054
Velocidad longitudinal (m/s)	$E_{s,u}$	0,11729	0,14216	0,01315	-0,09019
	$E_{c,u}$	0,11661	0,043266	0,01312	-0,09260
	Sesgo	0,00068	0,09888	0,000029	0,00240
	<b>Error (%)</b>	<b>0,580</b>	<b>69,555</b>	<b>0,221</b>	<b>2,661</b>
	S. día-noche	-0,00145	-0,04192	0,000036	-0,00448
Velocidad latitudinal (m/s)	$E_{s,v}$	0,07949	0,12565	-0,058395	-0,07236
	$E_{c,v}$	0,08019	0,05096	-0,058577	-0,07465
	Sesgo	0,00069	0,074689	-0,000181	0,00228
	<b>Error (%)</b>	<b>0,868</b>	<b>59,442</b>	<b>0,310</b>	<b>3,151</b>
	S. día-noche	0,00085	0,023264	0,000016	-0,00415
Velocidad vertical (m/s)	$E_{s,w}$	0,10358	0,21198	0,19185	0,10294
	$E_{c,w}$	0,10322	0,20092	0,19172	0,10006
	Sesgo	0,000367	0,011067	0,000134	0,00287
	<b>Error (%)</b>	<b>0,354</b>	<b>5,221</b>	<b>0,070</b>	<b>2,788</b>
	S. día-noche	0,000253	0,0050963	-0,000369	-0,004992
Densidad de $CO_2$ ( $mg/m^3$ )	$E_{s,CO_2}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			0,0025947
	$E_{c,CO_2}$				0,0024568
	Sesgo				0,0001379
	<b>Error (%)</b>				<b>5,315</b>
	S. día-noche				0,0000437
Densidad de $H_2O$ ( $g/m^3$ )	$E_{s,q}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			0,13728
	$E_{c,q}$				0,13726
	Sesgo				0,000016
	<b>Error (%)</b>				<b>0,012</b>
	S. día-noche				0,0003177

Tabla 3. RESUMEN DEL ESTADÍSTICO DE TERCER ORDEN RESPECTO A LA MEDIA (ASIMETRÍA). En esta tabla, se representa tanto los resultados de la asimetría aritmética ( $E_s$ ) como la asimetría ponderada ( $E_c$ ) de los distintos datos micrometeorológicos. La línea de ‘Sesgo’ corresponde a la resta de los de la asimetría aritmética menos la asimetría ponderada de cada uno de los parámetros medidos individualmente. La línea coloreada con fondo azul determina el error porcentual con respecto a la sobrestimación de la medida aritmética. En color rojo se recalcan los errores más grandes. La franja denominada como ‘S.día-noche’ es la diferencia que hay entre el sesgo de día menos el sesgo producido por la noche.

En estas figuras se muestra de una forma sencilla como en el caso de Balsa Blanca la distribución no se acerca ni mucho menos a la simetría y que además, se muestran la falta de datos. Por el contrario, en Gádor, se observa una distribución homogénea y prácticamente simétrica.

Eso sí, a pesar del error de la muestra, continúa habiendo un sesgo en las medidas, habiendo en todo momento sobrestimación de las medidas aritméticas.

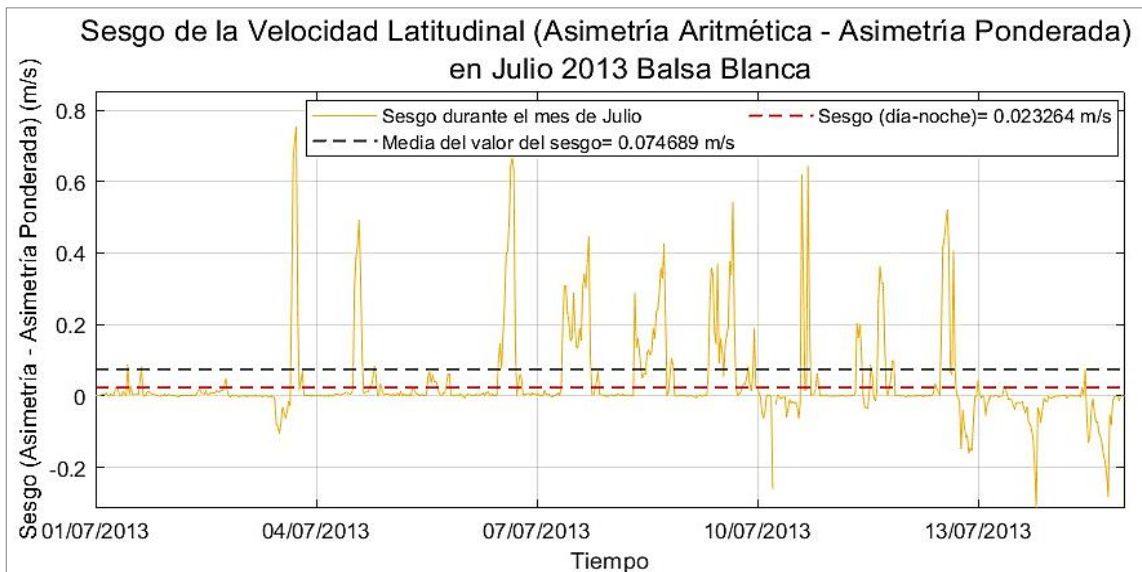


Figura 12. Sesgo producido en la asimetría de la velocidad latitudinal en julio de 2013 en Balsa Blanca. Se muestra el sesgo producido por la sobrestimación de la asimetría calculada de forma aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color amarillo que va variando en m/s (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal negra la media de este sesgo en el mes de julio que es de 0,074689 m/s y con una la línea discontinua roja que representa la diferencia entre el sesgo producido durante el día menos el producido en la noche que toma una valor de 0,023264 m/s.

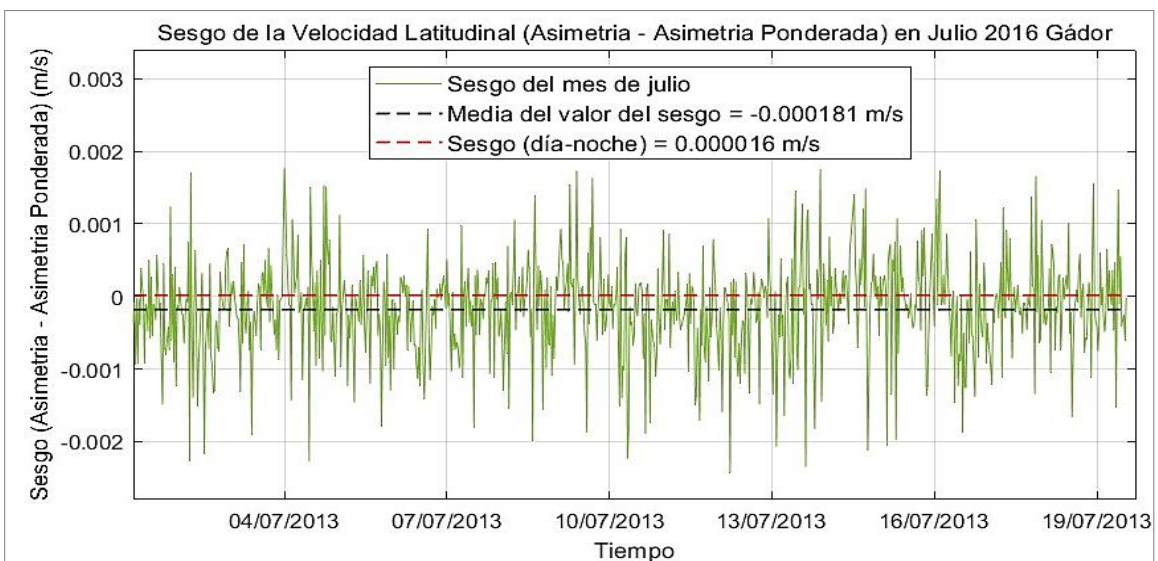


Figura 13. Sesgo producido en la asimetría de la velocidad latitudinal en julio de 2016 en Gádor. Se muestra el sesgo producido por la sobrestimación de la asimetría calculada de forma aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color verde que va variando en m/s (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal negra la media de este sesgo en el mes de julio que es de -0,000181 m/s y con una la línea discontinua roja que representa la diferencia entre el sesgo producido durante el día menos el producido en la noche que toma una valor de 0,000016 m/s.

Además, también es notorio el sesgo que se produce en la densidad del  $CO_2$  en el Olivar de Úbeda ya que es de aproximadamente un 5,6% de error con respecto a la medida aritmética, lo cierto es, que ocurre un caso parecido a Gádor, ya que, al tratar con cifras tan pequeñas y cercanas a 0, el porcentaje de error del sesgo se vuelve más grande. Pero, realmente el sesgo es de  $0,0001379 \text{ mg/m}^3$  en la asimetría, con lo cual, es muy pequeño. En la *figura 14*, se muestra la distribución de este para realmente visualizar que los datos son casi simétricos y que el sesgo es mínimo e irrelevante.

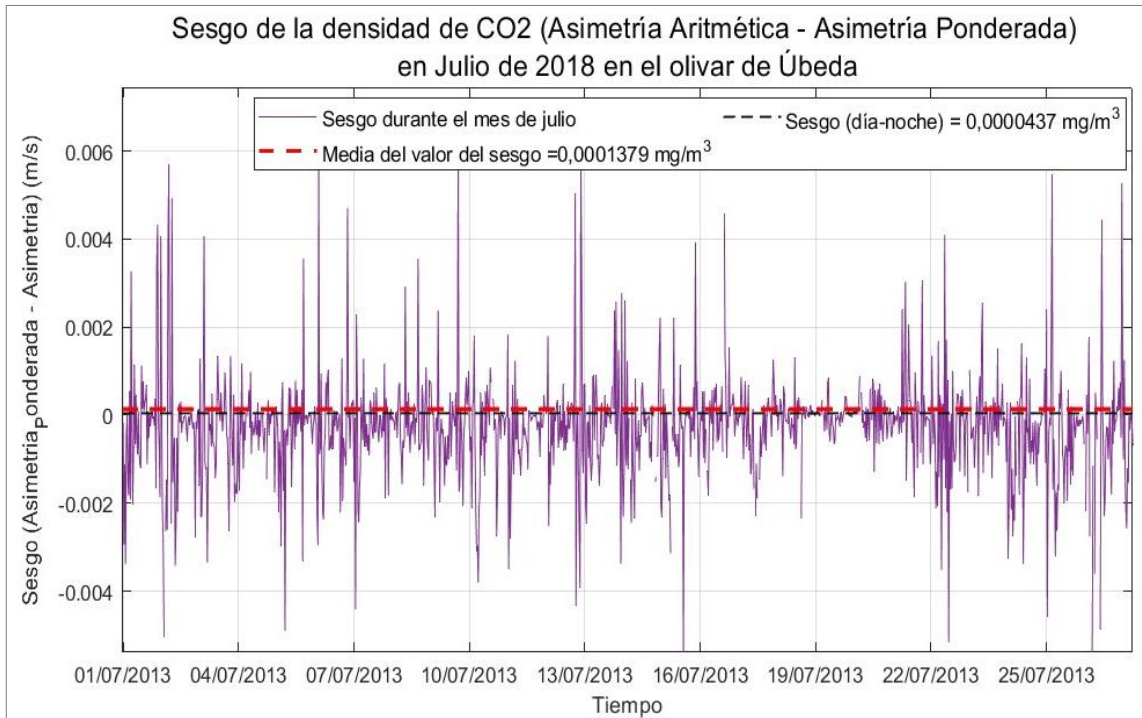


Figura 14. Sesgo producido en la asimetría de la densidad del  $CO_2$  en julio de 2018 en el olivar de Úbeda. Se muestra el sesgo producido por la sobrestimación de la asimetría calculada de forma aritmética frente a tener en cuenta los factores de ponderación. Este sesgo se representa con una línea continua de color morado que va variando en m/s (eje y) frente al tiempo del mes de julio (eje x). Además, en la gráfica, también se representa con una línea discontinua horizontal roja la media de este sesgo en el mes de julio que es de  $0,0000437 \text{ mg/m}^3$  y con una línea discontinua negra que representa la diferencia entre el sesgo producido durante el día menos el producido en la noche que toma un valor de  $0,0001379 \text{ mg/m}^3$

#### 4.4 Sesgo de la curtosis

En la Tabla 4 se muestra un resumen de los resultados obtenidos para la curtosis. Como se explicó en el punto 1.3.1.4, este estadístico indica la prominencia de la distribución en relación con la distribución normal y sus colas. Si tiene un valor de 3, será mesocúrtica, y por tanto tendrá una distribución homogénea.

De forma general, todos los datos siguen una distribución más o menos ‘normal’, en ningún ecosistema y en ningún parámetro medido se sobrepasa un dato de curtosis alto.

En el caso de Balsa Blanca, que es el ecosistema donde se producen mayores sesgos, igualmente continúa ocurriendo esto, teniendo un error de aproximadamente un 7% en la curtosis de la temperatura, pero realmente, al observar los datos de la asimetría no puede ser un resultado concluyente para dar validez a este error.

En cambio, los sesgos que se produjeron en la asimetría de las velocidades del Olivar de Úbeda, también se producen en la curtosis, de hecho, básicamente con el mismo error. Esto es un dato curioso, ya que nos está determinando que realmente el error aumenta y es notorio en los estadísticos de ordenes mayores. Es decir, en la asimetría y en la curtosis, el error se vuelve más evidente que en el caso de la media y de la desviación típica. Esto se debe a que son más sensibles a los valores atípicos o las variaciones y por tanto, se hace más evidente el factor de ponderación. Por ejemplo, el valor atípico en la media tiende a tener un efecto limitado, por el contrario, en la curtosis que se mide el grado de apuntamiento, los valores extremos toman un mayor peso.

RESUMEN DEL ESTADÍSTICO DE CUARTO ORDEN RESPECTO A LA MEDIA (CURTOSIS)					
		PADUL	BALSA BLANCA	GÁDOR	OLIVAR ÚBEDA
Datos micrometeorológicos	Año (julio)	2013	2013	2016	2018
Temperatura (°C)	$K_{s,T}$	2,90028	2,86417	3,0084	3,9368
	$K_{c,T}$	2,90024	2,64303	3,0043	3,9341
	Sesgo	0,00004	0,22114	0,00404	0,00271
	<b>Error (%)</b>	<b>0,001</b>	<b>7,721</b>	<b>0,134</b>	<b>0,069</b>
	S. día-noche	0,00234	0,24995	0,04189	0,00225
Velocidad longitudinal (m/s)	$K_{s,u}$	2,8753	2,9366	3,0546	4,1438
	$K_{c,u}$	2,8750	2,9331	3,0542	4,0375
	Sesgo	0,00028	0,0035265	0,000389	0,10627
	<b>Error (%)</b>	<b>0,010</b>	<b>0,120</b>	<b>0,013</b>	<b>2,565</b>
	S. día-noche	0,00219	0,0099307	-0,000205	0,19764
Velocidad latitudinal (m/s)	$K_{s,v}$	2,9304	2,8948	3,0514	4,1636
	$K_{c,v}$	2,9299	2,8869	3,0513	4,0471
	Sesgo	0,00049	0,0079312	0,000079	0,11657
	<b>Error (%)</b>	<b>0,0167</b>	<b>0,2740</b>	<b>0,0026</b>	<b>2,7997</b>
	S. día-noche	0,00109	0,015529	-0,000233	0,21589
Velocidad vertical (m/s)	$K_{s,w}$	4,5154	3,9727	4,755	5,2098
	$K_{c,w}$	4,5153	3,9168	4,7487	5,0936
	Sesgo	0,000058	0,055978	0,006317	0,11622
	<b>Error (%)</b>	<b>0,0013</b>	<b>1,4091</b>	<b>0,1328</b>	<b>2,2308</b>
	S. día-noche	0,007851	0,017996	-0,002378	0,21589
Densidad de $CO_2$ ( $mg/m^3$ )	$K_{s,CO_2}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			4,2654
	$K_{c,CO_2}$				4,2176
	Sesgo				0,04776
	<b>Error (%)</b>				<b>1,1197</b>
	S. día-noche				0,8308
Densidad de $H_2O$ ( $g/m^3$ )	$K_{s,q}$	Incorporado el factor de ponderación y por tanto, no se puede analizar el sesgo			2,8745
	$K_{c,q}$				2,8735
	Sesgo				0,00107
	<b>Error (%)</b>				<b>0,037</b>
	S. día-noche				0,0010211

Tabla 4. RESUMEN DEL ESTADÍSTICO DE CUARTO ORDEN RESPECTO A LA MEDIA (CURTOSIS) . En esta tabla, se representa tanto los resultados de la curtosis aritmética ( $K_s$ ) como la curtosis ponderada ( $K_c$ ) de los distintos datos micrometeorológicos. La línea de ‘Sesgo’ corresponde a la resta de los de la curtosis aritmética menos la curtosis ponderada de cada uno de los parámetros medidos individualmente. La línea coloreada con fondo azul determina el error porcentual con respecto a la sobrestimación de la medida aritmética. En color rojo se recalcan los errores más grandes. La franja denominada como ‘S.día-noche’ es la diferencia que hay entre el sesgo de día menos el sesgo producido por la noche.

En el resto de los datos, no se produce un error grande, son irrelevantes. Aún así, en todos los casos y en todos los estadísticos se ha comprobado que realmente existe un sesgo, es decir que el muestreo de los datos en la atmósfera no es aleatorio.

## 5. CONCLUSIONES

Tras realizar un análisis exhaustivo de los momentos estadísticos en cuatro ecosistemas diferentes durante el mes de julio, por la necesidad de comprender y corregir el sesgo que resulta al no ser un muestreo aleatorio el tomado. Resulta las siguientes conclusiones:

- El muestreo en la atmósfera no es aleatorio, se cumple que realmente los datos dependen de una ponderación.
- Existe un sesgo muestral producido entre las medidas aritméticas y ponderadas en todos los parámetros medidos. Siendo siempre el resultado de las medias aritméticas sobrestimado con respecto al ponderado.
- Hay una diferencia entre el sesgo producido por el día y por la noche en varios parámetros. En el caso de la temperatura, el sesgo es muy notorio debido a que durante el día, el forzamiento solar produce variaciones de temperatura y convección, influyendo estas fluctuaciones en la temperatura haciendo que las variaciones en el volumen de una parcela (Ley Charles) sean más importantes. También ocurre en las densidades de  $CO_2$  y vapor de agua probablemente por los factores antropogénicos, humedad relativa, presión atmosférica y temperatura entre otros.
- En el caso de la velocidad vertical, en el estadístico de la media, se muestra una ligera tendencia a tener mayor sesgo que en el resto de las velocidades, probablemente porque sus fluctuaciones son muy grandes relativas a su valor promedio
- Se puede intuir debido a la diversidad de medidas y ecosistemas, que al tener un menor número de datos de muestra, la probabilidad de mayor error aumenta, ya que los datos tienen un mayor impacto en los estadísticos.
- En el caso del Olivar de Úbeda se hace evidente que en la asimetría y en la curtosis, el error se vuelve más grande que en el caso de la media y de la desviación típica. Esto se debe a que son más sensibles a los valores atípicos o las variaciones y por tanto, se hace más evidente el factor de ponderación.
- De forma general, se concluye que el sesgo debido a la falta de ponderación no es de gran relevancia. Si se busca una exactitud en los datos, es recomendable ponderar. Pero debido a la dificultad de tratar los datos ponderados, se hace irrelevante.
- Al carecer de investigaciones de este mismo sesgo se hace imposible una comparativa con otros autores, por tanto, es importante continuar con este tipo de estudios para buscar errores que se están cometiendo en los cálculos y así, acercarnos más a una meteorología más precisa.

## **6. AGRADECIMIENTOS**

Transmitir mis sinceros agradecimientos a todos aquellos que me han ayudado en esta etapa y han hecho posible esta investigación.

En primer lugar, agradecer a mi tutor, Andrew Kowalski ya que la dificultad de una investigación completamente nueva y novedosa se ha visto compensada con su apoyo, paciencia y amabilidad en la explicación de cada una de las dudas y conceptos claves necesarios. Y sobre todo, por su rapidez y disponibilidad para responder cada uno de los mensajes lo antes posible.

En segundo lugar, a mi cotutora Matilde García-Valdecasas Ojeda por ayudarme en cada una de las tutorías a aprender a programar e intentar salir adelante con cada una de mis dificultades.

Por supuesto, a Penélope Serrano Ortiz por proporcionarme todos los datos para esta investigación y además, ser una gran profesora.

A mi familia, amigos y compañeros, que han sido el apoyo diario en toda la etapa universitaria, que siempre me han animado y apoyado en todo momento a seguir adelante.

A todos los profesores del máster que han hecho posible mi aprendizaje formativo.

Agradecer a la Universidad de Granada, por prestarme sus aulas, su material y su tiempo para hacer posible este proyecto y sobre todo por hacerme sentir como en casa.

Muchas gracias a todos vosotros, por hacerme sentir orgullosa de lo que soy, de lo que quiero ser en un futuro y de este proyecto, por conseguir que me encante la meteorología y por ello, esté opositando para la Agencia Estatal de Meteorología.



## 7. REFERENCIAS BIBLIOGRÁFICAS

- Acosta, M., Dušek, J., Chamizo, S., Serrano-Ortiz, P., & Pavelka, M. (2019). Autumnal fluxes of CH<sub>4</sub> and CO<sub>2</sub> from Mediterranean reed wetland based on eddy covariance and chamber methods. *Catena*, 183, 104191.
- Alberghi, S., Maurizi, A., & Tampieri, F. (2002). Relationship between the vertical velocity skewness and kurtosis observed during sea-breeze convection. *Journal of Applied Meteorology and Climatology*, 41(8), 885-889.
- American Meteorological Society. (2015). *Glossary of meteorology* (2nd ed.). American Meteorological Society.
- Aranda-Barranco, S., Serrano-Ortiz, P., Kowalski, A. S., & Sánchez-Cañete, E. P. (2023). The temporary effect of weed-cover maintenance on transpiration and carbon assimilation of olive trees. *Agricultural and Forest Meteorology*, 329, 109266.
- Arjona, L. (s.f.). De pajareo en Cortijo Guadiana. Olivares Vivos. Recuperado de: <https://www.olivaresvivos.com/de-pajareo-en-cortijo-guadiana/>
- Arnold, B. C., & Groeneveld, R. A. (1995). Measuring skewness with respect to the mode. *The American Statistician*, 49(1), 34-38.
- Chamizo, S., Serrano-Ortiz, P., López-Ballesteros, A., Sánchez-Cañete, E. P., Vicente-Vicente, J. L., & Kowalski, A. S. (2017). Net ecosystem CO<sub>2</sub> exchange in an irrigated olive orchard of SE Spain: Influence of weed cover. *Agriculture, Ecosystems & Environment*, 239, 51-64.
- Checa-Garcia, R. (2017). *Statistics for Weather and Climate: Introduction*.
- Chen, J., Brissette, F. P., Lucas-Picher, P., & Caya, D. (2017). Impacts of weighting climate models for hydro-meteorological climate change studies. *Journal of Hydrology*, 549, 534-546.
- Doane, D. P., & Seward, L. E. (2011). Measuring skewness: a forgotten statistic? *Journal of statistics education*, 19(2).
- Franzese, M., y Iuliano, A. (2019). *Descriptive Statistics*.
- Hollinger, D. Y., y Richardson, A. D. (2005). Uncertainty in eddy covariance measurements and its application to physiological models. *Tree physiology*, 25(7), 873-885.

- IFAPA- Instituto Andaluz de Investigación y Formación Agraria, Pesquera, Alimentaria y de la producción ecológica [20/04/2023 a las 19:00 horas]
- Kaur, P., Stoltzfus, J., y Yellapu, V. (2018). Descriptive statistics. *International Journal of Academic Medicine*, 4(1), 60.
- Kowalski, A. S. (2012). Exact averaging of atmospheric state and flow variables. *Journal of the Atmospheric Sciences*, 69(5), 1750-1757.
- Kowalski, A., y Ojeda, M. G. V. (2022). Widespread sampling bias in atmospheric data.
- Lee, D. K., In, J., & Lee, S. (2015). Standard deviation and standard error of the mean. *Korean journal of anesthesiology*, 68(3), 220-223.
- Lenschow, D. H., Wulfmeyer, V., & Senff, C. (2000). Measuring second-through fourth-order moments in noisy data. *Journal of Atmospheric and Oceanic technology*, 17(10), 1330-1347.
- López Ballesteros, A. Carbon balance and determinant processes in semiarid ecosystems of southeast Spain. Granada: Universidad de Granada, 2017.
- López-Ballesteros, A., Oyonarte, C., Kowalski, A. S., Serrano-Ortiz, P., Sánchez-Cañete, E. P., Moya, M. R., y Domingo, F. (2018). Can land degradation drive differences in the C exchange of two similar semiarid ecosystems? *Biogeosciences*, 15(1), 263-278.
- López-Ballesteros, A., Serrano-Ortiz, P., Sánchez-Cañete, E. P., Oyonarte, C., Kowalski, A. S., Pérez-Priego, Ó., y Domingo, F. (2016). Enhancement of the net CO<sub>2</sub> release of a semiarid grassland in SE Spain by rain pulses. *Journal of Geophysical Research: Biogeosciences*, 121(1), 52-66.
- Mortarini, L., Ferrero, E., Richiardone, R., Falabino, S., Anfossi, D., Trini Castelli, S., y Carretto, E. (2009). Assessment of dispersion parameterizations through wind data measured by three sonic anemometers in a urban canopy. *Advances in Science and Research*, 3(1), 91-98.
- Nick, T. G. (2007). Descriptive statistics. *Topics in biostatistics*, 33-52
- Ostashev, V. E., y Wilson, D. K. (2000). Relative contributions from temperature and wind velocity fluctuations to the statistical moments of a sound field in a turbulent atmosphere. *Acta Acustica united with Acustica*, 86(2), 260-268.
- Pérez, J. (2016). Estadística descriptiva e inferencial.

- Rai, R. K., Berg, L. K., Kosović, B., Mirocha, J. D., Pekour, M. S., & Shaw, W. J. (2017). Comparison of measured and numerically simulated turbulence statistics in a convective boundary layer over complex terrain. *Boundary-Layer Meteorology*, 163, 69-89.
- Ríos, L. G., e Iza, Y. M. (2007). Gases ideales: diagramas termodinámicos. *Scientia et technica*, 1(35).
- Royston, P. (1992). Which measures of skewness and kurtosis are best?. *Statistics in Medicine*, 11(3), 333-343.
- Sánchez-Cañete, E. P. (2009). Comportamiento de los flujos gaseosos de CO<sub>2</sub> en el suelo de un ecosistema kárstico. Factores que afectan a su ventilación.
- Sánchez-Cañete, E. P., Kowalski, A. S., Serrano-Ortiz, P., Pérez-Priego, O., y Domingo, F. (2013). Deep CO<sub>2</sub> soil inhalation/exhalation induced by synoptic pressure changes and atmospheric tides in a carbonated semiarid steppe. *Biogeosciences*, 10(10), 6591-6600.
- Serrano-Ortiz, P., Aranda-Barranco, S., López-Ballesteros, A., Lopez-Canfin, C., Sánchez-Cañete, E. P., Mejjide, A., & Kowalski, A. S. (2020). Transition period between vegetation growth and senescence controlling interannual variability of C fluxes in a Mediterranean reed wetland. *Journal of Geophysical Research: Biogeosciences*, 125(1), e2019JG005169.
- Serrano-Ortiz, P., Oyonarte, C., Pérez-Priego, O., Reverter, B. R., Sánchez-Cañete, E. P., Were, A., ... y Domingo, F. (2014). Ecological functioning in grass–shrub Mediterranean ecosystems measured by eddy covariance. *Oecologia*, 175, 1005-1017
- Tipler, P. A., y Mosca, G. (2021). Física para la ciencia y la tecnología, Vol. 1A: Mecánica. Reverté.
- Turney, S. (2022). What Is Kurtosis? | Definition, Examples & Formula. Scribbr.
- Vinnichenko, N. (2013). Turbulence in the free atmosphere. Springer Science & Business Media.
- Xu, Y., Knudby, A., Ho, H. C., Shen, Y., & Liu, Y. (2017). Warming over the Tibetan Plateau in the last 55 years based on area-weighted average temperature. *Regional Environmental Change*, 17, 2339-2347.
- Yeh, T. C. (1949). On energy dispersion in the atmosphere. *Journal of Atmospheric Sciences*, 6(1), 1-16.

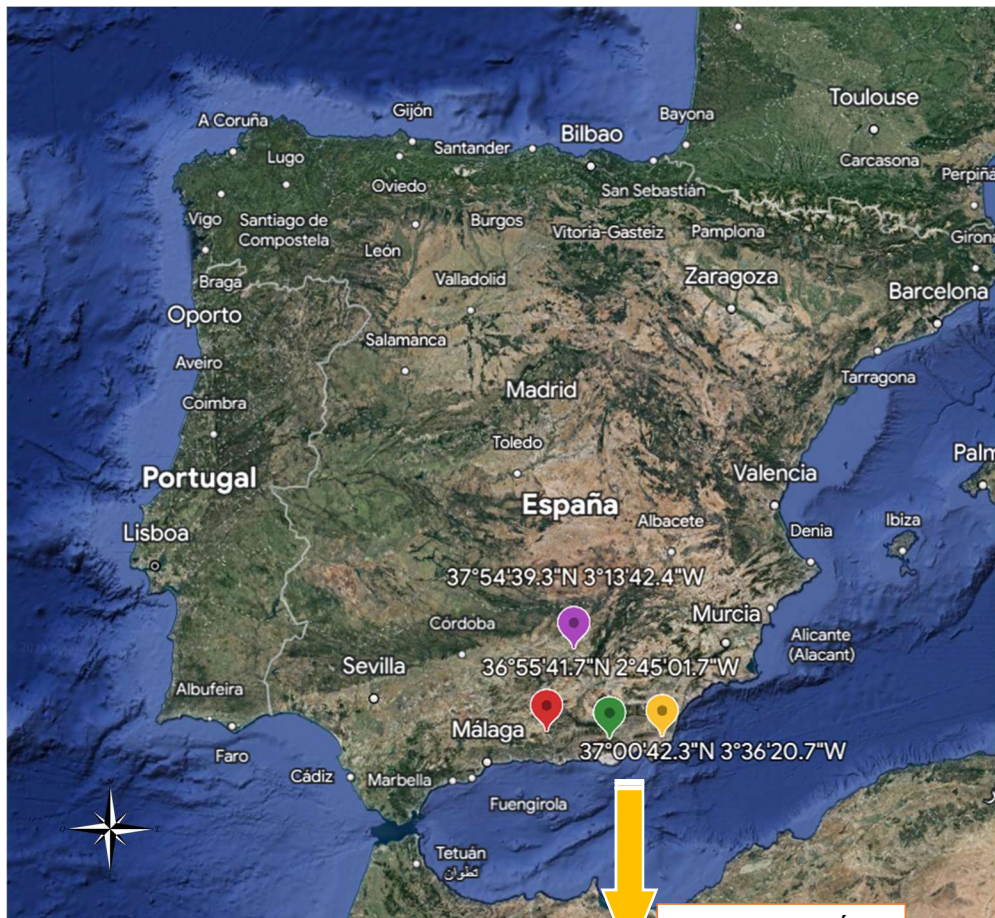
## ANEXO I. FORMULARIO

Mención	Nombre	Ecuación	Parámetros
[Ec.1]	Ley general de los gases ideales	$PV=nRT$	<ul style="list-style-type: none"> <li>- P=presión</li> <li>- V= volumen</li> <li>- n= cantidad de sustancia</li> <li>- R = constante de los gases</li> <li>- T= temperatura</li> </ul>
[Ec.2]	Media sesgada (media aritmética)	$\mu_s = \frac{1}{N} \sum_{i=1}^N x_i$	<ul style="list-style-type: none"> <li>- <math>\mu_s</math> = media sesgada (media aritmética)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math>= datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> </ul>
[Ec.3]	Varianza sesgada (varianza aritmética)	$\sigma_s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_s)^2$	<ul style="list-style-type: none"> <li>- <math>\sigma_s^2</math>= Varianza sesgada (varianza aritmética)</li> <li>- <math>\mu_s</math> = media sesgada (media aritmética)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math>= datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> </ul>
[Ec.4]	Desviación estándar sesgada (desviación estándar aritmética)	$\sigma_s = \sqrt{\sigma_s^2}$	<ul style="list-style-type: none"> <li>- <math>\sigma_s</math>= desviación estándar sesgada (desviación estándar aritmética)</li> <li>- <math>\sigma_s^2</math>= varianza sesgada (varianza aritmética)</li> </ul>
[Ec.5]	Skewness sesgado (skewness aritmético)	$E_s = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_s)^3}{(\sigma_s)^3}$	<ul style="list-style-type: none"> <li>- <math>E_s</math>= Asimetría o skewness sesgado (skewness aritmético)</li> <li>- <math>\sigma_s</math>= desviación estándar sesgada (desviación estándar aritmética)</li> <li>- <math>\mu_s</math> = media sesgada (media aritmética)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math>= datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> </ul>
[Ec.6]	Kurtosis sesgado (kurtosis aritmético)	$K_s = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_s)^4}{(\sigma_s)^4}$	<ul style="list-style-type: none"> <li>- <math>K_s</math>= Kurtosis sesgado (kurtosis aritmético)</li> <li>- <math>\sigma_s</math>= desviación estándar sesgada (desviación estándar aritmética)</li> <li>- <math>\mu_s</math> = media sesgada (media aritmética)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math>= datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> </ul>
[Ec.7]	Media corregida (media ponderada)	$\mu_c = \frac{\frac{1}{N} \sum_{i=1}^N x_i \cdot a_i}{\frac{1}{N} \sum_{i=1}^N a_i}$	<ul style="list-style-type: none"> <li>- <math>\mu_c</math> = media corregida (media ponderada)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math>= datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> <li>- <math>a_i</math> = peso asignado para cada dato de temperatura</li> </ul>

Mención	Nombre	Ecuación	Parámetros
[Ec.8]	Desviación estándar corregida (desviación estándar ponderada)	$\sigma_c = \sqrt{\frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_c)^2 \cdot a_i}{\frac{1}{N} \sum_{i=1}^N a_i}}$	<ul style="list-style-type: none"> <li>- <math>\sigma_c</math> = Desviación estándar corregida (desviación estándar ponderada)</li> <li>- <math>\mu_c</math> = media corregida (media ponderada)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math> = datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> <li>- <math>a_i</math> = peso asignado para cada dato de temperatura</li> </ul>
[Ec.9]	Skewness corregido (skewness ponderado)	$E_c = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_c)^3 \cdot a_i}{\frac{1}{N} (\sigma_c)^3 \cdot a_i}$	<ul style="list-style-type: none"> <li>- <math>E_c</math> =Skewness corregido (skewness ponderado)</li> <li>- <math>\mu_c</math> = media corregida (media ponderada)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math> = datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> <li>- <math>\sigma_c</math> = desviación estándar corregida (desviación estándar ponderada)</li> <li>- <math>a_i</math> = peso asignado para cada dato de temperatura</li> </ul>
[Ec.10]	Kurtosis corregido (kurtosis ponderado)	$K_c = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \mu_c)^4 \cdot a_i}{\frac{1}{N} (\sigma_c)^4 \cdot a_i}$	<ul style="list-style-type: none"> <li>- <math>K_c</math> = Kurtosis corregido (kurtosis ponderado)</li> <li>- <math>\mu_c</math> = media corregida (media ponderada)</li> <li>- N= número total de datos</li> <li>- <math>x_i</math> = datos micrometeorológicos tomados por torres Eddy covariance (<math>x_1, x_2, x_3 \dots x_N</math>)</li> <li>- <math>\sigma_c</math> = desviación estándar corregida (desviación estándar ponderada)</li> <li>- <math>a_i</math> = peso asignado para cada dato de temperatura</li> </ul>
[Ec.11]	Temperatura ponderada	$\bar{T} \equiv \frac{\frac{1}{N} \sum_{i=1}^N \rho_i C_{p,i} T_i}{\frac{1}{N} \sum_{i=1}^N \rho_i C_{p,i}}$	<ul style="list-style-type: none"> <li>- <math>\bar{T}</math> = temperatura ponderada</li> <li>- N= número total de datos</li> <li>- <math>T_i</math> = datos tomados de temperatura</li> <li>- <math>\rho_i</math> = densidad del aire seco</li> <li>- <math>C_{p,i}</math> = calor específico</li> </ul>
[Ec.12]	Velocidades ponderadas	$\tilde{u}, \tilde{v}, \tilde{w} \equiv \frac{\frac{1}{N} \sum_{i=1}^N \rho_i (u, v, w)_i}{\frac{1}{N} \sum_{i=1}^N \rho_i}$	<ul style="list-style-type: none"> <li>- <math>\tilde{u}, \tilde{v}, \tilde{w}</math> = velocidad longitudinal (u), latitudinal (v) y vertical (w)</li> <li>- N= número total de datos</li> <li>- <math>(u, v, w)_i</math> = datos tomados de velocidad longitudinal (<math>u_i</math>), latitudinal (<math>v_i</math>) y vertical (<math>w_i</math>).</li> <li>- <math>\rho_i</math> = densidad del aire seco</li> </ul>
[Ec.13]	Densidades ponderadas	$\tilde{c}, \tilde{q} \equiv \frac{\frac{1}{N} \sum_{i=1}^N \rho_i (c, q)_i}{\frac{1}{N} \sum_{i=1}^N \rho_i}$	<ul style="list-style-type: none"> <li>- <math>\tilde{c}, \tilde{q}</math> = densidad de <math>CO_2</math> (<math>\tilde{c}</math>) y densidad de <math>H_2O</math> (<math>\tilde{q}</math>)</li> <li>- N= número total de datos</li> <li>- <math>(c, q)_i</math> = datos tomados de densidad de <math>CO_2</math> (<math>c_i</math>) y densidad de <math>H_2O</math> (<math>q_i</math>).</li> <li>- <math>\rho_i</math> = densidad del aire seco</li> </ul>

## ANEXO II. UBICACIÓN DE LAS ZONAS DE ESTUDIO

Con un indicador rojo se muestra Padul, amarillo Balsa Blanca, verde Gádor y de color morado el olivar ubicado en Úbeda:



AMPLIACIÓN



## ANEXO III. EJEMPLO DE DESARROLLO DE SCRIPT DE MATLAB

Dentro del programa se desarrolló el siguiente script para la temperatura:

1. Primero se permite dar opciones al usuario para futuras investigaciones o comprobaciones de otros meses del año. En este caso para este proyecto se utilizó el mes de Julio, correspondiente al mes 7. Se pasa al directorio de trabajo que se esté utilizando en este momento, en el ejemplo es el del olivar del Cortijo Guadiana. Por último se determina un umbral para los futuros paquetes de datos de medias horas, para que al menos tengan el 90% de los datos y asegurarnos así, que son óptimos.

```
clear; clc

% OPCIONES DE USUARIO%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
MES = '07'; % número del mes en formato string
mes = 7;
iniday = 1; %día de inicio
endday = 31; %día de final |
cd('Olivar_Guadiana\')%se cambia del directorio actual a Padul,
% Balsa Blanca, Gádor o Olivar de Guadiana

umbral = 0.9*18000; % el 90% de los datos de media hora deben estar
%para considerarse óptimos
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

2. A continuación, se seleccionan los archivos del mes (file\_month) que son 18000 (como máximo). Luego, se generan los intervalos de tiempo, es decir los timestamps (dt) y por último, antes de iniciar el bucle, se crea una matriz de datos del mes con NaNs que se irá rellenando con los datos correspondientes.

```
% se seleccionan los archivos correspondientes al mes seleccionado:
file_month = dir(['TOA5_5726.ts_data_2018_' convertStringsToChars(MES) '_*.dat']);

% se generan los timestamps correspondientes a los datos a analizar:
dt = (datenum(2018,mes,iniday,0,0,0):1/48:datenum(2018,mes,endday,23,30,0))';

% se inicializa la matriz de datos del mes con NaNs para luego ir llenándola con
% los datos correspondientes
data_month = NaN(size(file_month,1),10);
```

3. Se crea un bucle for para iterar sobre el vector file\_month y realizar los cálculos para cada elemento de file\_month(i). Se almacenan los datos de un archivo de tabla especificado por file\_month(i) y se almacenan en la variable x. Se extraen porciones de las columnas "Ts", "diag\_csar" y "diag\_irga\_raw" de la tabla 'x' y se asignan a las variables t, diag y diag\_irga, respectivamente. Después, con estas condiciones se crean máscaras (mascara\_1 y mascara\_2) para cada condición, que resultaran en una máscara final (mascara) donde se cumplen las dos condiciones y elimina el resto de los datos. Y además, se realizan conversiones de algunas variables (presión, temperatura, densidad y masa molar aparente).

```

tic
for i=1:size(file_month,1)
    % CONSTANTES
    R = 8.314; % J/(K mol)
    % Crear mascara para el diagnóstico de las variable
    x = readtable(file_month(i).name);
    t = x.Ts_Avg(3:end,:);
    diag = x.diag_csat_Avg(3:end,:);%utilizamos cellfun
    % para poder convertir de celda a numero
    mascara= diag == 0;
    %hacemos el mismo procedimiento para no operar con celda

    % conversión de variables
    % Paso de Kpa a Pa y cambio lo de celda:
    P = x.Ptotal_Avg(3:end,:)*1000;
    % Paso de g/m³ a kg/m³ densidad del vapor
    rho_v = x.H2O_Avg(3:end,:)/1000;
    T = t+ 273.15; % Paso de °C a °K
    M = P./(R*T);

```

4. A continuación se le añade NaN a los datos cuya mascara tenga 0. Después, se calculan los estadísticos que se colocarán en cada columna de la matriz:
- $\text{data\_month}(i,1) = \text{dt}(i)$ . **Primera columna.** Media de temperatura de los 18000 datos agrupados cada media hora.
  - $\text{data\_month}(i,2) = \text{sum}(\sim\text{isnan}(T))$ . **Segunda columna.** Número de datos en el paquete de media hora (quitando los NaN).
  - $\text{data\_month}(i,3) = \text{nanmean}(T)$ . **Tercera columna.** Se calcula la media aritmética de la temperatura, como en la ecuación Ec.4.
  - $\text{data\_month}(i,4) = \text{nansum}((M.*T)/\text{nansum}(M))$ . **Cuarta columna.** Se calcula la media ponderada de la temperatura, como en la ecuación Ec.9.
  - $\text{vararit}(i,1) = \text{nansum}(((T-\text{data\_month}(i,3)).^2)/\text{data\_month}(i,2))$ .  
No se asigna a ninguna columna pero se calcula la varianza como en la ecuación Ec.5 para facilitar los cálculos posteriores.
  - $\text{vararitpon}(i,1) = \text{nansum}(((M.*(T-\text{data\_month}(i,4)).^2)/\text{data\_month}(i,2)) / (\text{nansum}(M)/\text{data\_month}(i,2)))$ . No se asigna a ninguna columna pero se calcula la varianza ponderada para facilitar los cálculos posteriores.
  - $\text{data\_month}(i,5) = \text{sqrt}(\text{vararit}(i))$ . **Quinta columna.** Desviación aritmética de la temperatura como en la ecuación Ec.6.
  - $\text{data\_month}(i,6) = \text{sqrt}(\text{vararitpon}(i))$ . **Sexta columna.** Desviación estándar ponderada para la temperatura como en la ecuación Ec.10.
  - $\text{data\_month}(i,7) = \text{nansum}(((T-\text{data\_month}(i,3)).^3)/\text{data\_month}(i,2)) / \text{data\_month}(i,5)^3$ . **Séptima columna.** Asimetría o coeficiente de Fisher de la temperatura de forma aritmética como en la ecuación Ec.7
  - $\text{data\_month}(i,8) = \text{nansum}(((M.*(T-\text{data\_month}(i,4)).^3)/\text{data\_month}(i,2)) / (\text{nansum}(M) / \text{data\_month}(i,2))) / \text{data\_month}(i,5)^3$ . **Octava columna.** Asimetría o coeficiente de Fisher de la temperatura de forma ponderada donde se calcula la asimetría ponderada como en la ecuación Ec.11.
  - $\text{data\_month}(i,9) = \text{nansum}(((T-\text{data\_month}(i,3)).^4)/\text{data\_month}(i,2)) / \text{data\_month}(i,5)^4$ . **Novena columna.** Curtosis o coeficiente de Pearson de la temperatura de forma aritmética como en la ecuación Ec.8.
  - $\text{data\_month}(i,10) = \text{nansum}(((M.*(T-\text{data\_month}(i,4)).^4)/\text{data\_month}(i,2)) / (\text{nansum}(M) / \text{data\_month}(i,2))) / \text{data\_month}(i,5)^4$ . **Décima columna.** Curtosis o coeficiente de Pearson de la temperatura de forma ponderada como en la ecuación Ec.12.



Antes de cerrar el bucle for, se añaden también los umbrales que debe permitir y descartar.

```
% Añadimos máscara a los datos
T(mascara == 0) = NaN;
data_month(i,1) = dt(i); % fecha
data_month(i,2) = sum(~isnan(T)); % n
data_month(i,3) = nanmean(T); % Media aritmética
data_month(i,4) = nansum((M.*T)/nansum(M)); % Media ponderada
vararit(i,1)=nansum(((T-data_month(i,3)).^2)/data_month(i,2)); % Varianza
vararitpon(i,1)=nansum(((M.*(T-data_month(i,4)).^2)/data_month(i,2))/(nansum(M)/data_month(i,2)));
% Varianza ponderada
data_month(i,5) = sqrt(vararit(i)); % Desviación estándar
data_month(i,6) = sqrt(vararitpon(i)); % Desviación estándar ponderada
data_month(i,7) = nansum((((T-data_month(i,3)).^3)/data_month(i,2))/data_month(i,5)^3); %Asimetría
data_month(i,8) = nansum(((M.*(T-data_month(i,4)).^3)/data_month(i,2))/(nansum(M)/data_month(i,2))/data_month(i,5)^3);
% Asimetría ponderada
data_month(i,9) = nansum((((T-data_month(i,3)).^4)/data_month(i,2))/data_month(i,5)^4); % Curtosis
data_month(i,10) = nansum(((M.*(T-data_month(i,4)).^4)/data_month(i,2))/(nansum(M)/data_month(i,2))/data_month(i,5)^4);
% Curtosis ponderada

if data_month(i,2)< umbral
    data_month(i,4) = NaN;
    data_month(i,5) = NaN;
end
clear n
```

5. Por último, se ajusta la tabla para que se rellenen adecuadamente los datos y se modifica la fecha para conseguir el formato apropiado.

```
%% guardar en formato .mat
cd('..datos_tratados_guadiana/');
% Crear tabla a partir de mi matriz
data_month_table = array2table(data_month, 'VariableNames', {'Fecha','n','Media_aritmetica','Media_ponderada','Desviacion','Desviacion_ponderada','Asimetria','Asimetria_ponderada','Curtosis','Curtosis_ponderada'});
% Cambiar formato de fecha a algo entendible
%data_month_table.Fecha = datetime(data_month(:,1));
data_month_table.Fecha = datetime(data_month(:,1), 'ConvertFrom','datenum');

save(['TOA5_PAD_2013_' convertStringsToChars(MES) '.mat'],'data_month_table'); %guardar en formato MAT
```

El resultado que se obtiene de este script da una tabla con el número de filas que se hayan conseguido tras el filtrado y 10 columnas con formato como a continuación:

	1	2	3	4	5	6	7	8	9	10
	Fecha	n	Media_aritmetica	Media_ponderada	Desviacion	Desviacion_ponderada	Asimetria	Asimetria_ponderada	Curtosis	Curtosis_ponderada
1	01-Jul-2013 00:00:...	17996	294.2861	294.2525	0.5578	0.5587	-5.2614e-04	0.1758	2.3610	2.3806
2	01-Jul-2013 00:30:...	17996	294.2217	294.1889	0.4106	0.4119	-0.2546	-0.0183	2.7539	2.7114
3	01-Jul-2013 01:00:...	17997	294.2887	294.2724	0.2517	0.2522	0.1375	0.3301	3.0495	3.1086
4	01-Jul-2013 01:30:...	17995	294.7553	294.7054	0.4727	0.4754	-0.4440	-0.1304	2.6512	2.5332
5	01-Jul-2013 02:00:...	17993	295.4920	295.4109	0.2619	0.2741	-0.2299	0.7262	2.8164	3.1141
6	01-Jul-2013 02:30:...	17987	295.5835	295.4032	0.2225	0.2862	0.1413	3.0990	2.8249	7.6402
7	01-Jul-2013 03:00:...	17994	295.4189	295.3533	0.2488	0.2572	0.3010	1.1068	3.0186	3.7517
8	01-Jul-2013 03:30:...	17985	295.4811	295.2687	0.2687	0.3422	0.3131	3.1696	3.2547	8.3603
9	01-Jul-2013 04:00:...	17997	294.7838	294.7666	0.4660	0.4662	0.1429	0.2471	3.7181	3.7475
10	01-Jul-2013 04:30:...	17998	294.4540	294.4529	0.5538	0.5538	-0.1086	-0.1080	2.6867	2.6871
11	01-Jul-2013 05:00:...	17991	295.6906	295.5759	0.2856	0.3077	-0.5084	0.7579	3.4606	3.6362
12	01-Jul-2013 05:30:...	17996	296.3447	296.3115	0.2740	0.2760	-0.3169	0.0462	2.0794	2.0147
13	01-Jul-2013 06:00:...	17998	297.2173	297.2167	0.4227	0.4226	0.5004	0.5014	2.2729	2.2719
14	01-Jul-2013 06:30:...	17951	298.6575	297.8780	0.3797	0.8656	-0.4659	14.2927	2.8637	41.9267
15	01-Jul-2013 07:00:...	17960	299.2118	298.5802	0.3847	0.7383	0.1528	9.4687	2.6947	27.0291
16	01-Jul-2013 07:30:...	17981	299.7875	299.5033	0.5044	0.5782	0.1523	2.0130	2.8036	5.1334
17	01-Jul-2013 08:00:...	17998	300.2582	300.2569	0.6432	0.6432	0.1126	0.1132	2.7399	2.7385
18	01-Jul-2013 08:30:...	17998	300.5823	300.5803	0.7641	0.7641	-3.4448e-04	9.1016e-04	2.5030	2.5026

Para la obtención de gráficas se utilizó los comandos facilitados por la propia página oficial de Matlab ([https://es.mathworks.com/help/matlab/creating\\_plots/types-of-matlab-plots.html](https://es.mathworks.com/help/matlab/creating_plots/types-of-matlab-plots.html)).