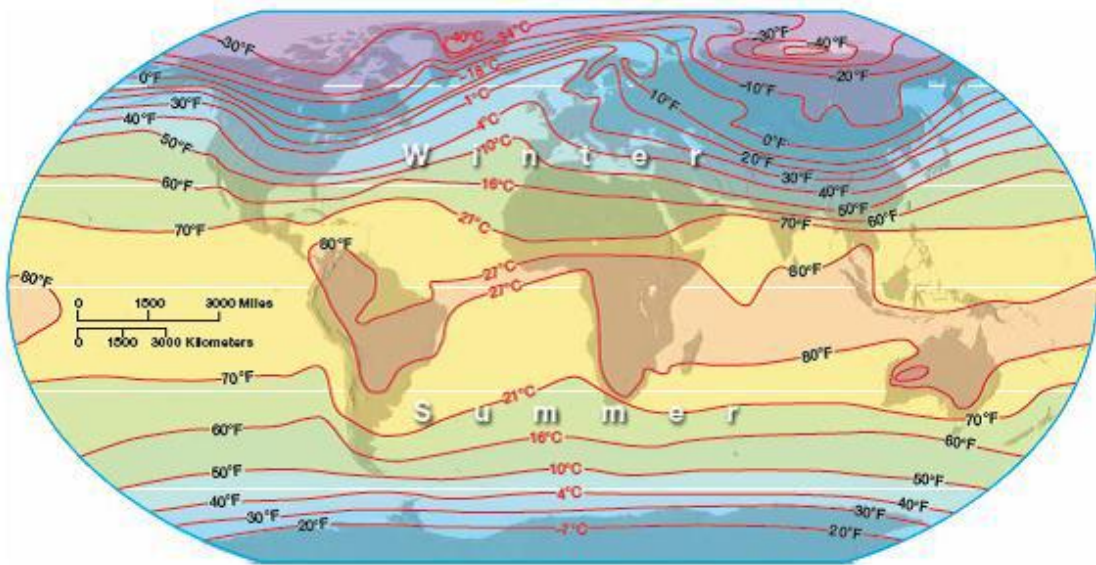


Contrastes de temperaturas en la superficie terrestre



23/12/2020 – 10/01/2021

Javier Olcoz Macayo

mrjolma@gmail.com

THE BRIDGE DIGITAL
TALENT
ACCELERATOR

1. Índice

1. Índice.....	2
2. Resumen y objetivo	3
3. Introducción.....	5
3.1. Introducción climatológica.....	5
3.2. Introducción estadística	8
4. Métodos	9
5. Resultados	10
5.1. Resultados globales diurnos	10
5.2. Resultados globales nocturnos	16
5.3. Resultados por meses.....	22
6. Conclusiones	28

2. Resumen y objetivo

Este proyecto se ha centrado en hacer un tratamiento de datos provenientes de la página oficial de la NASA ‘<https://neo.sci.gsfc.nasa.gov/>’ con el fin de demostrar las siguientes hipótesis:

1. A la hora de estudiar la zona caliente de planeta, es mejor hacerlo con estadísticos robustos (mediana, rango intercuartílico) en vez de usar la media y la desviación típica.
2. La enorme variedad climatológica del hemisferio sur hace que la latitud no tenga la misma importancia para la temperatura que en el hemisferio norte.

Una vez entendidas las hipótesis, se limpiaron los datos y se empezó a obtener resultados acerca de nuestras hipótesis.

Respecto a la hipótesis número 1, se vio que en cuanto a estadísticos de valores centrales nos era indiferente trabajar con mediana o media dado que los resultados era muy similares y los resultados obtenidos de cada uno tenían una correlación de entre 0.98 - 1 para valores diurnos y entre 0.89 - 1 para valores nocturnos. Si bien es cierto que la menor correlación se dio en la zona caliente, llegamos a la conclusión de que no era suficiente para verificar nuestra hipótesis. En cuanto a los estadísticos de dispersión de datos vimos que la diferencia más notable se daba en el Ecuador, por lo que nuestra hipótesis no se cumple para toda la región cálida sino sólo en la ecuatorial. En horario nocturno esto se daba a lo largo de todo el hemisferio sur del planeta.

Respecto a la hipótesis número 2, fue muy sencilla de verificar dado que en cada una de las gráficas obtenidas para el hemisferio sur,

sobre todo entre 30 y 60 grados sur, se veía que no seguía una linealidad muy marcada, a diferencia del hemisferio norte.

3. Introducción

3.1. Introducción climatológica

La temperatura en los últimos 100 años ha cambiado mucho debido al cambio climático, todo esto ha hecho que el estudio de la climatología terrestre se haya convertido en un sector muy importante de cara a ver qué medidas se pueden tomar para que la vida en la Tierra sea lo mejor posible para el ser humano.

Para entender este estudio hay que entender lo que es la temperatura y qué factores la alteran. Se define como temperatura a la energía cinética de las partículas que forman un sistema. Al aumentar esta energía aumentará la temperatura del sistema. La forma más común de aumento de energía en partículas es por colisiones o por incidencia de radiación sobre las mismas, la cual las excita.

La forma de medida más usual de temperatura es a través de instrumentos de medida llamados termómetros.

En cuanto a la temperatura terrestre las variables que más le afectan son: las precipitaciones, la nubosidad, la humedad, el viento (entendiendo al mismo como el desplazamiento de partículas en el aire) y, siendo la más importantes y en la que vamos a hacer hincapié, la radiación solar.

La radiación solar es el conjunto de radiaciones electromagnéticas emitidas por el Sol. Estas llegan a la Tierra y pueden ser reflejadas por la atmósfera o pueden incidir en la misma y llegar a la superficie terrestre, dónde son absorbidas y posteriormente emitidas como

rayos infrarrojos. Esta radiación no es uniforme a lo largo de todo el año sino que varía según el momento en el que se encuentre la Tierra en su órbita de traslación. Esto se debe a que el eje sobre el cual el planeta realiza su rotación no es perpendicular a su órbita alrededor del sol sino que tiene una inclinación de casi 23.5 grados. Así es como se originan las estaciones. Las regiones que reciban más radiación serán las que tengan una mayor temperatura, es el caso de los trópicos (según la estación será uno u otro). La región que recibe una radiación más constante es el Ecuador.

La inclinación sobre el eje de rotación es lo que hace que la temperatura varíe en función de la latitud y es, sobre esta premisa, sobre la que hemos basado y centrado este proyecto.

La radiación emitida por la superficie terrestre como rayos infrarrojos es atrapada por la atmósfera por el muy conocido ‘Efecto Invernadero’ producido por ciertos gases. Este es el mecanismo por el cual la Tierra no se enfría por las noches, momento en el cual no hay incidencia directa de luz solar, hasta el punto de ser inhabitable.

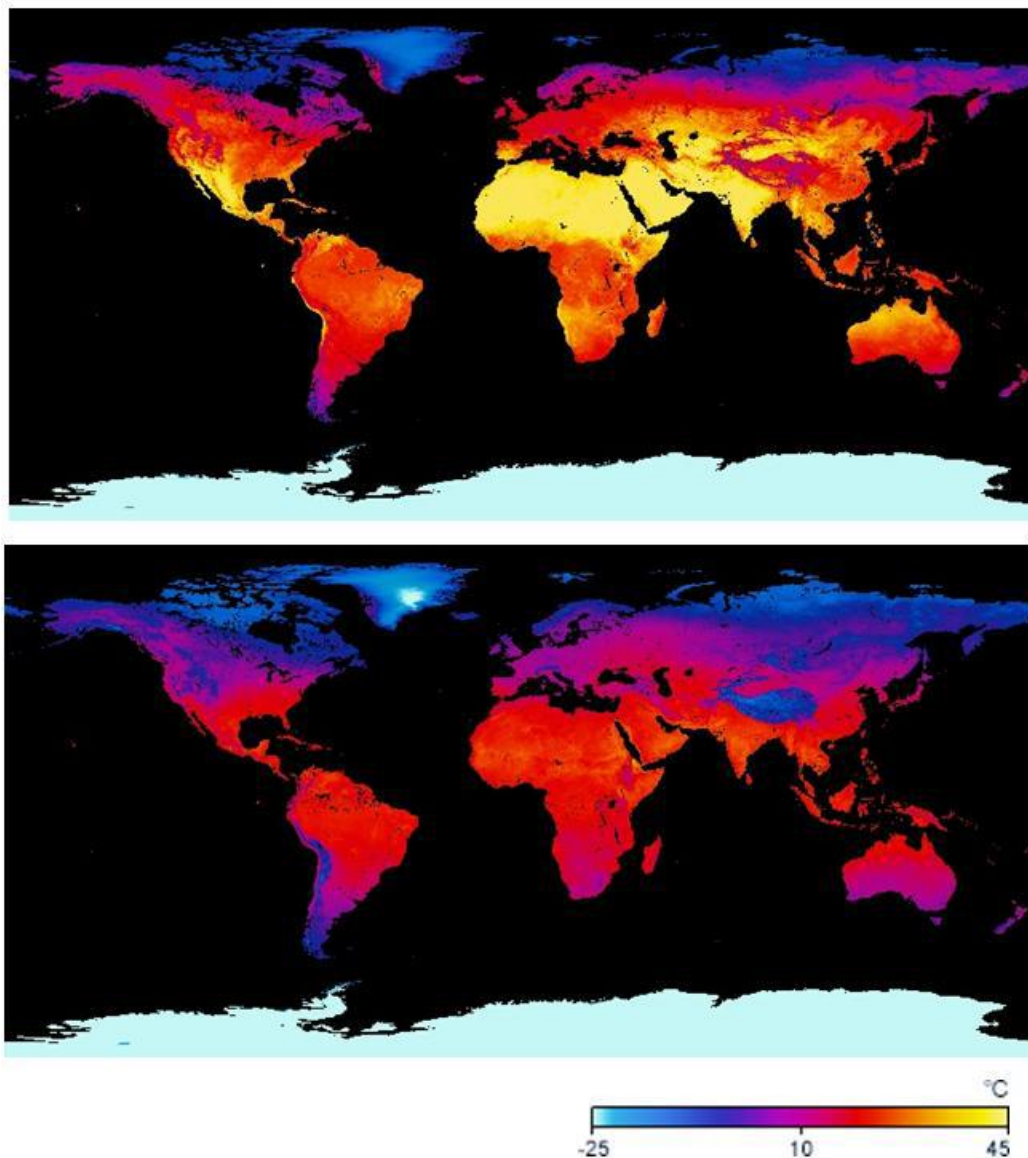


Figura 1. Diferencia de temperatura Mayo 2020 día/noche

En este proyecto estudiaremos ambos casos, valores diurnos y nocturnos, para poder tratar con más solidez las hipótesis planteadas.

El estudio se hará desde el 02/2000 – 12/2019.

3.2. Introducción estadística

Vamos a decir que un estadístico posee mayor robustez cuando las anomalías dentro de su distribución de datos le afectan menos que a otros estadísticos, haciendo así que sus resultados sean más fiables.

Como estadísticos menos robustos usaremos la media, como estadístico promediador, y la desviación estándar, como estadístico de dispersión.

Como estadísticos robustos usaremos la mediana, como estadístico promediador, y el rango intercuartílico, como estadístico de dispersión.

La principal diferencia entre media y mediana es que la media tiene en cuenta los valores de los datos mientras que la mediana solo tiene en cuenta la frecuencia de los datos. La mediana tiene como contra que la distribución de datos tiene que estar ordenada.

Veamos un ejemplo de robustez. Si tenemos una colección de datos tal que la siguiente: $x = (1, 2, 3, 4, 5000)$. La media de x sería 1002 mientras que la mediana sería el valor que ocupa la posición central de la distribución, o sea, 3.

Sobre esto fundamento mi primera hipótesis. En las zonas frías y templadas, las temperaturas varían más con el paso del año por lo que una anomalía no va a notarse tanto como en la zona caliente del planeta, donde la dispersión de datos va a ser menor y por tanto un dato anómalo modifica más el promediado.

4. Métodos

En esta sección se va a explicar brevemente como se ha trabajado.

Una vez descargados los datos y volcados a Excel para comprobar su utilidad, se organizó para tener datos cada 0.1 grados de latitud. Se vio que había una enorme cantidad de datos NaN. Esto es debido a que hay latitudes con poca superficie terrestre y entonces no se puede recoger datos de dichas latitudes. Visto esto se limpió el dataframe para que solo tuviéramos las latitudes con un número de resultados significativos para estudio, 50. Una vez limpiados los datos se calcularon los estadísticos de ambos horarios y se procedió a la visualización e interpretación de resultados.

Posteriormente se quiso realizar lo mismo pero a nivel mensual, para ver como afectaban las estaciones según la latitud.

“El resto de métodos de trabajo vienen especificados en la guía de trabajo del proyecto proporcionada por la academia ‘The Bridge’”

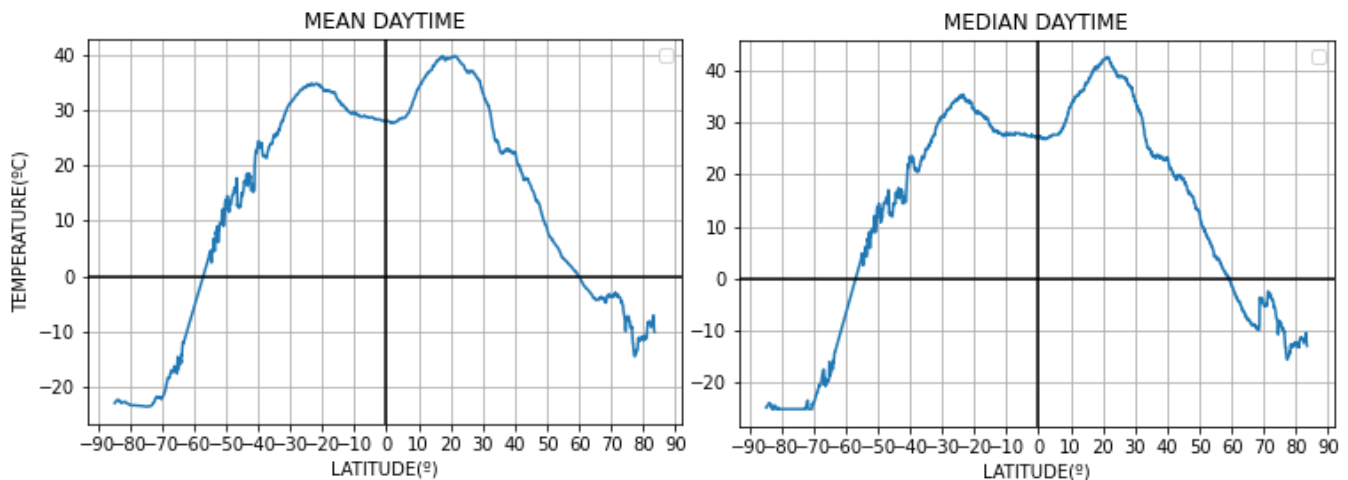
5. Resultados

En esta sección vamos a presentar los resultados globales primero explicando cómo influye cada resultado en nuestras hipótesis.

5.1. Resultados globales diurnos

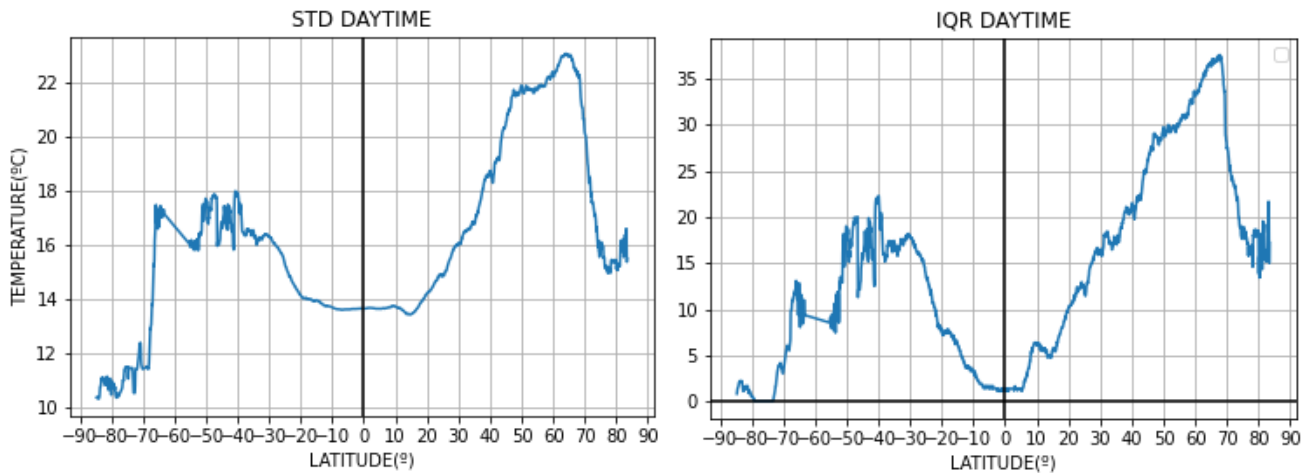
Lo primero que vamos a hacer es mostrar las gráficas obtenidas de los estadísticos estudiados para comparar los resultados robustos con los menos robustos.

Lo primero que vamos a ver es la comparativa entre la media y la mediana para datos diurnos.



Vemos que entre ambas no hay una gran diferencia. Ambas cortan al eje de 0 grados Celsius a 60 grados sur y 60 grados norte. Cada hemisferio tiene su máximo en el trópico pertinente y con valores muy similares. La diferencia más notable está en la forma de los máximos, siendo los de la media menos puntiagudos que los de la mediana.

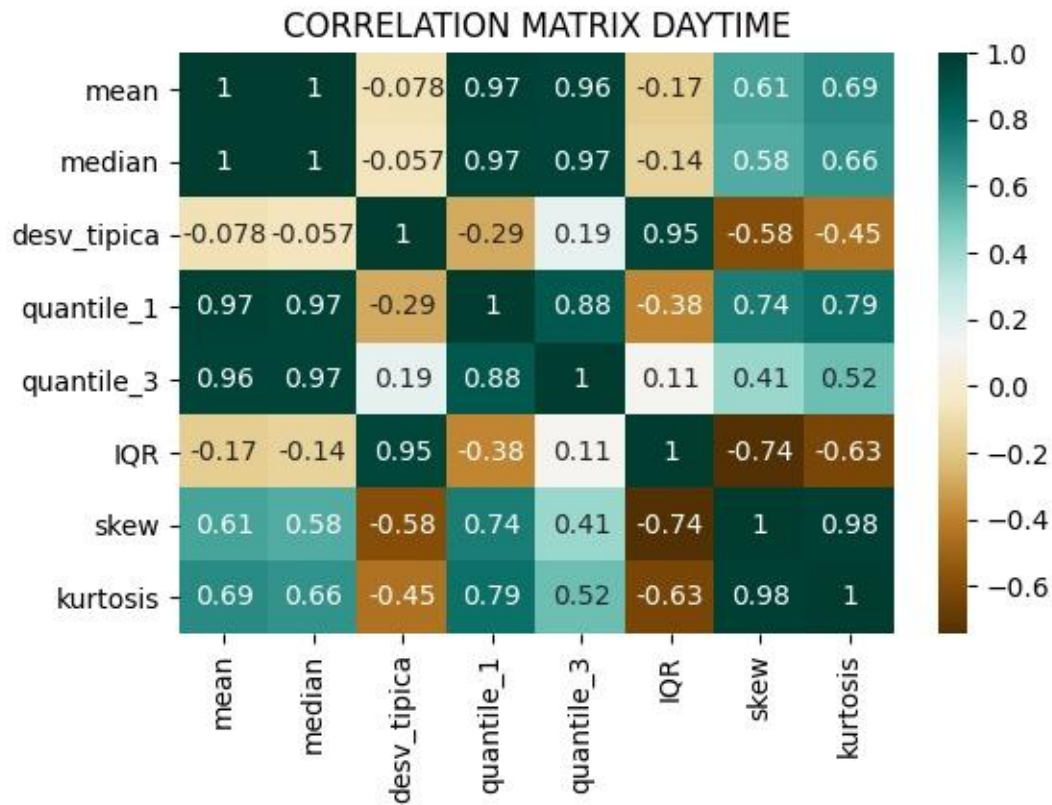
Pasemos ahora a analizar los indicadores de dispersión.



Vemos que ahora observamos una diferencia entre los estadísticos en la región ecuatorial. El hemisferio norte en ambos estadísticos posee una traza mucho más clara y lineal, lo que nos hace entender que en el hemisferio sur la latitud no es la variable más relevante sino que el estudio en cuanto a la latitud es confuso por la enorme variedad de climas de la región (hipótesis 2). Vemos como ambos gráficos nos muestran que la dispersión de datos en las zonas templadas es muchos mayor a la dispersión de las zonas cálidas. Esto es debido a la mayor afectación de las estaciones en dichas regiones. Los mínimos, en ambos gráficos, de las zonas cálidas los encontramos en el Ecuador, sin embargo, hay una diferencia de casi 12 grados Celsius entre ellos. Esta diferencia proviene de los valores extremos registrados en esta latitud, los cuales, son ignorados por el IQR y no por el STD.

Vamos a ver ahora como de importante son estas diferencias y si nos permiten corroborar o no nuestra hipótesis número 1.

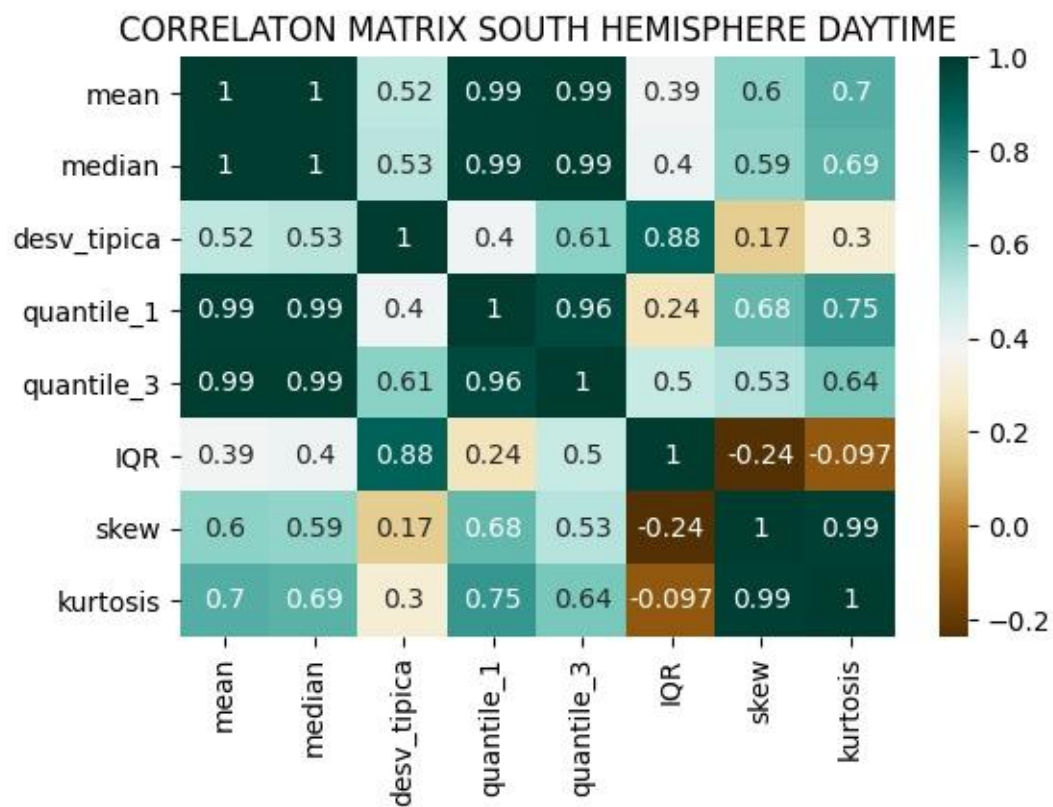
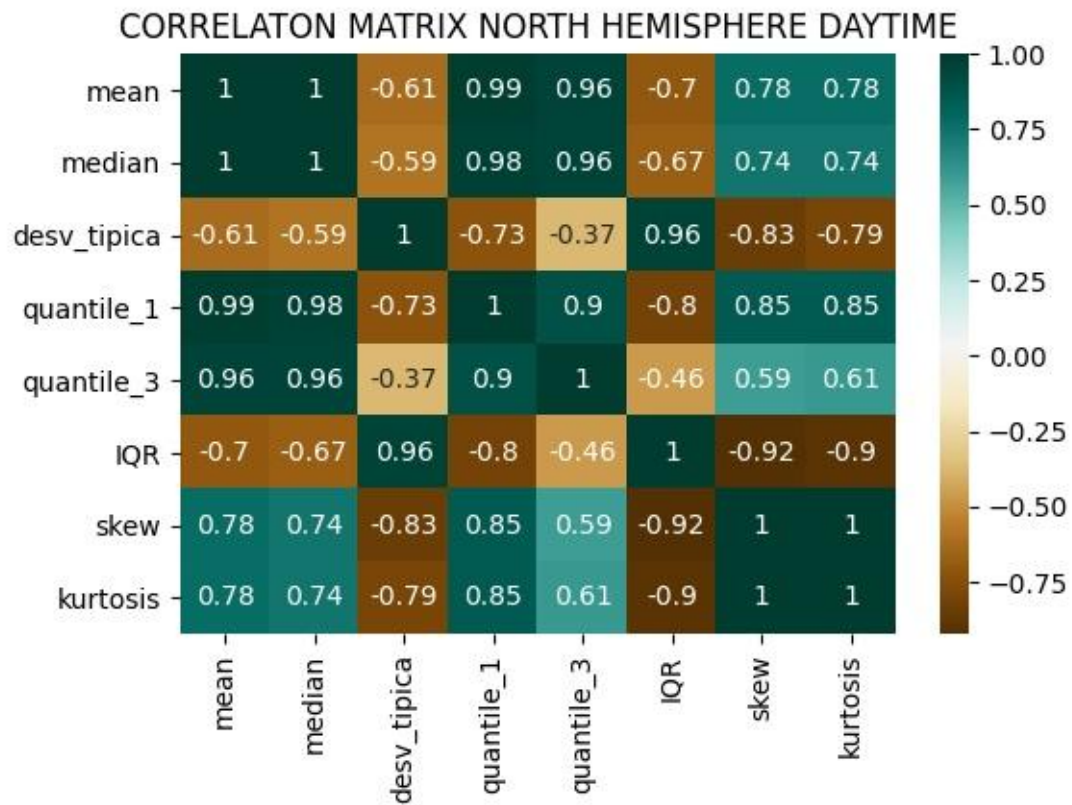
Esto lo haremos a través de las matrices de correlación entre los resultados obtenidos. Vamos a ello.



En esta primera matriz vemos que la correlación entre media y mediana es 1; y entre IQR y STD es 0.95. Esto nos indica que si nuestro estudio es a nivel global, no necesitaríamos diferenciar entre estadísticos robustos y no robustos.

Veamos qué ocurre si realizamos este estudio por hemisferios.

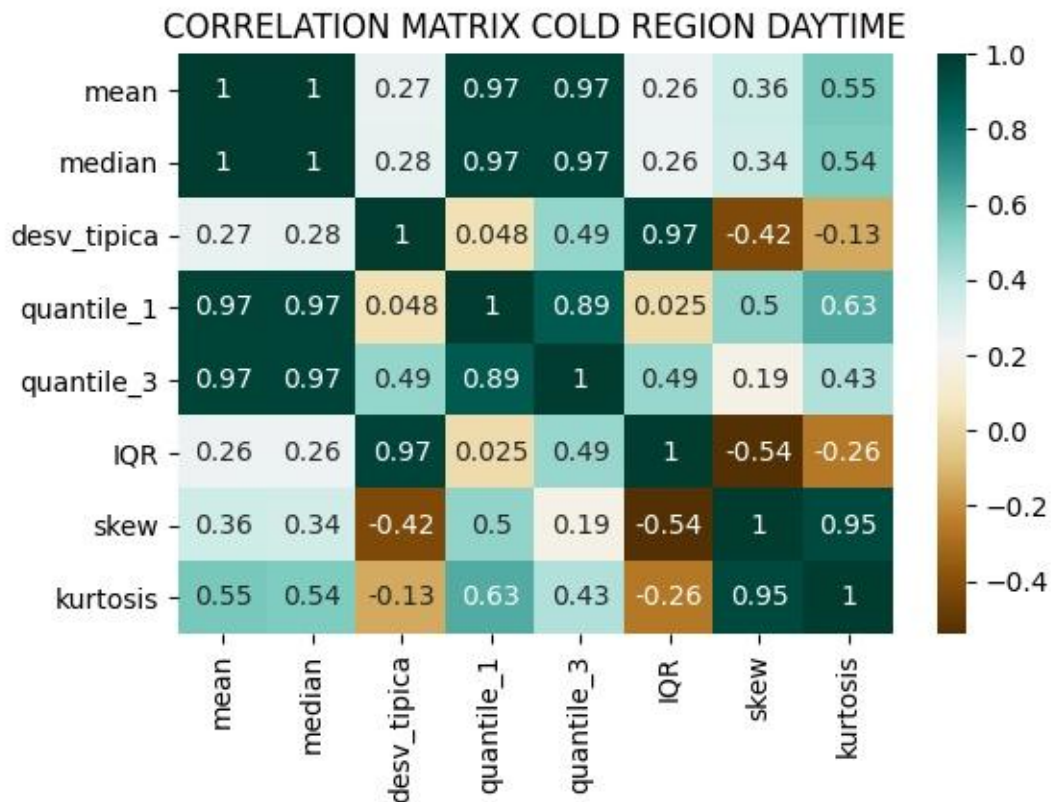
Para el hemisferio sur entendemos que su matriz de correlación será menos fiable debido a su menor claridad en las gráficas previamente mostradas.



Vemos que ambos hemisferios la correlación de datos entre media y mediana es total, para los indicadores de dispersión es

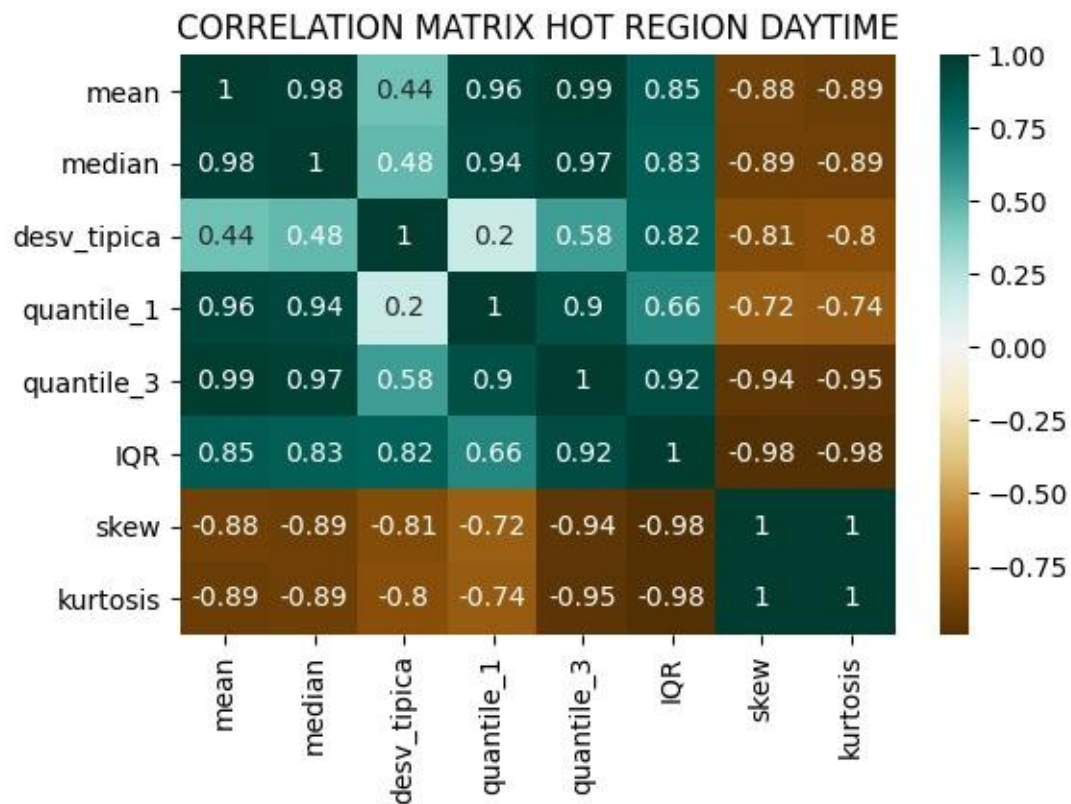
suficientemente elevada como para decir que son datos bien correlacionados.

Por último, vamos a ver estas correlaciones para las zonas cálidas y para las zonas templadas y frías el planeta.



Para las zonas templadas del planeta vemos que la correlación entre estadísticos de misma utilidad es excepcional, siendo 1 para los promedidores y 0.97 para los indicadores de dispersión. Esto nos hace concluir que para estas regiones los resultados van a ser invariantes si se usan estadísticos robustos o no.

Pasemos ahora a ver la zona cálida, la cual, es la relacionada con nuestra hipótesis número 1.

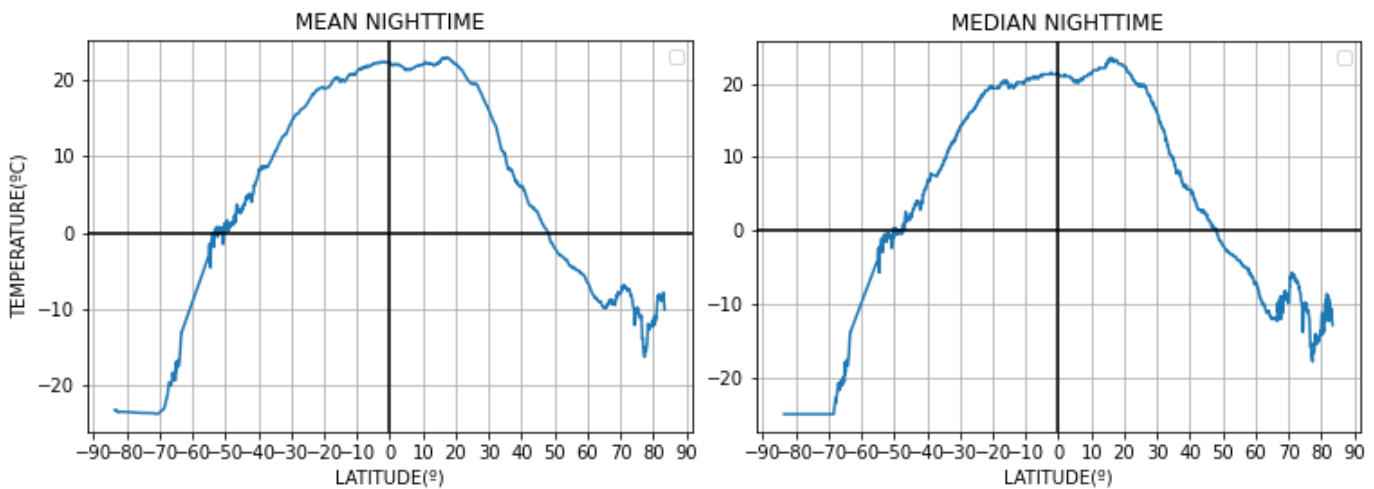


La matriz de correlación nos permite ver que la correlación entre media y mediana es prácticamente total (0.98) mientras que la correlación entre los indicadores de dispersión es de 0.82, mucho más baja que para las regiones templadas pero no lo suficiente como para decir que existe un sesgo enorme que nos imposibilite el estudio.

Así concluimos que, para el uso diurno, el estudio a nivel latitudinal de la temperatura se puede hacer tanto con estadísticos robustos como con estadísticos de menor robustez, únicamente teniendo precaución en el estudio de la dispersión de datos en las zonas cálidas del planeta.

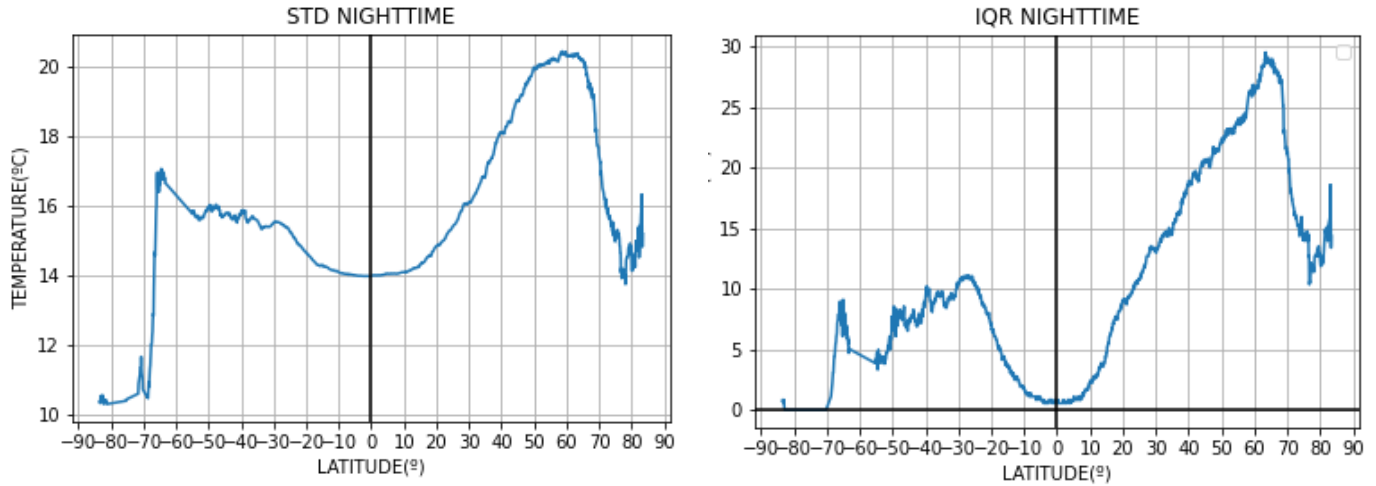
5.2. Resultados globales nocturnos

Una vez visto los resultados diurnos vamos a ver si los nocturnos son análogos o no. Este apartado se abordará más como una comparativa con dichos resultados que como explicación de los mismos, por ello la estructura va a ser la misma.



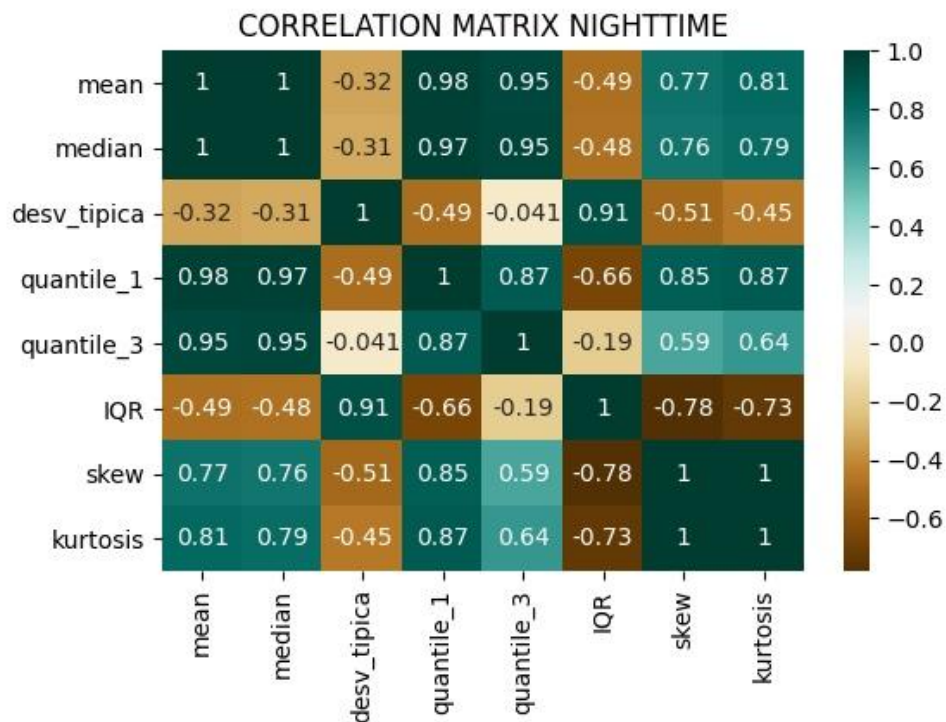
Vemos que, al igual que en el caso diurno, ambas gráficas cortan al eje x en 0 a la misma latitud, 50 grados sur y 50 grados norte. Son 10 grados más cerca del Ecuador respecto al caso diurno, lo que tiene sentido, dado que al ser de noche, más superficie estará por debajo de 0 grados Celsius. En ambas gráficas el máximo se encuentra en el trópico de Cáncer para el hemisferio norte mientras que para el hemisferio sur se encuentra en el Ecuador. Esto se debe a que la temperatura nocturna tiene relación con la cantidad de energía liberada por la Tierra y retenida por la atmósfera. Al haber mucha menor cantidad de superficie en el hemisferio sur, esta liberará menos energía.

Pasemos ahora a analizar los indicadores de dispersión.



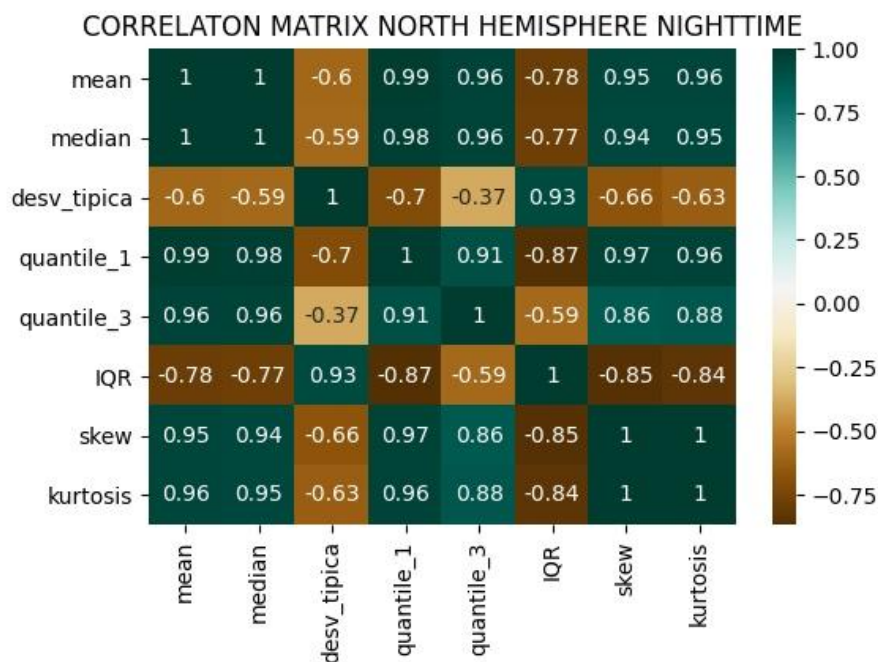
Los resultados son análogos a los obtenidos para el caso diurno. El Ecuador es lo más notable junto con la diferencia de máximos en el hemisferio norte, punto no comentado en el caso diurno. Es como si la desviación típica magnificase menos las dispersiones en las regiones, siendo su máximo 21 grados Celsius en comparación con el máximo del IQR, 29 grados Celsius; y habiendo una diferencia en la dispersión en el Ecuador de casi 14 grados Celsius.

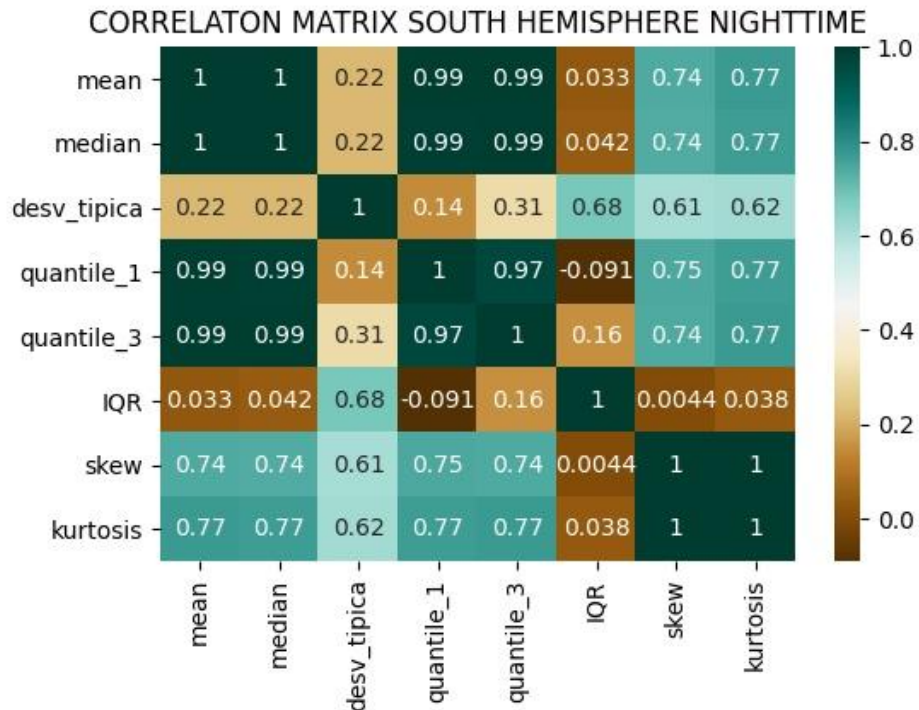
Analicemos las matrices de correlación para ver si estos resultados son significativos.



Al igual que para el uso diurno la correlación entre media, mediana e indicadores de dispersión es superior a 0.9. Por lo que si hacemos el estudio global no habrá problema al utilizar unos estadísticos u otros.

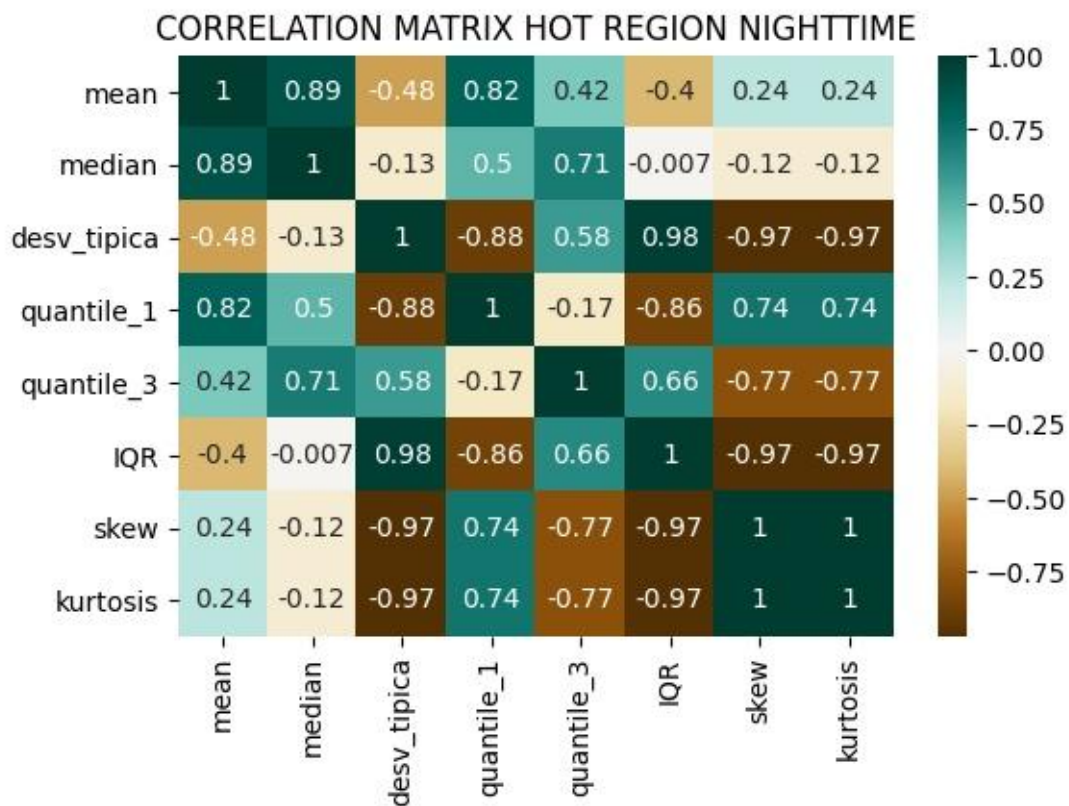
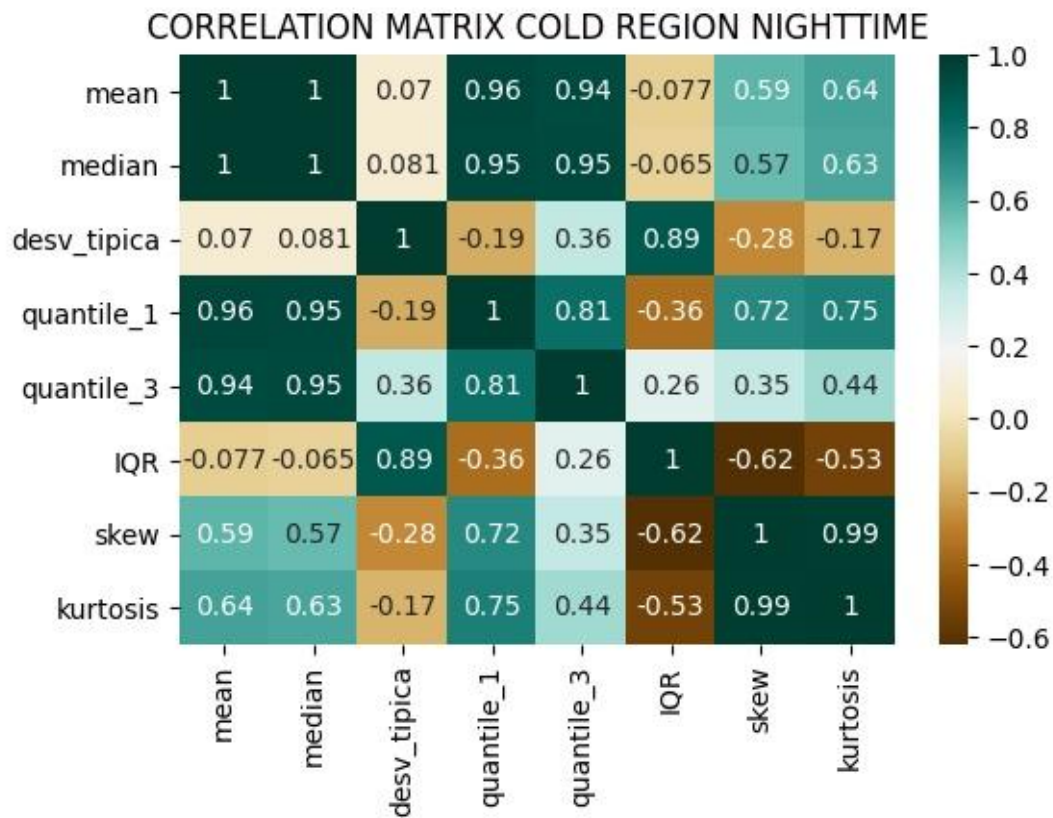
Pasemos ahora al estudio por hemisferios:





Tanto para el hemisferio norte como para el hemisferio sur las medias y medianas tienen una correlación perfecta, sin embargo, dicha correlación solo se mantiene para los indicadores de dispersión del hemisferio norte. En el hemisferio sur esa correlación cae hasta ser de 0.68, magnitud insuficiente para hacer el estudio indistintamente. Encontramos aquí un resultado que puede apoyar nuestra hipótesis, y es que, si ya ese hemisferio es complicado de estudiar por la baja superficie y la variedad climatológica, hay que tener en cuanto a estos resultados a la hora de interpretar la dispersión en este horario.

Veamos ahora las matrices de las zonas cálidas y templadas.



Vemos ahora resultados que no nos esperábamos en absoluto. La correlación de los indicadores de dispersión en las zonas templadas

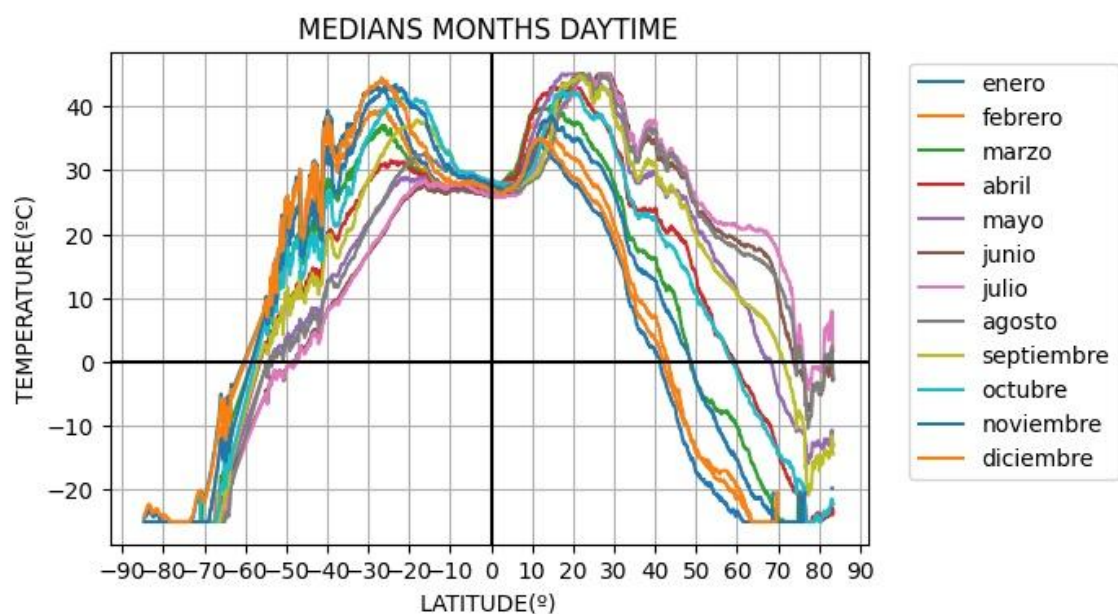
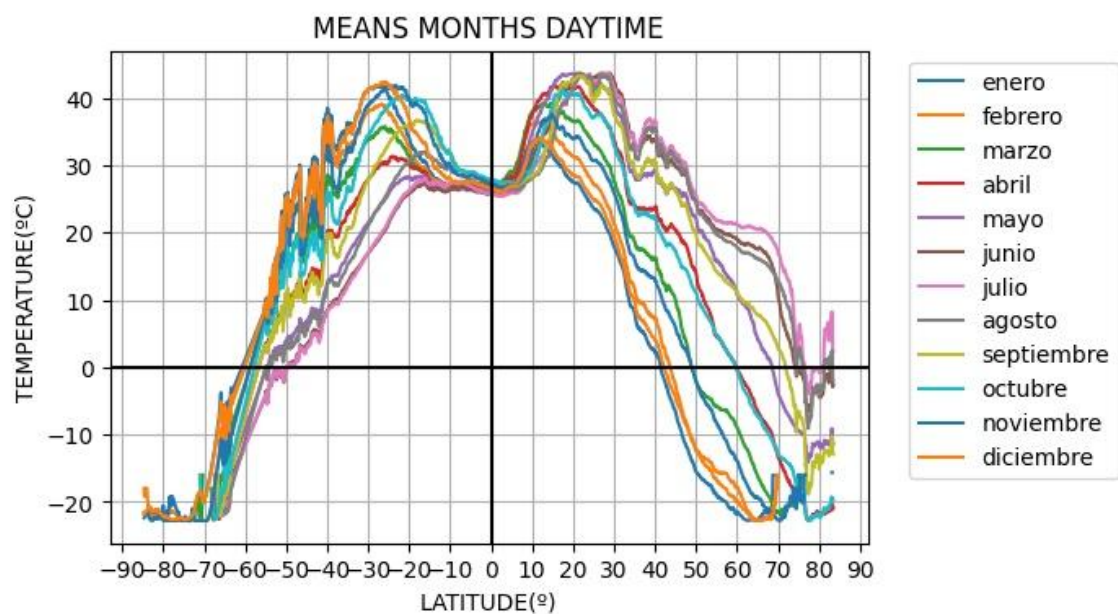
del planeta es inferior a la de las zonas cálidas. Un hecho que se opone a nuestra hipótesis y que nos hace ver que, para horario nocturno, el problema de usar un tipo de estadísticos u otros, está en relación con el hemisferio en el que nos encontremos y no con la zona (cálida, templada).

Así pues, para el horario nocturno, vemos que en el hemisferio norte no hay distinción entre usar estadísticos robustos y no robustos, sin embargo, sí la habrá en el hemisferio sur a la hora de estudiar la dispersión de datos de la distribución.

5.3. Resultados por meses

Estudiaremos en este caso como varía la temperatura no solo con la latitud sino también con la época del año en la que nos encontremos.

Para empezar vamos a presentar tanto las medias como las medianas en ambos horarios y ver la variación según los meses. Esto se hará con el fin de entender mejor la climatología por hemisferios de cara a reforzar la hipótesis número 2.



Empezamos con el horario diurno. Una vez entendido que podríamos trabajar tanto con medias como con medianas graficamos ambas para corroborar su correlación y se ve que así es, las gráficas son casi iguales.

Vamos a interpretar las gráficas, las cuales, nos van a proporcionar conclusiones muy interesantes más allá de las hipótesis planteadas.

Empezamos viendo que las temperaturas en el Ecuador a lo largo de todo el año son siempre las mismas, estando en torno a 27 grados Celsius. A medida que avanzamos hacia los trópicos vemos como la diferencia entre meses va aumentando, siendo julio y agosto los meses más cálidos en el trópico de Cáncer y enero, febrero, noviembre y diciembre los más cálidos en el trópico de Capricornio.

Analicemos ahora el hemisferio norte. A medida que avanzamos desde el trópico de Cáncer hasta latitudes altas como los 80 grados norte, la diferencia entre los meses va aumentando, teniendo su máxima diferencia entre los 50 y los 70 grados norte. A estas latitudes las estaciones son muy notables. El hemisferio norte se comporta de una forma bastante lineal debido a que en su gran superficie terrestre, el clima más abundante es el continental, lo que permite que la latitud sea la variable de mayor peso en cuanto a la variación de temperatura.

Tratemos ahora el problema de ártico. Desde los 79 hasta los 85 grados norte (límite de datos recogidos), vemos que hay 2 meses del año que superan los 0 grados Celsius, julio y agosto. Esto implica que, durante esos meses, se producirá un deshielo en esas latitudes. Esto implica que la única región del ártico que podrá tener hielo durante todo el año es la conformada entre 74 y 79 grados norte, lo

que implica una subida del nivel de los océanos y un cambio enorme en los ecosistemas de la región.

Estudiemos ahora el hemisferio sur. La mayor diferencia entre meses se da entre 23 y 27 grados sur. A partir de los 37 grados sur y hasta los 55 grados sur, la gráfica es muy borrosa, esto es debido a la hipótesis 2, la diversidad de climas hace que la latitud no sea la única variable a tener en cuenta. Esto lo ilustramos en la figura 2.

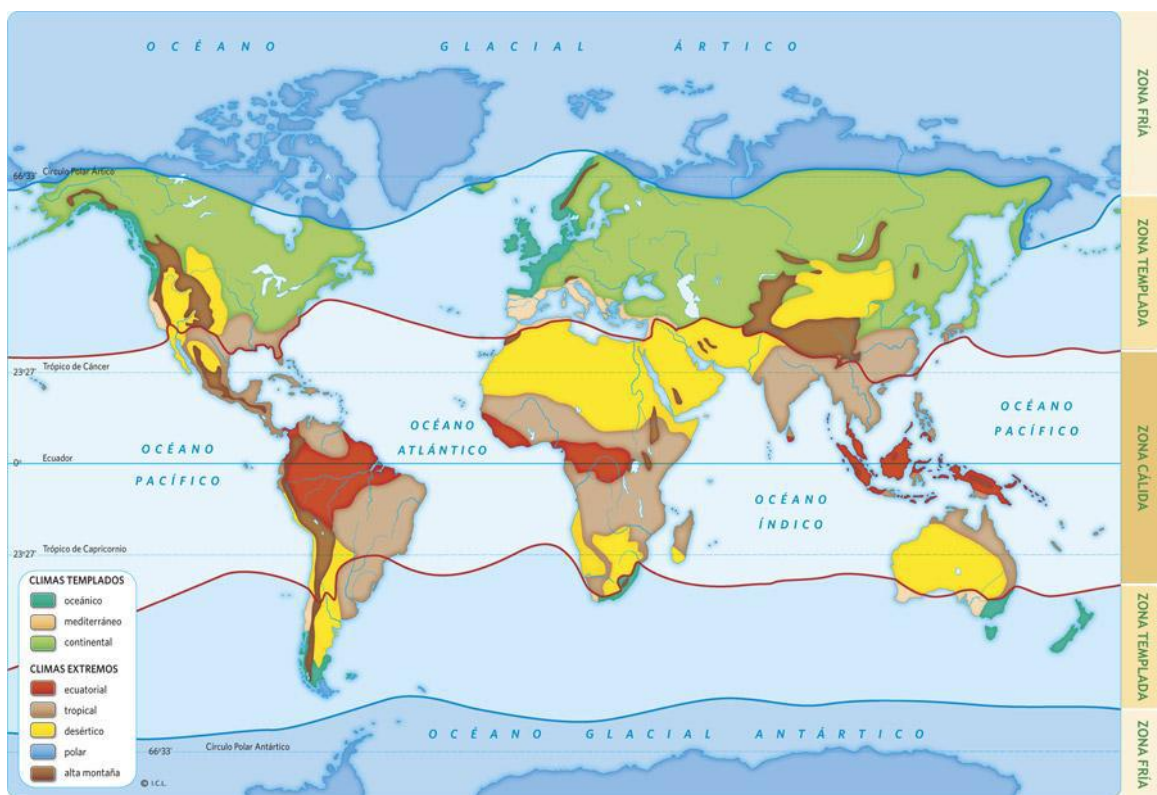
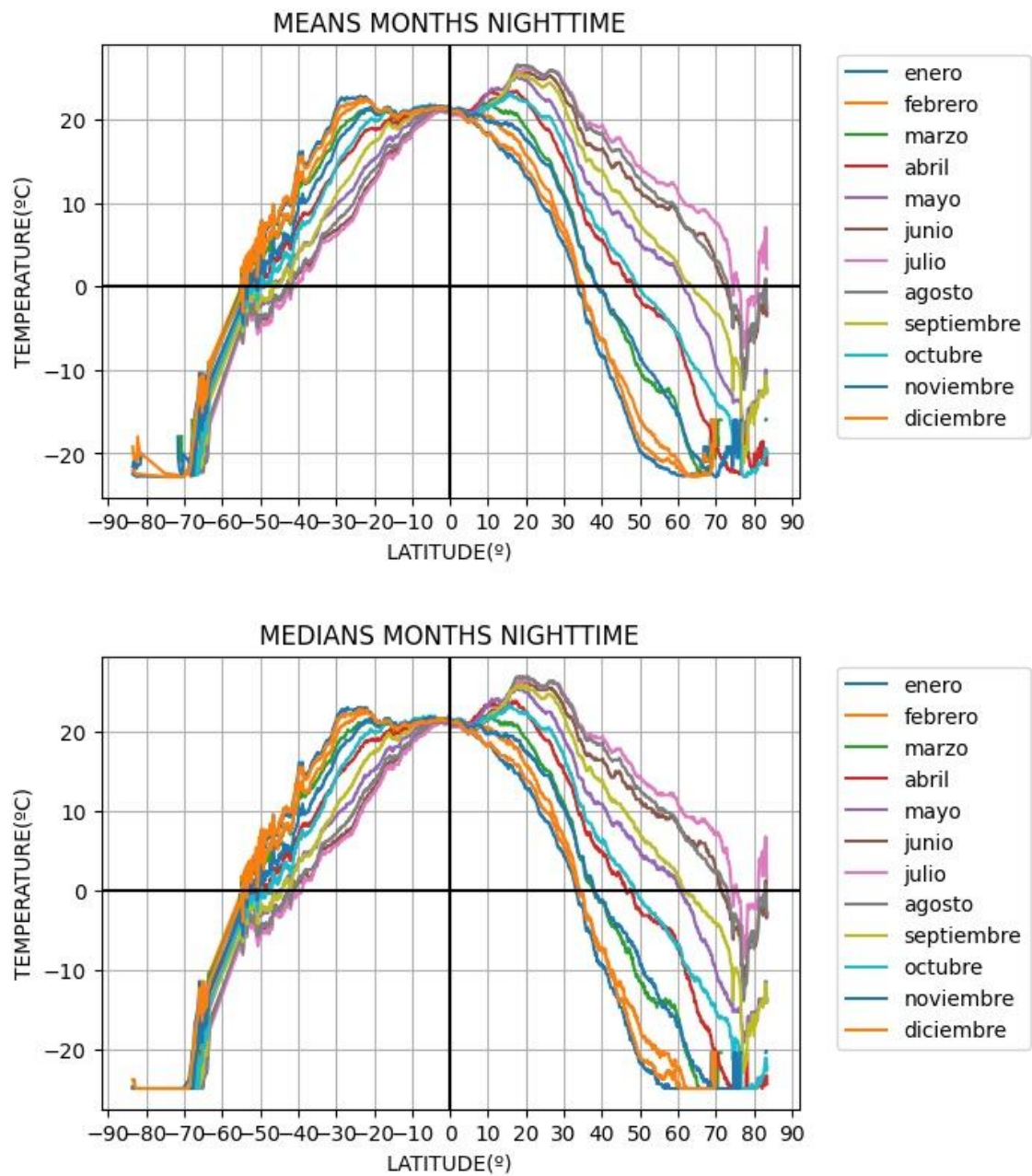


Figura 2. Climas hemisferio sur

Por debajo de los 60 grados sur las temperaturas estarían por debajo de 0 grados Celsius todo el año.

Pasemos ahora a analizar las gráficas obtenidas para el horario nocturno y ver si son análogas al diurno.

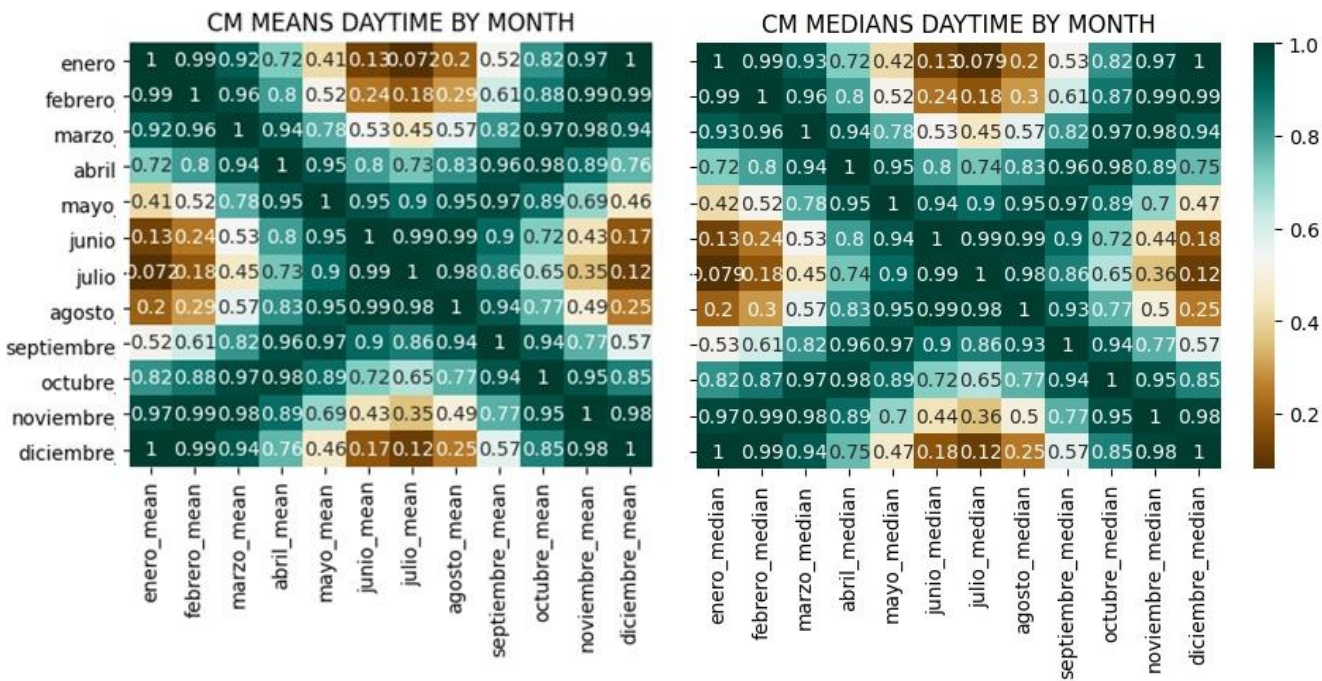


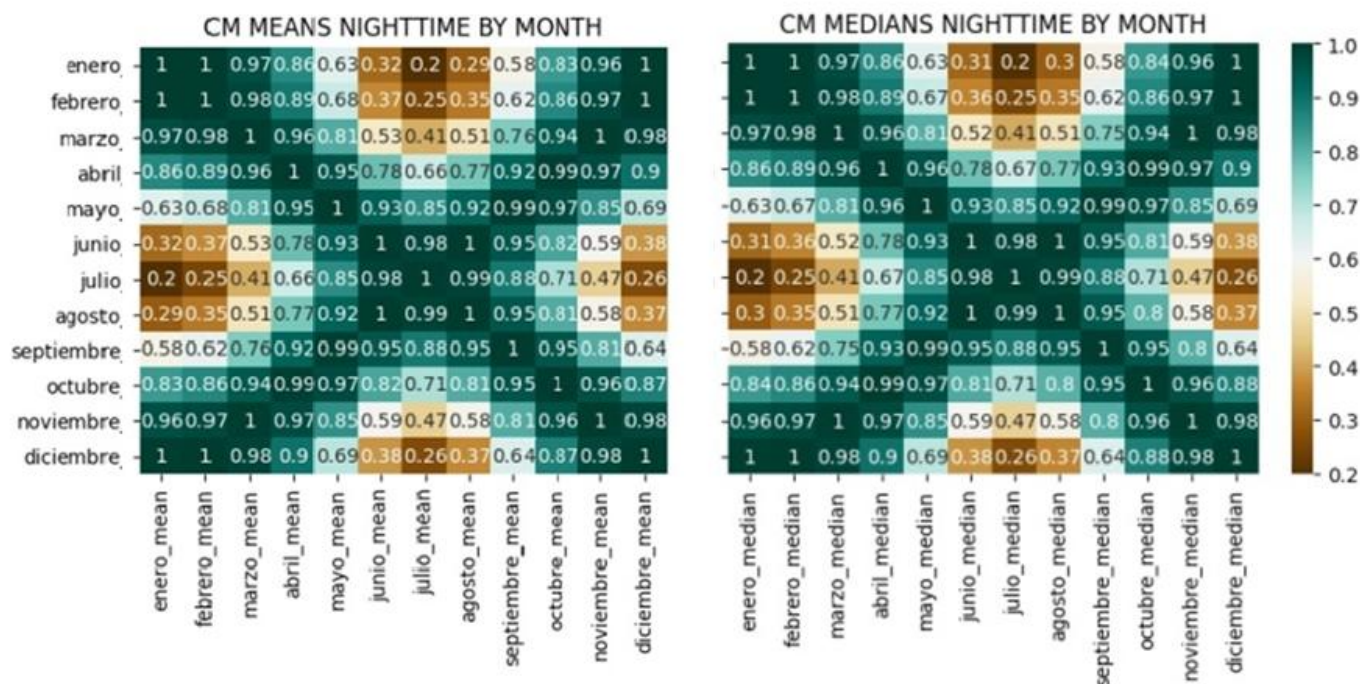
Vemos otra vez que las gráficas son prácticamente exactas. Las principales diferencias con el horario diurno se encuentran en el

trópico de Capricornio, donde ahora la temperatura vuelve a bajar mucho más que en el trópico de Cáncer.

Entre los 79 y los 85 grados norte, sigue habiendo meses donde la temperatura aumenta de los 0 grados Celsius.

Para finalizar vamos a adjuntar las correlaciones existentes entre las medias de los meses del año y también entre sus medianas.





Como se puede observar, la máxima correlación siempre se da en los meses contiguos, esto era de esperar dado que son los meses que tienen la incidencia de luz solar más similar. La menor correlación se da entre julio y enero siendo esta de 0.072 en horario diurno y 0.2 en horario nocturno.

6. Conclusiones

Las conclusiones respecto a las hipótesis son las siguientes.

Hipótesis 1: No es cierto que en la zona cálida haya que usar estadísticos de mayor robustez dado que la correlación entre media y mediana ha estado en todos los casos entre 0.98 - 1. Para el horario diurno, la mayor discrepancia la encontramos en los indicadores de dispersión para la zona cálida del planeta. Para el resto se puede usar indistintamente estadísticos de menor o mayor robustez. Para el horario nocturno la mayor discrepancia se encuentra en los indicadores de dispersión en el hemisferio sur, esta discrepancia es bastante relevante y hay que tener mucho cuidado a la hora de estudiar dicha región.

Hipótesis 2: Totalmente corroborada, tanto en el estudio global como en el estudio por meses realizado. La enorme variedad climatológica hace que la latitud sea más relevante en el estudio de temperaturas en el hemisferio norte que en el hemisferio sur.