



**Universidad de Cienfuegos
“Carlos Rafael Rodríguez”**

Tesis presentada en opción al Título Académico de
Máster en Matemática Aplicada
(Mención Estadística Matemática)

Título:

Técnicas estadísticas para la elaboración de un calendario térmico.

Autor:

Lic. Daynier Escalante Pérez

Tutores:

Dr. C. Yohanna de la C. Morales Díaz
MSc. María M. Martín García

Cienfuegos, 2017

PENSAMIENTO

*“La estadística demuestra que suele ser más fácil hacer algo bien
que explicar por qué se hizo mal.”*

Allen L. Webster

Dedico esta investigación a mis padres Olga Lidia y Domingo por el amor infinito, la preocupación constante de que me supere y la ayuda incondicional para lograr mis metas.

A Marilú, Mamita y Mayito, pues sin ellos no hubiera sido posible estudiar en esta maestría y poder deleitarme en la bella ciudad del mar.

A mi hermana, mi familia, mi ahijado y mis amigos, que siempre han estado al tanto de cada paso profesional y personal que doy.

A todos los profesores que me han guiado hasta aquí, esta tesis va especialmente dedicado a ellos.

AGRADECIMIENTOS

A mis padres por todo el apoyo que me han brindado para poder superarme cada día.

A mi familia de Cienfuegos, por acogerme. Sin ellos no hubiera sido posible este resultado. Gracias Marilú y Mayito por permitirme ser una parte más de la familia.

A María Carla, por ser una excelente (pero finalista) compañera de estudios hace 7 años.

A Alina por toda la ayuda durante el transcurso de la maestría, bueno, desde antes.

A las tutoras por acogerme y poder llegar a los resultados alcanzados.

A Alia y Nivaldo por sus consejos y orientaciones en la realización de la tesis.

A José Raúl por su ayuda metodológica y brindarme parte de su tiempo.

A los trabajadores del CIEC por permitirme comenzar la maestría y a Mary por continuarla.

A la profe Lourdes por toda la ayuda brindada en el transcurso de la Maestría.

RESUMEN

Este trabajo propone elaborar un calendario térmico a través de procesamientos estadísticos de manera que permita que se puedan trazar estrategias que minimicen efectos negativos en los cultivos tomando como caso de estudio el cultivo de la piña en Ciego de Ávila. Se seleccionan las medidas de tendencia central para la elaboración de un calendario térmico y se identifica la más apropiada. Se establecen comparaciones entre meses de comportamiento normal de la temperatura para evidenciar la efectividad de la técnica estadística seleccionada y se compara el calendario térmico propuesto con datos del Modelo de Circulación Regional PRECIS CARIBE para estimar la posible existencia de variaciones en la variable durante los próximos años. Los resultados proyectan que la mediana es la medida de tendencia central más factible para la elaboración de un calendario térmico por su efectividad en comparaciones con observaciones reales. De 2016 a 2020 el calendario propuesto tendrá un elevado por ciento de certidumbre, mostrando escasa variación en comparación con los datos que modela el PRECIS CARIBE, los que fueron trabajados en el paquete estadístico STATGRAPHICS.

ÍNDICE

INTRODUCCIÓN.....	1
CAPÍTULO 1: FUNDAMENTOS TEÓRICOS Y METODOLÓGICOS PARA LA ELABORACIÓN DE CALENDARIOS CLIMÁTICOS	6
1.1 CALENDARIOS CLIMÁTICOS. ESTUDIOS DE TEMPERATURA.....	6
1.2 ANTECEDENTES HISTÓRICOS DE LA ESTADÍSTICA. SUS APLICACIONES....	12
1.3 TÉCNICAS ESTADÍSTICAS. DEFINICIONES Y CONSIDERACIONES.....	16
1.3.1 SERIES TEMPORALES	16
1.3.2 MEDIDAS DE TENDENCIA CENTRAL.....	17
1.3.3 OTRAS MEDIDAS A CONSIDERAR	20
1.4 MODELO DE CIRCULACIÓN REGIONAL PRECIS CARIBE. VERIFICACIÓN DE MODELOS METEOROLÓGICOS	24
1.5 PRUEBA DE HIPÓTESIS. ANÁLISIS DE VARIANZA	28
1.5.1 COMPARACIÓN DE VARIAS MUESTRAS EN EL STATGRAPHICS	29
CONCLUSIONES DEL CAPÍTULO	32
CAPÍTULO 2: MATERIALES Y MÉTODOS PARA LA ELABORACIÓN DEL CALENDARIO TÉRMICO.....	33
2.1 CARACTERIZACIÓN DE LA ZONA ESCOGIDA.....	33
2.2 OBTENCIÓN DE LOS DATOS. CONTEXTUALIZACIÓN DE LA METODOLOGÍA PROPUESTA POR SOLER Y MARTÍN.....	38
2.3 EMPLEO DE OTRAS TÉCNICAS ESTADÍSTICAS. SELECCIÓN DE LA MÁS EFECTIVA	39
2.4 TRABAJO CON EL MODELO DE CIRCULACIÓN REGIONAL PRECIS CARIBE..	41
2.4.1 TRATAMIENTO DE LOS DATOS EN EL PERÍODO BASE	42
2.4.2 EXACTITUD DE LOS DATOS	45
2.4.3 AJUSTE DEL MODELO.....	46
2.4.4 TRATAMIENTO DE LOS DATOS EN EL PERÍODO 2016-2020.....	49
CONCLUSIONES DEL CAPÍTULO	50

CAPÍTULO 3: DISCUSIÓN DE LOS RESULTADOS	51
3.1 CALENDARIO TÉRMICO SEGÚN LA METODOLOGÍA POR SOLER & MARTÍN (2002).....	51
3.1.1 CALENDARIO DE TEMPERATURA MÍNIMA	51
3.1.2 CALENDARIO DE TEMPERATURA MEDIA.....	55
3.1.3 CALENDARIO DE TEMPERATURA MÁXIMA	59
3.2 CALENDARIO TÉRMICO UTILIZANDO OTRAS TÉCNICAS ESTADÍSTICAS	63
3.2.1 CALENDARIO DE TEMPERATURA MÍNIMA	63
3.2.2 CALENDARIO DE TEMPERATURA MEDIA.....	64
3.2.3 CALENDARIO DE TEMPERATURA MÁXIMA	65
3.3 COMPROBACIÓN DE LOS RESULTADOS CON PERÍODOS DE COMPORTAMIENTO NORMAL.....	66
3.3.1 TEMPERATURA MÍNIMA	67
3.3.2 TEMPERATURA MEDIA.....	67
3.3.3 TEMPERATURA MÁXIMA.....	68
3.4 VERIFICACIÓN DEL PRECIS-CARIBE	70
3.5 COMPARACIÓN DEL CALENDARIO PROPUESTO CON EL PERÍODO 2016-2020	71
3.5.1 TEMPERATURA MÍNIMA	71
3.5.2 TEMPERATURA MEDIA.....	75
3.5.3 TEMPERATURA MÁXIMA.....	80
CONCLUSIONES PARCIALES DEL CAPÍTULO	83
CONCLUSIONES.....	84
RECOMENDACIONES.....	85
BIBLIOGRAFÍA.....	86
ANEXOS	90

INTRODUCCIÓN

El estudio de las variables meteorológicas contribuye al mejor aprovechamiento de los recursos brindados por la naturaleza. La información meteorológica aplicada es utilizada en distintos sectores, como por ejemplo en la agricultura, dadas las evidencias del calentamiento del sistema climático. Trabajar con herramientas estadísticas que permitan prever los cambios futuros en las variables climáticas, ayudaría a los hombres de campo a aumentar la capacidad de respuesta ante la ocurrencia de anomalías, como es el caso de la elaboración de calendarios climáticos.

Los calendarios de variables climáticas resultan de una notable utilidad en la planificación temporal de actividades socioeconómicas muy diversas, en especial, las desarrolladas total o parcialmente al aire libre. La construcción de un verdadero calendario climático exige disponer de series climáticas, es decir, de registros meteorológicos suficientemente largos para que tengan significación climática. Los calendarios pueden presentarse en formato numérico o gráfico de la variable o fenómeno analizado para cada día del año, los que se deducen del análisis estadístico de la serie correspondiente (Soler & Martín, 2002).

Al desarrollar un calendario climático se debe recurrir a un método estadístico de mayor precisión en sus resultados con la expectativa de mejorar el conocimiento del entorno y que la predicción del tiempo sea más confiable. La estadística es una herramienta que permite orientar y obtener resultados en cualquier tipo de estudio. Según Barón (1998), esta se ocupa de recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso, formular predicciones.

En la estadística un tema importante es el estudio de las medidas de tendencia central, las que se presentan como conceptos estadísticos básicos y de gran aplicación práctica en la vida cotidiana y profesional. La estadística proporciona herramientas fundamentales para el estudio del clima, permitiendo evaluar y cuantificar de forma objetiva el cambio de patrones climatológicos, así como estimar variaciones probables en el futuro. Alcanzar predecir, o al menos obtener algún tipo de información sobre cómo se van a comportar ciertos fenómenos climáticos, es de gran importancia para tomar medidas que permitan mitigar sus consecuencias.

Las evidencias del calentamiento del sistema climático son una preocupación y objeto de estudio en muchas investigaciones, sobre todo las desarrolladas en el sector agrícola. La provincia de Ciego de Ávila posee cultivos con características especiales, como la papa, la piña, el plátano, entre otros, donde ya se han ido notando la influencia de las variaciones del clima en la producción de estos, evidenciadas en distintos estudios. Es por eso que es imperante desarrollar instrumentos para poder contrarrestar, mediante la acción oportuna, los efectos negativos del clima en las áreas agrícolas, como es el caso de calendarios climáticos, que pueden ser propicios para la toma de decisiones.

En el 2002, Soler y Martín desarrollaron una propuesta metodológica para la confección de calendarios climáticos con el fin de ampliar sus niveles de aplicación, haciendo referencia a los calendarios pluviométricos. En el artículo presentaron a la media móvil como el tratamiento estadístico de los datos para eliminar los “dientes de sierra”, aunque para realizar las suavizaciones no propusieron ni indicaron cuál fue el software que utilizaron para ello.

Pérez *et al.* (2008) elaboraron calendarios pluviométricos en el margen oriental del área agrícola argentina. Utilizando como herramienta de ayuda la metodología de Soler & Martín (2002). Sin embargo, no se refirieron a un proceso matemático para la confección del calendario ni a un procesador de datos para el tratamiento de las observaciones.

En Cuba se han desarrollado calendarios climáticos con la variable precipitación, como es el caso de Matos y Batista (2005), para la zona litoral de Cayo Coco, Ciego de Ávila; Martín y Matos (2009), para la zona de Júcaro, al sur de esta provincia; Martín (2009), que realizó su tesis de maestría en Calendarios Pluviométricos en la provincia Ciego de Ávila. En la bibliografía consultada no se ha encontrado estudios relacionados con calendarios térmicos.

Todos estos autores recurrieron a la metodología propuesta por Soler y Martín, utilizando la media móvil y sin referirse a un procesador estadístico para el tratamiento de los datos. Los procesadores estadísticos son capaces de proporcionar análisis exigentes y experimentados y algunos facilitan la simplicidad suficiente para realizar procedimientos complejos. El uso de estos programas sería de gran importancia cuando se desea realizar análisis de variables meteorológicas.

Si se elabora un calendario térmico a partir de la metodología propuesta por Soler y Martín, para la confección de calendarios climáticos, sería utilizando la media móvil como el procedimiento estadístico. Sin embargo, las medidas de centralización o de posición central

se pueden utilizar con esta finalidad, siendo técnicas simples y con fácil nivel de interpretación. A partir del análisis estadístico estas pueden ser una herramienta de mejor efectividad en los pronósticos y que permitan una certidumbre superior en los servicios.

Por todo lo anteriormente descrito es que se propone el siguiente **Problema científico**:

¿Cómo elaborar un calendario térmico que permita mejorar la eficiencia en el pronóstico de la variable temperatura?

El **Objeto de estudio** se centra en el proceso de elaboración de calendarios climáticos. El **Campo de acción** está en función de la eficiencia en el pronóstico de la temperatura.

Para la solución del problema se plantea la siguiente **Hipótesis**:

Si se elabora un calendario térmico utilizando técnicas estadísticas apropiadas entonces se incrementa la eficiencia del pronóstico en la variable temperatura y de esta manera se proporciona una herramienta para la toma de decisiones que contribuya a minimizar los efectos negativos en los cultivos.

Se asume como

Variable independiente: calendario térmico utilizando técnicas estadísticas.

Variable dependiente: la predicción más confiable de la variable temperatura.

Para demostrar la validez de la hipótesis se propone como **Objetivo General**:

Elaborar un calendario térmico utilizando técnicas estadísticas que permitan mejorar la eficiencia en el pronóstico de la variable temperatura y que de esta manera se proporcione una herramienta para la toma de decisiones que contribuya a minimizar efectos negativos en los cultivos.

Para su obtención se tiene como **Tareas de investigación**:

- Revisión bibliográfica de los antecedentes teóricos y metodológicos para la elaboración de calendarios climáticos en la aplicación de técnicas estadísticas.
- Selección de las metodologías y técnicas estadísticas más apropiadas para la elaboración de un calendario térmico.
- Comparación entre meses de comportamiento normal de la variable temperatura para evidenciar la efectividad de la técnica estadística seleccionada.

- Comparación entre el calendario térmico propuesto y los datos del Modelo de Circulación Regional PRECIS CARIBE para estimar la posible existencia de variaciones en la variable temperatura durante los próximos años.

La **novedad** de esa investigación radica en la contextualización de la metodología para la confección de calendarios climáticos, propuesta por Soler & Martín (2002), para elaborar un calendario térmico aplicando técnicas estadísticas, lo que se considera una herramienta para la toma de decisiones que contribuya a minimizar efectos negativos en los cultivos.

Para el desarrollo del trabajo, se utilizaron los siguientes **métodos y técnicas**:

Del nivel teórico

- El método histórico-lógico: para conocer la evolución y desarrollo de los estudios asociados a la aplicación de las técnicas estadísticas a variables meteorológicas, determinar regularidades en función de la selección de las más apropiadas para elaborar un calendario climático.
- El método analítico-sintético: en la determinación de los fundamentos teóricos y metodológicos la investigación, la elaboración del calendario térmico a partir de técnicas estadísticas y la determinación de conclusiones finales.
- El método inductivo-deductivo: en la generalización de propuestas relacionadas con la elaboración de calendarios climáticos, determinar potencialidades, limitaciones y la determinación de las características distintivas de las principales categorías conceptuales a emplear para la elaboración del calendario térmico.

Del nivel empírico:

- Entrevista: a directivos y trabajadores de la Brigada Estatal de Piña de la Empresa Agroindustrial Ceballos, lo que permitió obtener información para la caracterización de la zona agrícola de estudio.
- Análisis documental: en el análisis de calendarios climáticos, así como en la base de datos de las variables meteorológicas de la estación 78346 del municipio de Venezuela en función de la caracterización, la comparación de datos y elaboración del calendario térmico.

Técnicas y procedimientos de la estadística:

- Estadística descriptiva: para la confección de tablas, gráficos, en el uso de las medidas de tendencia central para la comparación e interpretación de datos, asociados a la variable meteorológica analizada en cada una de las técnicas estadísticas.
- Estadística inferencial: en el contraste de hipótesis al decidir si dos muestras son estadísticamente diferentes, en el análisis de las suavizaciones de las series temporales y en la estimación de variaciones que puedan existir entre el calendario térmico y futuras observaciones en la variable temperatura.

El trabajo está estructurado en introducción, tres capítulos, conclusiones, recomendaciones, bibliografía y anexos.

En el Capítulo 1 se realiza una revisión bibliográfica del tema; se asumen presupuestos teóricos asociados al objeto y campo de investigación aportados por diferentes autores y se toman consideraciones sobre métodos estadísticos vinculados al estudio de las variables meteorológicas. También se presenta información sobre la verificación de modelos meteorológicos y aspectos para el conocimiento del Modelo de Circulación Regional PRECIS CARIBE.

En el capítulo 2 se realiza una caracterización del área seleccionada para el estudio. Se procede para la elaboración del calendario térmico a partir de la metodología propuesta por Soler & Martín (2002) para la confección de calendarios climáticos, así como el empleo de otras técnicas estadísticas para la confección del mismo. Se detallan los procedimientos realizados con el ajuste y manejo de las salidas del Modelo de Circulación Regional PRECIS CARIBE, así como los métodos para la verificación del modelo.

En el Capítulo 3 se presentan los resultados relacionados con la aplicación contextualizada de la metodología propuesta por Soler & Martín (2002) para la confección del calendario térmico. Se establecen comparaciones entre meses de comportamiento normal de la variable temperatura para evidenciar la efectividad del calendario propuesto. Además, se comprueba la exactitud del Modelo de Circulación Regional PRECIS CARIBE y se compara el calendario con las salidas del modelo para analizar variaciones de la variable en los próximos 5 años.

CAPÍTULO 1: FUNDAMENTOS TEÓRICOS Y METODOLÓGICOS PARA LA ELABORACIÓN DE CALENDARIOS CLIMÁTICOS

En este capítulo se realiza una revisión bibliográfica sobre los fundamentos teóricos y metodológicos para la elaboración de calendarios climáticos, se toman consideraciones sobre métodos estadísticos vinculados al estudio de las variables meteorológicas y se explica sobre la verificación de modelos meteorológicos y aspectos para el conocimiento del Modelo de Circulación Regional PRECIS CARIBE.

1.1 CALENDARIOS CLIMÁTICOS. ESTUDIOS DE TEMPERATURA

El Calendario (del latín *calenda*) es una cuenta sistematizada del tiempo para la organización de las actividades humanas (Real Academia Española, 2014). Antiguamente estaba instituido por los ciclos lunares. En la actualidad, los diversos calendarios están basados en el ciclo que describe la Tierra alrededor del Sol y se denominan calendarios solares. El calendario sideral se basa en el movimiento de otros astros diferentes al Sol (Prensa Latina, 2013).

Martín (2003) define el Calendario Climático a través del comportamiento de una variable meteorológica determinada, que bien puede ser la precipitación, la temperatura, la presión atmosférica, etc., en cada una de las jornadas del año, recibiendo el nombre según la variable tratada en cuestión. Es por eso que se puede indicar que el calendario térmico, como bien dice su nombre, estudia y ordena a resolución diaria la distribución estacional de la temperatura. Sus resultados (valores de temperatura a régimen diario), desde el punto de vista de la predicción y la climatología brindan información detallada de la variable en cuestión.

Los calendarios de variables climáticas resultan de una notable utilidad en la planificación temporal de actividades socioeconómicas muy diversas, en especial, las desarrolladas total o parcialmente al aire libre. La construcción de un verdadero calendario climático exige disponer de series climáticas, es decir, de registros meteorológicos suficientemente largos para que tengan significación climática, para cada día del año. En consecuencia, hay que poseer en realidad 365 ó 366 series climáticas, una relativa a cada fecha del año.

Los calendarios climáticos pueden presentarse en formato numérico o gráfico, como valores medios, frecuenciales, extremos, probabilidades de ocurrencia, índices, entre otros, de la variable o fenómeno analizado para cada día del año, deducidos del análisis estadístico de la serie correspondiente. De esta manera se sabrá cuándo, a lo largo del año, es más

probable la presencia de cierto fenómeno, como por ejemplo la lluvia, o qué valor medio cabe esperar en una fecha determinada, como la temperatura media; así podrán elegirse las fechas más adecuadas para la realización de ciertas actividades al aire libre (Soler & Martín, 2002).

A nivel mundial, Javier Martín Vide es un pionero en aplicar los calendarios climáticos para conocer el comportamiento histórico de una variable meteorológica a régimen diario. Uno de sus primeros trabajos fue “Metodología para la construcción de un calendario de nubosidad. El caso de Barcelona”. Martín & Esteban (1998) proponen que para eliminar las irregularidades sin significación climática que producían las observaciones, se hallan medias móviles centradas de cada fecha del año en diferentes intervalos, de 5 en 5, de 11 en 11, de 15 en 15, etc. En el estudio, los autores exponen que la media de 15 en 15 fue la que proporcionó el suavizado más conveniente y lo suficientemente fino para eliminar esas irregularidades. Además, definieron con la confección del calendario, fechas con normal, mucha y poca nubosidad, así como la correlación de la nubosidad con la precipitación, aunque no precisaron un procesador matemático para el trabajo de los datos.

Soler & Martín (2002) desarrollaron una propuesta metodológica como tal, con el fin de ampliar sus niveles de aplicación, trabajo que fue reforzado por Martín (2003) en la obra científica “El tiempo y el clima”. Los autores hicieron su propuesta metodológica para la construcción de calendarios climáticos haciendo referencia a los pluviométricos, que pueden tener en cuenta tanto la frecuencia como la cantidad de precipitación media para cada día del año. Construyeron calendarios pluviométricos gráficos de 8 observatorios de la fachada mediterránea de la Península Ibérica, que revelaron una gran diversidad de comportamientos pluviométricos temporales medios. En el artículo presentaron a la media móvil como el tratamiento estadístico de los datos para eliminar los dientes de sierra, que son pequeñas irregularidades frutos del azar, aunque para realizar las diferentes suavizaciones no propusieron ni indicaron cuál fue el software que utilizaron para ello.

Pérez *et al.* (2008) elaboraron calendarios pluviométricos en el margen oriental del área agrícola argentina. La metodología para la elaboración de los calendarios fue desarrollada teniendo en cuenta las particulares condiciones de régimen hídrico (precipitaciones superiores a los 1000 mm principalmente en los meses de verano y otoño), suelos (pesados con baja capacidad de infiltración) y relieve (ondulado) de la zona. Identificaron fechas con baja probabilidad de que ocurran precipitaciones y con alta probabilidad de precipitaciones intensas.

Los autores no se refirieron a un proceso matemático para la confección del mismo ni a un procesador de datos para el tratamiento de las observaciones, solo que las series utilizadas tenían representación climática, concepto ofrecido por la Organización Meteorológica Mundial cuando las observaciones exceden los 30 años establecidos y que estadísticamente son adecuadas para los fines del análisis propuesto. En este trabajo, se pudo valorar a partir de las figuras representadas y comentarios, que utilizaron el procedimiento de medias móviles de Soler y Martín, pero no lo definieron como el método principal en la elaboración de los calendarios, por lo que se supone que fue una herramienta de ayuda, pues en la metodología sí dejan claros otros aspectos para su confección.

En Cuba, en lo referente a Calendario climático, se han desarrollado investigaciones con la variable precipitación, como es el caso de Matos & Batista (2005), Martín & Matos (2009) y Martín (2009). Matos & Batista (2005) con su estudio no hacen más que reproducir la metodología de Soler y Martín y aplicarla para la zona norte de Ciego de Ávila, utilizando los datos de la estación meteorológica 78339 de Cayo Coco, con pocos datos para una significación climática, pues solo contaron con 12 años de registro de precipitación.

Martín & Matos (2009) realizan la construcción de un calendario de tipo pluviométrico, a partir de la información referida a 22 años (1984 – 2005) de la estación meteorológica 78345 de Júcaro, al sur del municipio de Venezuela, en Ciego de Ávila, considerando no solo la cantidad media de precipitación para cada día del año, sino también el número de días en que están presentes los procesos de lluvia, con el propósito de conocer su distribución a resolución diaria. Con el estudio se tiene una determinación más precisa de los períodos poco lluvioso, normal y lluvioso. No se hace nuevas contribuciones desde el punto de vista de una nueva metodología para realizar un calendario climático, ni tampoco describen un procedimiento matemático, solo referencian al método de la media móvil propuesto por Soler y Martín.

Martín (2009) realizó un estudio a régimen diario con el objetivo de definir la climatología de la precipitación en la provincia Ciego de Ávila. Para ello elaboró los calendarios pluviométricos de las estaciones meteorológicas de la provincia avileña con una data de 22 años (1984 – 2005), propiciando la determinación de las características espacio - temporales de la precipitación diaria en ese territorio. Además, balanceó los resultados entre estas y comparó con la climatología de esta variable en la provincia, aplicando otras escalas temporales. Los resultados no refirieron una propuesta para la confección de calendarios pluviométricos, pues no era el objetivo, sino más bien una herramienta para

lograr caracterizar la precipitación en esta provincia. El autor utilizó la metodología de medias móviles propuesta por Soler y Martín.

En la bibliografía consultada no se ha encontrado estudios relacionados con calendarios térmicos.

Metodología para la elaboración de un calendario climático

Soler & Martín (2002) desarrollaron una propuesta metodológica para la confección de calendarios climáticos, la cual ha sido utilizada en varias investigaciones sobre el tema. En la bibliografía consultada esta es la única metodología en la que se detalla un procedimiento estadístico para la conformación del mismo.

Disponiendo de 30 años de registros diarios se calcula para cada fecha del año el promedio de los 30 valores observados en ella, aunque para alguna variable podría usarse un período inferior en una primera aproximación. En el caso de que se quisiera expresar otro parámetro estadístico como la variabilidad o los valores extremos, se operaría también para cada día, del modo correspondiente. Se dispondrá de esta manera de 365 (ó 366) valores, uno por día. El caso singular del 29 de febrero puede resolverse simplemente eliminándolo o asignándole el promedio que se obtenga (hallado obviamente a partir de menos datos que en los días restantes) o en el caso de variables frecuenciales, como el número de días lluviosos, con la frecuencia relativa al número de días observados.

Al representar los 365 valores hallados se aprecian siempre numerosos dientes de sierra que se traducen en pequeñas irregularidades fruto del azar sin significación climática. Para eliminar estas irregularidades se hallan medias móviles centradas de cada día con diferentes intervalos, de 5 en 5, de 7 en 7, de 9 en 9, de 11 en 11, etc. Las medias móviles de las primeras y de las últimas fechas del calendario se calculan suponiendo que éste es cíclico, de modo que el valor que sigue al del 31 de diciembre es el del 1 de enero, etc.

En dependencia de las diferentes suavizaciones que producen, la media móvil más adecuada es la que realiza el suavizado más conveniente, suficientemente fino como para eliminar las irregularidades no significativas, menores, pero sin ocultar las que puedan reflejar comportamientos climáticos significativos, como el de determinar los períodos especiales, lluviosos y secos, en el caso de la precipitación, cálidos y fríos, en el de la temperatura. Para ello, un procedimiento consiste en hallar ciertos valores de probabilidad, como los correspondientes a los cuartiles primero y tercero del conjunto de 365 (ó 366) valores medios o, mejor, de los 365 (ó 366) valores de una determinada media móvil.

El primer cuartil es aquel valor que iguala o supera a la cuarta parte de los valores inferiores de una serie y es inferior a las tres cuartas partes restantes. Mientras que el tercer cuartil iguala o supera al 75 % de los valores, siendo inferior al cuarto restante. Su cálculo manual es muy simple, aunque si la serie es larga es algo laborioso. Para ello se ordenan todos los valores de la serie, mejor de la media móvil elegida, en orden creciente o decreciente y se busca aquel que iguala o supera los porcentajes reseñados.

Los valores de los cuartiles establecen los umbrales que definen un día normal, en la variable analizada, cuando quede entre el primero y el tercer cuartil, mientras que los que queden por debajo del primero y por encima del tercero son los días con un carácter alejado de lo normal, seco y lluvioso, frío y cálido, etc., respectivamente. Este procedimiento determinará en cualquier lugar, sea cual sea su clima, un 50 % de días normales, un 25 % con un carácter por “debajo” de lo normal y otro 25 % por “encima”. Se trata, pues, de una clasificación relativa a cada lugar.

Podrían definirse más categorías, por ejemplo, con el empleo de los quintiles, cinco en este caso marcadas por los porcentajes del 20, 40, 60 y 80 %. Como la ocurrencia de una fecha cálida, por ejemplo, entre fechas normales o frías no tiene ninguna significación climática, dada la imprecisión astronómica que concurre en cada fecha, pudiendo achacarse al azar, la definición de período lluvioso o seco, cálido o frío, etc., ha de exigir la existencia de varias fechas de ese mismo carácter consecutivas, al menos más de dos. Del mismo modo un conjunto de varias fechas consecutivas con un carácter definido no debe romperse por la aparición de una o dos sin ese carácter en medio. De esta manera, puede definirse un período cálido como el constituido por más de dos fechas cálidas consecutivas o, a lo sumo, con una o dos sin ese carácter en el conjunto de las fechas cálidas sucesivas.

Estudios de temperatura

Las variaciones en los regímenes de temperatura están presentes en Cuba dada la ubicación geográfica del país, situado en latitudes tropicales del Hemisferio Norte, que implica que la predicción climática de la temperatura tenga presente la marcada persistencia de los débiles gradientes horizontales de esta variable y de los patrones atmosféricos. Lecha *et al.* (1994) describen el marcado ciclo estacional de la radiación y la circulación atmosférica sobre nuestro país, y con ello, las variaciones mensuales de la temperatura, cuya variabilidad está asociada en gran medida a las condiciones locales del relieve y la distancia a la costa.

El pronóstico climático de temperaturas a corto plazo resulta de vital importancia para las actividades económicas, sociales y medioambientales. Los pronósticos de esta variable en la actualidad se realizan mediante modelos climáticos, algunos dinámicos, estadísticos y sus combinaciones. En Cuba se brindan hasta el momento, las predicciones mensuales de temperatura mediante la implementación de regresiones; específicamente se utiliza el software *Climate Predictability Tool* (CPT, por sus siglas en inglés) desarrollado por el *International Research Institute for Climate and Society* (IRI, por sus siglas en inglés). Este software es capaz de implementar diferentes técnicas estadísticas, entre ellas la regresión por Componentes Principales, muy utilizada a nivel internacional por la habilidad del método de reducir en Funciones Ortogonales Empíricas la mayor varianza en la muestra independiente. Estos pronósticos nacionales se informan por el Centro Nacional del Clima del Instituto de Meteorología para tres regiones de Cuba, Occidente, Centro y Oriente (Sorí *et al.*, 2014).

En estudio de pronósticos de temperatura, las regresiones y las técnicas de reducción de escala son muy utilizadas por los meteorólogos, como por ejemplo Gu & Jiang (2005), aplicaron un modelo complejo autoregresivo para el pronóstico de la anomalía de temperatura mensual y Gutiérrez *et al.* (2004) plantean la ventaja del uso de la regresión teniendo en cuenta la correlación canónica aunque detallan que la habilidad del método depende de la capacidad de los predictores para explicar de forma lineal los predictandos. Estos métodos posibilitan realizar pronósticos más locales; pues en algunos casos se utilizan salidas de modelos climáticos que, en el caso de Cuba, no ofrecen una detallada información pues tienen baja resolución y homogenizan el territorio cubano, entre ellos se pueden mencionar el ECHAM4, GFDL y CCM3v6.

Hernández *et al.* (2009), realizaron un estudio sobre la climatología de la temperatura en Ciego de Ávila, analizando la temperatura máxima, media y mínima en las zonas norte y centro-sur de la provincia, para el período 1970 – 2008. Con el estudio se determinó que las temperaturas tienen similar fluctuación en el tiempo, pero con diferencias que señalan un ambiente más cálido en el centro-sur del territorio. Las diferencias del comportamiento térmico que advierten, argumentan la necesidad e importancia de pronosticar para cada estación los valores de temperatura y lograr una mayor exactitud de acuerdo a la heterogeneidad en el comportamiento de la temperatura en la región.

Sorí *et al.* (2014) efectuaron un pronóstico climático mensual de la temperatura en Ciego de Ávila, elaborando un método estadístico para el pronóstico de cada temperatura (media,

máxima media y mínima media mensual). Se obtuvieron ecuaciones de predicción a escala mensual que amplían las posibilidades del servicio especializado de pronóstico climático del Centro Meteorológico Provincial. Se predijo adecuadamente la estacionalidad de la temperatura, sin encontrar influencia significativa de los índices ENOS, NAO y SOI en la predicción de la variable mediante las ecuaciones de regresión.

Conseguir predecir, o al menos obtener algún tipo de información sobre cómo se van a comportar ciertos fenómenos climáticos, es de gran importancia para tomar medidas que permitan mitigar sus consecuencias. De Aragón (2013), refiere que la estadística proporciona herramientas fundamentales para este objetivo y en general, para el estudio del clima, permitiendo evaluar y cuantificar de forma objetiva el cambio del clima, así como estimar variaciones probables en el futuro.

Al realizar una revisión bibliográfica sobre los calendarios climáticos y los estudios de temperatura, se considera que las investigaciones consultadas tenían poco enfoque estadístico en el análisis de sus datos. Es por eso que en el próximo epígrafe se hará una introducción a los antecedentes históricos de la estadística y sus aplicaciones alcanzadas en el desarrollo de los años, para después conocer sobre algunas técnicas estadísticas que pudieran considerarse en la confección de calendarios climáticos.

1.2 ANTECEDENTES HISTÓRICOS DE LA ESTADÍSTICA. SUS APLICACIONES

En las últimas décadas se ha hecho evidente los grandes cambios y avances vertiginosos en las ciencias, así como la importancia que va adquiriendo la Estadística dentro de éstas debido a su amplia aplicación en diversas áreas de estudio y en el quehacer cotidiano de las personas. Si se quisiera conocer cómo es que se distribuyen datos o se busca un valor que represente al conjunto de datos, se tendría que calcular una serie de valores llamados estadísticos.

En la antigüedad, los astrónomos babilonios asentaron en tablillas de arcillas (Figura 1.1) registros sobre los movimientos de los astros y planetas, resolviendo un problema de estimación mediante el cálculo de la suma total de las observaciones y dividiéndolo por el número de datos (Plackett, 1970, citado por Chan, 2009). Los comienzos de la estadística pueden ser hallados en el antiguo Egipto, cuyos faraones lograron recopilar, hacia el año 3050 antes de Cristo (a.C.), datos relativos a la población y la riqueza del país. De acuerdo al historiador griego Heródoto, dicho registro de riqueza y población se hizo con el objetivo de preparar la construcción de las pirámides. En el mismo Egipto, Ramsés II hizo un censo

de las tierras con el objeto de verificar un nuevo reparto, esto debido a que el nivel del río Nilo subió y grandes extensiones de tierra que se encontraban a su alrededor se perdieron.



Figura 1.1. Registro del movimiento de los astros y planetas realizada por los astrónomos babilonios. Fuente: Chan (2009).

Los griegos también efectuaron censos, pero de manera periódica, estos con fines tributarios, sociales (división de tierras) y militares (cálculo de recursos y hombres disponibles). Una investigación histórica ha revelado que se realizaron 69 censos para calcular impuestos, determinar los derechos de voto y ponderar la potencia de guerra. En esta misma época, los pitagóricos realizaron una teoría aritmética, la cual consiste en calcular valores medios o promedios de una serie de números, distinguiendo tres casos: la media aritmética, la media geométrica y la media armónica (Collette, 1979, citado por Chan, 2009).

También, en épocas pasadas, cuando las tormentas en el mar causaban en las embarcaciones la pérdida de bienes, estos tenían que ser pagados mediante un acuerdo de todos los que tenían mercancía en el mismo buque. El daño causado por estas tormentas al cargamento se le conocía como “havaría” y la palabra llegó a aplicarse naturalmente al dinero que cada individuo tenía que pagar como compensación. De esta palabra latina se deriva la palabra *avarage* (promedio). La idea de promedio tiene sus raíces, por lo tanto, en una clase primitiva de seguros, de modo que se compartía el riesgo, no sólo entre los que arriesgaban sus bienes en un viaje particular, sino entre grupos mayores de comerciantes. La distribución de este riesgo se desarrolló, más tarde, eventualmente en una profesión aparte, lucrativa y experimentada (Newman, 1968, citado por Chan, 2009).

Durante los siglos oscuros en los que se elaboró la antigua civilización humana, la matemática no pasa de los primeros problemas que en la vida práctica se presentan sobre

la cantidad de las cosas. Tan pronto como los hombres se reunieron en sociedad fijaron sus posesiones recíprocas, bien por leyes, por convenciones generales la necesidad y el interés, estos dos grandes móviles de la industria humana, no tardaron en producir las artes de primera necesidad. Se construyeron cabañas, se aprendió a medir la extensión de los campos, se observó el curso de los astros, se vio que la tierra producía de forma natural y espontáneamente muchos frutos apropiados para la alimentación, pero que los productos eran más abundantes y útiles cuando la producción era secundada por un cultivo subordinado a la sucesión de las estaciones y de aquí nació la agricultura. Todas estas observaciones y estas prácticas, aunque fueron puramente rutinarias e informales, encierran el primer germen de la matemática.

La asiduidad que exigía la caza, la pesca y los trabajos de campo, no permitían elevar ideas generales y reflexivas. Cuando la inteligencia humana salió de su letargo, despertó el espíritu de la curiosidad; entonces el hombre vio con otros ojos el magnífico espectáculo que la naturaleza ofrece por doquier. Se estudiaron con atención los fenómenos de la naturaleza y se quisieron conocer sus causas; la geometría se desarrolló notablemente; la astronomía se enriqueció con observaciones regulares y con varios instrumentos destinados a multiplicarlos y a darle toda la exactitud posible; el comercio y el trato social habían hecho adelantar notablemente el cálculo numérico: todas las partes de las matemáticas hicieron grandes progresos.

Resulta muy importante cultivar la utilización de la estadística a la nueva generación dada su amplia aplicación en distintas áreas de estudio, que permite, entre muchas otras cosas, saber interpretar de manera correcta las informaciones que se presentan a diario en los diferentes medios, para formar opiniones más fundamentadas y tomar decisiones con criterios más objetivos (Sayritupac, 2013).

La estadística es una herramienta que permite obtener resultados en cualquier tipo de estudio, cuyos movimientos y relaciones, por su variabilidad intrínseca, no puedan ser abordados desde la perspectiva de las leyes deterministas. Esta se ocupa de recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sea una causa específica de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y en su caso formular predicciones (Barón, 1998).

Aunque es difícil dividir la estadística en partes separadas, una división clásica ha sido entre estadística descriptiva y estadística inferencial. La estadística descriptiva se utiliza para

describir los datos, resumirlos y presentarlos de forma que sean fáciles de interpretar. El interés se centra en el conjunto de datos dados y no se plantea el extender las conclusiones a otros datos diferentes. La estadística inductiva o inferencial trata de obtener conocimientos sobre ciertos conjuntos extensos o poblaciones, a partir de la información disponible de un subconjunto de tal población llamada muestra.

Hasta 1900 la estadística se restringía a la estadística descriptiva, que a pesar de sus limitaciones hizo grandes aportaciones al desarrollo de la ciencia. A partir de esa época comenzaría la inferencia estadística, con los trabajos de Fisher, Pearson y sus colaboradores. Con el desarrollo de la informática en la segunda mitad del siglo XX, puesto que era fácil analizar grandes muestras, ya no había por qué limitarse a los métodos estadísticos basados en distribuciones conocidas, cuya principal aplicación eran las pequeñas muestras. Tampoco había por qué restringirse a analizar una o unas pocas variables, porque el tiempo de cálculo se había eliminado y era preferible aprovechar toda la información disponible.

Con todo ello surge una nueva filosofía en los estudios estadísticos: el análisis exploratorio de datos, introducida en 1977 por John W. Tukey en su libro de “Análisis exploratorio de datos”. Se pensaba que para obtener conclusiones de los datos era preciso recurrir a la inferencia (modelo confirmatorio). Si se parte de esta hipótesis, que es previa a la recogida de datos, se calculan los estadísticos (media, coeficiente de correlación en la muestra) que sirven para aceptar o rechazar ciertas hipótesis establecidas de antemano. Al contemplar solamente dos alternativas, (confirmación o no de la hipótesis), los datos no se exploraban para extraer cualquier otra información que pueda deducirse de los mismos.

El análisis de datos es sólo una parte (aunque importante) en el proceso de investigación. Este proceso comienza con la definición de un problema, el estudio de la bibliografía relacionada y el diseño del trabajo de campo, en el cual se recoge datos para el estudio, mediante encuestas, observación o mediciones. Una vez recogidos los datos y planteadas las preguntas de investigación, el análisis de datos permitirá contestar estas preguntas si están bien planteadas y se han recogido los datos necesarios (Batanero & Díaz, 2008).

Dentro de esta formación general para la buena cultura estadística de un ciudadano, las medidas de tendencia central, media, mediana y moda se presentan como conceptos que deben formar parte necesaria de dicha formación y más aún en los profesionales de todas las especialidades, por ser conceptos estadísticos básicos y de gran aplicación práctica en la vida cotidiana y profesional (Sayritupac, 2013).

Del estudio de Mayén (2009) se considera las siguientes razones generales sobre la importancia de hacer un estudio sobre la media, mediana y la moda: la complejidad de estos conceptos (a pesar de ser conceptos básicos), su importancia para la construcción posterior de otros conceptos estadísticos (resulta útil para alumnos que llevan cursos de estadística como parte de su formación profesional) y su uso frecuente en la vida diaria. El próximo epígrafe introducirá estos conceptos más profundamente.

1.3 TÉCNICAS ESTADÍSTICAS. DEFINICIONES Y CONSIDERACIONES.

En los trabajos estadísticos, cuando se tiene cierta información válida sobre una determinada variable de estudio, resulta interesante hacer un análisis de los datos obtenidos que permitan describirlos, resumirlos y presentarlos de manera que sean fáciles de interpretar. En este epígrafe se abordarán diferentes técnicas en el ámbito de la estadística que permiten corroborar lo anterior y que son utilizadas o pueden servir de herramientas para la confección de calendarios climáticos.

1.3.1 SERIES TEMPORALES

Una serie temporal es un conjunto de observaciones ordenadas en el tiempo, que pueden representar la evolución de una variable a lo largo de él. El objetivo del análisis de una serie temporal es el conocimiento de su patrón de comportamiento, para así prever su evolución futura, suponiendo que las condiciones no variarán. Dado que no se trata de fenómenos deterministas, sino sujetos a una aleatoriedad, el estudio del comportamiento pasado ayuda a inferir la estructura que permita predecir su comportamiento futuro, pero es necesaria una gran cautela en la previsión debido a la inestabilidad del modelo.

La particular forma de la información disponible de una serie cronológica (se dispone de datos en períodos regulares de tiempo) hace que las técnicas habituales de inferencia estadística no sean válidas para estos casos, ya que nos encontramos ante n muestras de tamaño 1 procedentes de otras tantas poblaciones de características y distribución desconocidas (Terrádez & Juan, 2002).

Los objetivos que persiguen las series temporales son: entender la estructura especial de la información en una serie temporal, comprender qué está sucediendo con los datos (patrón de comportamiento) y predecir valores futuros. La forma más sencilla de iniciar el análisis de una serie temporal es mediante su representación gráfica, usándose un sistema cartesiano en el que los valores o períodos de tiempo se llevan al eje de las abscisas y los valores de la serie, y_t , se llevan al eje de las ordenadas.

El resultado es un diagrama de dispersión, con la particularidad de que el eje de abscisas se reserva siempre a la misma variable: el tiempo. El objetivo del análisis de series temporales es doble. Por un lado, se busca explicar las variaciones observadas en la serie en el pasado, tratando de determinar si responden a un determinado patrón de comportamiento. Y por otro, si se consigue definir ese patrón o modelo, se intentará predecir el comportamiento futuro de la misma (Sánchez, 2012).

Una forma de visualizar la tendencia es mediante el suavizado de la serie. La idea central es definir a partir de la serie observada una nueva serie que suavice los efectos ajenos a la tendencia (estacionalidad, efectos aleatorios) de manera que se pueda determinar la dirección de la tendencia (Figura 1.2).

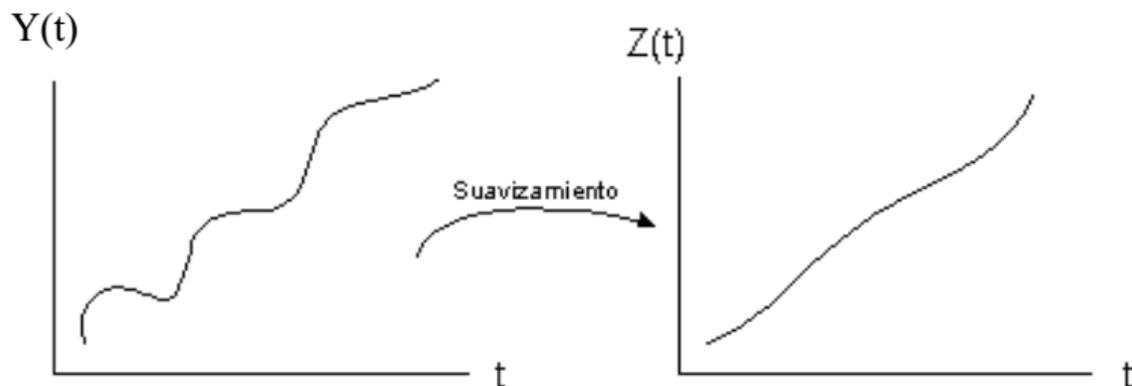


Figura 1.2. Suavizado de una curva. Fuente: Sánchez (2012).

Lo que se hace es usar una expresión lineal que transforma la serie $Y(t)$ en una serie suavizada $Z(t) = F(Y(t)), t = 1, \dots, n$

$$Y(t) \rightarrow F \rightarrow T(t)$$

De tal modo que $F(Y(t)) = T(t)$; donde la función F determina Filtro Lineal. El filtro lineal más usado es la media móvil, cuyo objetivo es eliminar de la serie los componentes estacionales y accidentales.

1.3.2 MEDIDAS DE TENDENCIA CENTRAL.

Dentro de los datos estadísticos se encuentran las medidas de tendencia central (llamados así por su característica principal de acumularse al centro de la distribución de datos), siendo las más comunes, la media, la mediana y la moda. Estas medidas describen un valor típico o representativo del conjunto de datos y señalan las tendencias o características del

mismo. Para un mayor conocimiento y precisión se muestra algunas definiciones en este epígrafe.

Media aritmética

Según Novaes (2011), la media (media aritmética) de un conjunto de datos N de números $X_1, X_2, X_3, \dots, X_N$ es el valor numérico que se obtiene dividiendo la suma total de los valores observados de una variable entre el número de observaciones, se representa por \bar{x} y se define como:

$$\bar{x} = \frac{X_1 + X_2 + X_3 + \dots + X_N}{N} = \frac{1}{N} \sum_{i=1}^N X_i$$

Strauss y Bichler (1988) proponen una clasificación de las propiedades de la media aritmética en tres aspectos:

Estadístico:

- ✓ La media está localizada entre los valores extremos.
- ✓ La suma de las desviaciones de la media es cero.
- ✓ La media se ve influenciada al añadir otros datos distintos de la media.

Abstracto:

- ✓ La media no es necesariamente igual a un valor que se haya sumado.
- ✓ La media puede ser una fracción que no sea posible en la realidad.
- ✓ Cuando calculamos la media, si aparece un valor cero, este se debe tener en cuenta.

Representativo:

- ✓ La media es un valor representativo de los valores que se están promediando.

La media aritmética puede ser presentada desde diversos puntos de vista: como el resultado de un cálculo, como operador que a una distribución asigna un número y como un resumen estadístico o parámetro que caracteriza una distribución (Cobo & Batanero, 2004, citado por Salazar, 2015).

Mediana

La mediana es el valor de los datos que ocupa la posición central cuando los datos se ordenan según su tamaño, esencialmente lo que hace es separar al conjunto de datos ordenados en dos subconjuntos de igual tamaño.

Según lo citado por Sayritupac (2013), dado que la mediana hace referencia al valor que ocupa la posición central o del medio, para que haya tal centro o medio por lo menos tiene que haber un orden (por eso no es posible calcular la mediana en variables cualitativas nominales). Esta medida es representada por Me y para su cálculo es preciso distinguir cuando se tiene datos agrupados y sin agrupar, así como también, si el número de datos “n” es par o impar. En efecto, para datos sin agrupar se tiene:

Si es “n” par, la mediana es el valor que resulta de calcular el promedio de los datos que ocupan las posiciones $\frac{n}{2}$ y $\frac{n+1}{2}$.

Si “n” fuera impar, Me resulta ser el dato que ocupa la posición $\frac{n+1}{2}$.

Moda

La moda es el valor más frecuente de la variable estadística, es decir, el valor que más se repite. La moda puede no existir, y si existe, no ser única. Si la distribución presenta una sola moda se llamará *unimodal*, si existen dos *bimodal* y si ningún valor de la variable se repite más veces que los otros o si hay más de dos que se repiten más veces, se dice que no existe moda (Murray R. *et al.*, 2009, citado por Sayritupac, 2013).

Consideraciones entre las medidas de tendencia central

La media, mediana y moda, como representantes de un conjunto de datos, brindan cierta información de cómo se comportan los datos en una distribución; sin embargo, es necesario saber cuándo una de estas medidas resulta ser más representativa o adecuada frente a otra, pues cada una de ellas posee ciertas características particulares. Para esto, si se conoce las propiedades de estas medidas, se brinda una mejor información para el buen manejo las mismas. Las propiedades se han clasificado como: numéricas, algebraicas y estadísticas (Cobo, 2003, citado por Sayritupac, 2013).

Propiedades Numéricas

- ✓ La media, mediana y moda de un conjunto de datos son siempre valores pertenecientes al rango de la variable (estas nunca resultan mayores al máximo de los datos, ni menores al mínimo de los mismos).
- ✓ Un representante de un conjunto de datos puede ser uno de ellos (en el caso de la moda o de la mediana, si hubiera un número impar de datos) o algún valor, que no necesariamente coincide con uno de los datos, obtenido por un promedio adecuado (la media aritmética o la mediana, cuando se tiene un número par de datos).

- ✓ En el cálculo de la media se tienen en cuenta todos los valores de los datos, pero no en la mediana y en la moda.
- ✓ El valor numérico de la media cambia cuando se cambia cualquier dato, o cuando se añade un dato (incluso el cero) al conjunto y todos los otros se mantienen igual, salvo que el dato añadido sea igual al de la media anterior.

Propiedades Algebraicas

- ✓ Las medidas de tendencia central conservan los cambios de origen y escala (si sumamos o multiplicamos a cada uno de los datos por cierto valor; la media, mediana y moda son también sumados o multiplicados por ese mismo valor).
- ✓ La media de la suma de dos o más variables, es igual a la suma de las medias de dichas variables (esta propiedad sólo se cumple cuando las variables tienen el mismo número de elementos).
- ✓ La moda puede no existir o, si existe, no ser única. La media y la mediana siempre existen en datos numéricos.

Propiedades Estadísticas

- ✓ La media, mediana y moda son representantes de un colectivo (estas medidas nos brindan información de todo el conjunto y no de uno de ellos en particular pues las medidas de tendencia central son valores numéricos que localizan, de alguna manera, el centro de un conjunto de datos).
- ✓ La media es un estadístico poco resistente, muy sensible a la variación de los datos, especialmente en los valores atípicos (en estos casos, la mediana se presenta como un mejor representante pues solo considera el o los valores centrales de la distribución. Tanto la mediana como la moda son estadísticos más resistentes).
- ✓ Para datos agrupados en intervalos con alguno de ellos abierto, también es preferible la mediana a la media.
- ✓ Existen modas tanto para variables cuantitativas como cualitativas. En variables cualitativas nominales, la moda es el único estadístico calculable y que adquiere sentido. Si la variable es continua sólo se puede hallar un valor aproximado de la moda.

1.3.3 OTRAS MEDIDAS A CONSIDERAR

Los promedios son valores representativos de un conjunto de datos y que tienden a encontrarse en el centro del conjunto; los más usados son la media aritmética, la mediana, y la moda. Como la media es un estadístico poco resistente, es por eso que aparecen otros

tipos de media que tienden a eliminar estos valores atípicos, entre las que se encuentran la media móvil y la media tijerada.

Media móvil

La herramienta de análisis media móvil proyecta valores en el período de pronósticos, basándose en el valor promedio de la variable calculada durante un número específico de períodos anteriores. Una media móvil proporciona información de tendencias que se vería enmascarada por una simple media de todos los datos históricos. Las medias móviles permiten suavizar las fluctuaciones de los valores, con la finalidad de dar una interpretación más compacta de la situación. El período de tiempo durante el cual se calcula la media, es vital para la representación que deriva de dicho cálculo. Así, cuanto mayor sea el período, la media representará un conjunto de valores mucho más general.

Las medias móviles, son un indicador que admiten doscientos mil parámetros distintos, para al final medir lo mismo: la tendencia y la fuerza de esta. Las medias móviles habituales son tres: simples, exponenciales y ponderadas (Illescas, 2015).

- Media móvil simple

Es una de las herramientas básicas del análisis técnico y de ella se derivan otros indicadores y conceptos de uso muy frecuente. La media simple otorga la misma relevancia a todos los datos. Como su nombre indica, la media móvil simple toma para su cálculo la suma de los datos a los que hace referencia su longitud y los divide por el período indicado, es decir, utiliza como pronóstico para el siguiente período, el promedio de los “n” valores de los datos más recientes de la serie de tiempo, matemáticamente:

$$Promedio\ Móvil = \frac{\sum(n\ \text{valores de datos recientes})}{n}$$

El término móvil indica que conforme se tenga disponible una nueva observación de la serie de tiempo, se reemplaza la ecuación más antigua de la ecuación y se calcula un nuevo pronóstico. La variable “n” es una indicación de cuántos períodos habrá que tomarse para calcular el promedio, dependiendo de cuántos elementos tiene la serie.

- Media móvil exponencial

La media móvil exponencial es un indicador que se deriva del uso de la Media Móvil Simple en un intento de otorgar una importancia progresiva a las cotizaciones más recientes utilizando un sistema de ponderación o suavizado exponencial y que tiene en cuenta todos los datos de la serie. Es más efectiva que la media móvil ponderada para seguir las

tendencias. En este caso todos los datos no tienen la misma importancia, se les da más peso a los datos recientes y reacciona más rápidamente a los cambios recientes, y resta importancia a los más antiguos. La fórmula es la siguiente:

$$MME = P \times K + MME \text{ (de ayer)} \times (1 - K)$$

P: el valor de hoy, que puede ser cierre, apertura, etc. Según se quiera aplicar el parámetro.

N: es el número de datos escogido.

K: es dos, dividido entre la suma de $N + 1$. $K = \frac{2}{(N+1)}$

A diferencia de la media móvil simple y la media móvil ponderada, la Media Móvil Exponencial efectúa sus giros antes y por tanto es más sensible a los cambios. Por su forma de cálculo la media móvil es un indicador seguidor, no líder y presenta un retraso de tantas sesiones como indica su período de cálculo.

- Media móvil ponderada

La media móvil ponderada, deriva de la media móvil simple. Intenta otorgar importancia progresiva a las cotizaciones más recientes, respecto a las primeras del período al que se refiere la media. Es muy parecida a la exponencial, pero da aún mayor importancia a los últimos datos, que los que otorga la exponencial. En la práctica se pega un poco más a los últimos datos que la media exponencial, pero gráficamente difiere muy poco.

La técnica de medias móviles construye una nueva serie a partir de la media de un número determinado de datos, en la que se va añadiendo sucesivamente un dato nuevo y quitando, al mismo tiempo, el más antiguo de los datos incluidos en la media anterior.

La expresión general de una media móvil de orden s consistiría en calcular una serie que para cada momento t toma el siguiente valor:

$$M_t = \frac{y_t + y_{t-1} + \dots + y_{t-s+1}}{s}$$

La técnica se aplica óptimamente en series sin tendencia ni estacionalidad. En ese caso, la predicción se realiza con la última media móvil calculada, es decir:

$$\hat{y}_{t+1} = M_t$$

Cuando hay tendencia, debe corregirse el efecto del incremento medio anual, o sea, para predecir valores mensuales:

$$\hat{y}_{t+1} = M_t + \frac{y_t - y_{t-12}}{12}$$

Este resultado es similar a calcular:

$$\hat{y}_{t+1} = y_t + \Delta\hat{y}_{t+1} = y_t + M_t = y_t + \frac{y_t - y_{t-12}}{12}$$

Cuando la serie presenta estacionalidad, la media móvil debe ser corregida por “coeficientes de estacionalidad”.

Versiones alternativas consistirían en atribuir ponderaciones distintas a cada valor incluido en la media móvil (normalmente más ponderación con el dato más próximo), es decir, calcular *medias móviles ponderadas*, y *dobles medias*. No existen criterios generales sobre el orden adecuado de la media móvil.

Media tijereada

La media tijereada, recortada o truncada tiene por finalidad evitar la distorsión que las puntuaciones extremas no compensadas causan a la media aritmética. Es una medida de tendencia central estadística, similar a una media aritmética y una mediana. Es un estimador útil porque es menos sensible a valores atípicos que la media y aun así da un razonable estimador de la tendencia central o promedio para numerosos modelos estadísticos. En este sentido es reconocido por ser un estimador robusto.

Para obtener la media recortada, se ordenan los datos por sectores en orden ascendente. Para el cálculo de la media en este caso, previamente se descartan porciones de la distribución de probabilidad o muestra en el extremo inferior y superior, típicamente se descarta igual cantidad en ambos extremos, para la mayoría de los usos en estadística se elimina del 5 % al 25 % de los elementos de la muestra en los extremos (Rothenberg *et al.*, 1966).

En algunas regiones de Europa Central a veces se le denomina promedio Windsor, pero no se debe confundir este nombre con un promedio Winsorizado: en este último, las observaciones que han sido descartadas son reemplazadas por el más grande o el más pequeño de los valores que restan.

Cuando se debe determinar la media truncada de una muestra, pero no es posible hacerlo en forma precisa, lo mejor es calcular las dos medias truncadas más próximas, y luego interpolar (por lo general en forma lineal). Por ejemplo, si se debe calcular la media truncada al 15 % de una muestra conteniendo diez elementos, primero se calcula la media truncada al 10 % (eliminando un elemento en cada extremo de la muestra), luego se calcula la media

truncada al 20 % (eliminando dos datos en cada extremo) y finalmente se interpola para calcular la media truncada al 15 % (Bloch, 1966).

En el tratamiento de datos meteorológicos las técnicas estadísticas presentadas, en dependencia de las características de cada una y del objetivo que persiga el investigador, permiten resultados aceptables al tratarse de estudios climáticos. Estas se pueden utilizar también para determinar un valor futuro a partir de las salidas de modelos climáticos. En el siguiente epígrafe se presentará la verificación de modelos meteorológicos y se explicará el PRECIS CARIBE, el cual fue utilizado en esta investigación.

1.4 MODELO DE CIRCULACIÓN REGIONAL PRECIS CARIBE. VERIFICACIÓN DE MODELOS METEOROLÓGICOS

El IPCC (2001) refiere a un “escenario” como una descripción coherente y consistente de cómo el sistema climático de la Tierra puede cambiar en el futuro. Jones *et al.* (2004) lo describen como un resultado plausible y consistente que ha sido construido para explicar las consecuencias potenciales de la influencia de las actividades humanas sobre el clima. De acuerdo con el IPCC (2007), los escenarios climáticos se definen como “una representación posible y simplificada del clima futuro, basadas en un conjunto de relaciones climatológicas que ha sido construida expresamente para investigar las posibles consecuencias del cambio climático antropogénico, y que en muchas ocasiones sirve como materia prima para modelos de impacto”.

Basados en determinadas condiciones futuras sobre emisiones de gases de efecto invernadero se generan los escenarios de cambio climático, que describen un posible estado futuro del mundo (Palma *et al.*, 2009). Los más utilizados son los construidos a partir de las salidas de los Modelos de Circulación General (MCG) debido a las diversas escalas espaciales y niveles de complejidad a los que pueden estar sometidos (Ruiz, 2007). En Cuba se utiliza las salidas del Modelo de Circulación General Regional denominado *Providing Regional Climates for Impacts Studies* (PRECIS, por sus siglas en inglés) para generar escenarios futuros. Se le llama PRECIS-CARIBE pues las corridas del modelo son utilizadas para la región de Centroamérica, México y el Caribe.

Según Palma *et al.* (2009), el Modelo de Circulación Regional PRECIS, es un modelo climático atmosférico-terrestre de área limitada y alta resolución, en él se encuentran descritos: el flujo dinámico, el ciclo de sulfuro atmosférico, nubes y precipitación, los procesos radiativos, la superficie terrestre y el espesor del suelo. Tiene la resolución típica de un modelo de clima regional de alrededor de 50 km en la horizontal (resolución de 0.44°

x 0.44°), y considera condiciones de la atmósfera bajo equilibrio hidrostático y tiene una representación completa de la fuerza de Coriolis.

Aunque la información de los MCG tiene muchas ventajas, la resolución espacial es de algunos cientos de kilómetros. Tal información no es totalmente útil para algunos territorios pequeños, como los estados insulares. Existe, sin embargo, la forma de reducir la escala de los resultados de los MCG, mediante técnicas estadísticas o dinámicas, las cuales se les conoce como regionalización o “*downscaling*”. Los modelos climáticos regionales son la herramienta fundamental utilizada para realizar la reducción de escala dinámica; de ahí que el método PRECIS utilice en uno de sus experimentos, salidas del MCG Echam4 desarrollado a partir del modelo para pronóstico del tiempo del *European Centre for Medium-Range Weather Forecasts* (ECMWF), como condiciones de frontera.

Los escenarios climáticos proporcionan representaciones verosímiles de contextos futuros, contruidos a partir de las relaciones entre las variables del clima observadas y proyectadas, y suelen utilizarse como fundamento para la elaboración de patrones de impacto futuro (IPCC, 2007), son un ingrediente indispensable para el desarrollo de estrategias de adaptación y mitigación ante el cambio climático (Conde *et al.*, 2008). Las principales fuentes de incertidumbre en los escenarios de cambio climático (ECC) se observan principalmente en las emisiones de gases de efecto invernadero, en la variabilidad natural y en los modelos climáticos.

Los ECC usados para describir un posible estado futuro del mundo están basados en las condiciones climáticas observadas (generalmente durante un espacio de treinta años) denominadas “escenario base” (Palma *et al.*, 2009). La Organización Meteorológica Mundial (OMM) recomendó el uso del período de 30 años disponible más reciente, que finalizara en el año más próximo que terminara en 0 (OMM, 2007, citado por SINAVEF, 2009), por lo que en esta investigación se trabaja como escenario base 1981-2010.

Debido a su alta resolución, se pueden crear escenarios de cambio climático nacionales para países pequeños como los que componen el Caribe. Además, con el modelo PRECIS se proveen software para el post procesamiento, análisis y graficación de los datos generados por el modelo (Palma *et al.*, 2009). En Cuba, se utiliza el modelo climático regional PRECIS CARIBE, con salidas del MCG Echam 4 y el Hadley, bajo condicionantes del tipo A2 y B2. El Informe Especial del IPCC sobre escenarios de emisiones refiere a los escenarios agrupados en cuatro familias (A1, A2, B1, B2) que exploran vías de desarrollo alternativas incorporando toda una serie de fuerzas originantes demográficas, económicas

y tecnológicas, junto con las emisiones de gases de efecto invernadero (GEI) resultantes (IPCC, 2000).

El Cuarto Informe del IPCC (2007) propone la línea argumental A1 como un crecimiento económico mundial muy rápido, un máximo de la población mundial hacia mediados de siglo, y una rápida introducción de tecnologías nuevas y más eficientes. Se divide en tres grupos, que reflejan tres direcciones alternativas de cambio tecnológico: intensiva en combustibles fósiles (A1FI), energías de origen no fósil (A1T), y equilibrio entre las distintas fuentes (A1B).

B1 describe un mundo convergente, con la misma población mundial que A1, pero con una evolución más rápida de las estructuras económicas hacia una economía de servicios y de información. B2 describe un planeta con una población intermedia y un crecimiento económico intermedio, más orientada a las soluciones locales para alcanzar la sostenibilidad económica, social y medioambiental. A2 describe un mundo muy heterogéneo con crecimiento de población fuerte, desarrollo económico lento, y cambio tecnológico lento (Figura 1.3).

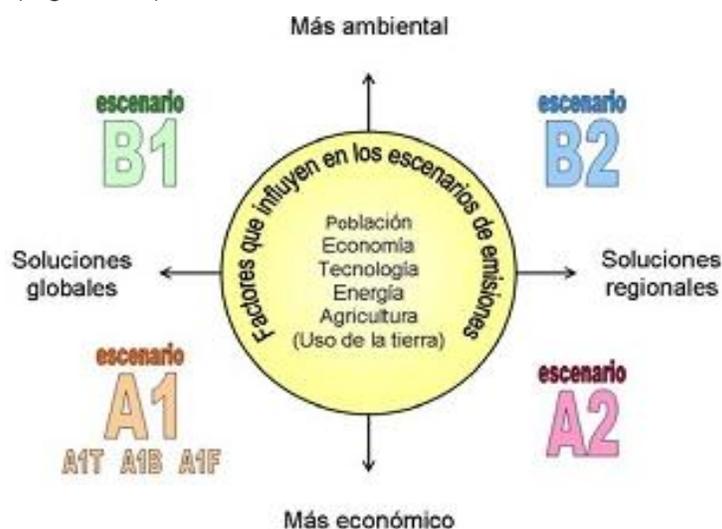


Figura 1.3. Familias de Escenarios de emisiones. Fuente: IPCC (2007).

Verificación de modelos meteorológicos

La evaluación de los modelos ha sido tradicionalmente subjetiva, teniendo en consideración la experiencia del pronosticador para determinar la exactitud de los mismos. Esta verificación se puede hacer por comparaciones simples, entre las salidas del modelo, con los análisis gráficos y las imágenes de satélites y diagramas de los datos de superficie y de altura. Sin embargo, estas comparaciones pueden ser engañosas puesto que contienen a menudo sesgos e impresiones personales.

En la medida en que los modelos son más complejos y cambian con más frecuencia, se requieren medios más objetivos de evaluar el funcionamiento de ellos, tanto para eliminar sesgos específicos como para determinar el impacto del modelo en el cambio de las correcciones de errores conocidos. Las herramientas de evaluación de la exactitud del modelo, también llamadas verificación del modelo de diagnóstico o del modelo, se han desarrollado para proporcionar una medida objetiva de su habilidad, que puede ser utilizada por los pronosticadores (León, 2005).

La autora también destaca que el método de verificación que se elige depende del tipo de pronóstico, del tipo de variable a predecir y del contexto en que se utilizarán los resultados de la verificación. Existen dos tipos de pronósticos: categóricos o determinísticos, que consisten en afirmaciones que indican que ciertos eventos van a ocurrir o no (por ejemplo “esta noche va a llover” o “mañana la temperatura subirá hasta 25 °C) y los probabilísticos, que establecen la probabilidad que ocurra un evento (por ejemplo “esta noche hay un 80 % de probabilidad que llueva” o “hay un 10 % de probabilidad que la temperatura suba a más de 3 °C sobre la normal”).

Las variables pueden ser escalares, cuando toman un valor en una escala continua y discreta (como la temperatura o la dirección del viento) o no escalares, cuando definen eventos que pueden ocurrir o no (como el tipo de precipitación: lluvia, nieve o aguanieve, o el estado del cielo: nublado, parcialmente nublado o despejado / soleado).

Los pronósticos de alta calidad exhiben una estrecha correspondencia con las observaciones. Las tres medidas más comunes utilizadas para valorar la calidad son:

- ✓ Fiabilidad: entre el grado de correspondencia entre los pronósticos promedio y las observaciones promedio; esto a veces también se denomina “sesgo” porque es una medida de la tendencia sistemática en un conjunto de pronósticos hacia la sobre-estimación o hacia la sub-estimación.
- ✓ Exactitud: el grado promedio de correspondencia entre pronósticos individuales y las correspondientes observaciones. Esto significa que incluso un sistema de pronóstico equilibrado que tiene una fiabilidad ideal puede tener todavía un grado de exactitud pobre.
- ✓ Habilidad: La exactitud de un pronóstico en comparación con la exactitud de un “pronóstico de referencia”, es decir, pronósticos basados en la climatología, la persistencia o la probabilidad.

Otros métodos para verificar modelos meteorológicos son a partir de inferencias estadísticas entre las corridas del modelo con variables reales. De los distintos métodos de estimación se recurrirá al contraste de hipótesis en el trabajo con los datos del PRECIS CARIBE y de la estación meteorológica. En el epígrafe a continuación se hará un breve resumen de técnicas para el análisis de experimentos: el análisis de varianza y la prueba de hipótesis.

1.5 PRUEBA DE HIPÓTESIS. ANÁLISIS DE VARIANZA

En ocasiones, el objetivo de la estadística es hacer inferencias con respecto a parámetros poblacionales desconocidos, basadas en la información obtenida mediante datos muestrales. Estas inferencias se expresan en una de dos maneras, como estimaciones de los parámetros respectivos o como pruebas de hipótesis referentes a sus valores al tomar decisiones con base en datos experimentales. Si se toma decisiones es porque hay alternativas y cada una de estas es formalizada como una “hipótesis estadística” y el proceso mediante el cual se enfrentan o confrontan las hipótesis al tomar como punto de apoyo los datos muestrales constituye lo que se denomina “prueba o contraste de hipótesis” (Dante, 2002).

Para evaluar si un conjunto (dos o más) de medias poblacionales son iguales y en caso que no lo sean, identificar cuál o cuáles son diferentes, desde un punto de vista estadístico y a partir de la información muestral o experimental que se tiene sobre poblaciones o conjuntos de datos generados por un criterio de clasificación, se formulan dos hipótesis: la hipótesis nula (H_0) y la hipótesis alternativa (H_1), las que se escriben de la siguiente forma:

$$H_0: \mu_1 = \dots = \mu_a \text{ Las medias de los } a \text{ grupos son todas iguales}$$

$$H_1: \text{al menos una de las "a" medias poblacionales es distinta}$$

donde a representa la cantidad de poblaciones a comparar a través de sus medias. Estas a poblaciones que están involucradas en el estudio, deben distinguirse o estar separadas en base a algún criterio de clasificación. De esta manera, si se encuentran diferencias entre los valores esperados de todas o de al menos un par de medias, se podrá inferir sobre la causa de los efectos que generan las diferencias (UNC, 2017).

Cuando se tiene que contrastar una hipótesis estadística es conveniente seguir un esquema el cual incluye las siguientes etapas:

1. Enunciado de la hipótesis nula y alternativa

2. Elección del nivel de significación (α)
3. Selección del estadístico de prueba.
4. Cálculo del estadístico.
5. Exposición de las conclusiones.

Análisis de varianza

El análisis de la varianza (ANOVA) se debe al estadístico-genético Sir Ronald Aylmer Fisher (1890-1962), autor del libro *Statistics Methods for Research Workers* publicado en 1925 y pionero de la aplicación de métodos estadísticos en el diseño de experimentos. El objetivo principal del ANOVA es contrastar si existen diferencias entre las distintas medias de los niveles de las variables (factores). Muchas veces interesa saber qué medias difieren entre sí después de realizar el ANOVA y los contrastes de comparaciones múltiples que proporcionan información detallada sobre las diferencias entre las medias.

Para este objetivo una primera intuición llevaría a realizar las correspondientes pruebas *t-Student* entre todas las posibles parejas de grupos. Existen diversos métodos para ajustar el tipo de error y conseguir que efectivamente el error conjunto no sea superior al 5 %. La primera aproximación es debida a Fisher, quien propuso que sólo se han de comparar las diferencias entre medias 2 a 2 si el precedente ANOVA ha resultado significativo. Estas comparaciones *a posteriori* se realizan sin corrección alguna. Este método es conocido como LSD (*Least Significant Difference*).

En el caso de que no se cumplan las suposiciones del análisis de la varianza, es necesario aplicar la prueba de Kruskal-Wallis para el contraste de k medianas. La prueba de Kruskal-Wallis fue propuesta por William H. Kruskal (1919-2005) y W. Allen Wallis (1912- 1998) en el artículo "*Use of ranks in one-criterion variance analysis*" publicado en el "*Journal of American Statistics Association*" en 1952 (Salvador, 2003).

1.5.1 COMPARACIÓN DE VARIAS MUESTRAS EN EL STATGRAPHICS

En el procesador estadístico STATGRAPHICS CENTURION se pueden obtener intervalos de confianza y contrastes de hipótesis para poblaciones normales (medias y desviaciones). Con el contraste de hipótesis se realizan inferencias sobre la media y la varianza poblacionales de una variable a partir de los datos muestrales. Además, permite discrepar la suposición de que la media y la mediana tomen un valor concreto (hipótesis nula) frente a que tomen un valor diferente (hipótesis alternativa). Uno de los test utilizados para la

comparación de medias en el STATGRAPHICS se basa en la distribución *t-Student* (García, 2002).

La prueba ANOVA permite comparar las medias de r grupos ($r \geq 2$) y presupone que las varianzas de los grupos son iguales y que los residuos o errores son aleatorios, independientes e idénticamente distribuidos siguiendo una ley normal con media 0 y desviación constante. La hipótesis nula de la prueba ANOVA es H_0 y la alternativa H_1 . Esta prueba se basa en la comparación de las sumas de cuadrados medias, debidas a la variabilidad entre grupos y a la variabilidad intra grupos (dentro de los grupos). Ambas sumas son estimaciones independientes de la variabilidad global, de manera que si el cociente entre la primera y la segunda es grande, se tendrá mayor probabilidad de rechazar la hipótesis nula. Este cociente sigue una distribución F con $r - 1$ y $n - r$ grados de libertad (Salvador, 2003).

Las sumas de cuadrados son un paso previo para el cálculo del ANOVA. La suma de cuadrados entre grupos *SCE*, la suma de cuadrados dentro de grupos *SCD* y la suma de cuadrados total *SCT* se calculan del siguiente modo:

$$SCE = \sum_{j=1}^r n_j (\bar{x}_j - \bar{x})^2$$

$$SCD = \sum_{j=1}^r \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 = \sum_{j=1}^r \sum_{i=1}^{n_j} x_{ij}^2 - \sum_{j=1}^r n_j \bar{x}_j^2$$

$$SCT = \sum_{j=1}^r \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2$$

Donde r al número de grupos, n_j el número de individuos en cada grupo ($j = 1, \dots, r$), \bar{x}_j la media de cada grupo y \bar{x} la media global.

Los grados libertad entre grupos *GLE*, dentro de los grupos *GLD* y total *GLT* se calculan:

$$GLE = r - 1, GLD = n - r, GLT = n - 1$$

El cuadrado medio entre grupos *CME* y el cuadrado medio dentro de grupos *CMD* se determinan por:

$$CME = \frac{SCE}{GLE}$$

$$CMD = \frac{SCD}{GLD}$$

El estadístico de contraste para realizar la prueba ANOVA se construye de la forma siguiente:

$$F = \frac{CME}{CMD}$$

que se distribuye según una F-Snedecor con GLE grados de libertad del numerador y GLD grados de libertad del denominador.

Una medida relativa de la variabilidad explicada por los grupos es el cociente:

$$R^2 = \frac{SCE}{SCT}$$

que se denomina coeficiente de determinación y estará entre cero y uno. Mientras más próximo esté de uno, más variabilidad explica el modelo, y por tanto menos variabilidad no explicada o residual.

Con la información anterior se conforma la tabla del ANOVA:

	Suma de cuadrados	G.L.	Cuadrado medio	F-valor	P-valor
Entre Grupos	SCE	GLE	CME	F	p
Dentro Grupos	SCD	GLD	CMD		
Total	SCT	GLT			

La prueba de Kruskal-Wallis es el método más adecuado para comparar poblaciones cuyas distribuciones no son normales. Incluso cuando las poblaciones son normales, este contraste funciona muy bien. También es adecuado cuando las desviaciones típicas de los diferentes grupos no son iguales entre sí. Las hipótesis de la prueba de Kruskal-Wallis son:

H_0 : las k medianas son todas iguales

H_1 : al menos una de las medianas es diferente

Para cada observación se le asigna el rango según el orden que ocupa la observación en el conjunto total de los datos, asignando el rango medio en caso de empates. Para cada grupo $m = 1, \dots, r$, siendo r el número de grupos, se define R_m como la suma de rangos de cada grupo m . El valor medio de los rangos $E[R_m]$ se calcula como:

$$E[Rm] = \frac{n_m(n + 1)}{2}$$

Y el rango medio \bar{R}_m como:

$$\bar{R}_m = \frac{Rm}{n_m}$$

El estadístico de contraste de Kruskal-Wallis (H') se calcula como:

$$H' = \frac{\frac{12}{n(n+1)} \sum_{m=1}^r \frac{1}{n_m} [Rm - E[Rm]]^2}{1 - \frac{\sum_{j=1}^k (d_j^3 - d_j)}{n^3 - n}}$$

siendo d_j el número de empates en $j = 1, \dots, k$ siendo k el número de valores distintos de la variable respuesta, que sigue una distribución Chi-Cuadrado con $r - 1$ grados de libertad.

CONCLUSIONES DEL CAPÍTULO

- Los calendarios climáticos realizados utilizan como metodología o herramienta la propuesta de Soler y Martín a través de la media móvil, sin encontrarse otras proposiciones matemáticas y estadísticas para la confección de los mismos.
- La moda y la mediana son estadísticos más resistentes que la media, la cual es muy sensible a la variación de los datos, pero con el uso de la media móvil y la media tijerada se eliminan los valores atípicos, se proporciona mayor relevancia a todos los datos y se evita la distorsión de las puntuaciones extremas.
- Los ECC tienen fuentes de incertidumbre sobre todo los elaborados para territorios pequeños o estados insulares, por lo que la evaluación de la exactitud del modelo y la comparación de sus salidas con datos reales son herramientas importantes para descartar sesgos.

CAPÍTULO 2: MATERIALES Y MÉTODOS PARA LA ELABORACIÓN DEL CALENDARIO TÉRMICO

En este capítulo se realiza una caracterización del área seleccionada para el estudio. Se procede para la elaboración del calendario térmico a partir de la metodología propuesta por Soler & Martín (2002) para la confección de calendarios climáticos, así como el empleo de otras técnicas estadísticas para la confección del mismo. También se detallan los procedimientos realizados con el ajuste y manejo de las salidas del Modelo de Circulación Regional PRECIS CARIBE, así como los métodos para la verificación del modelo.

2.1 CARACTERIZACIÓN DE LA ZONA ESCOGIDA

El cultivo de la piña fue el seleccionado para esta investigación, pues ha sido una siembra tradicional e importante en la economía de la provincia Ciego de Ávila, utilizándose como fruta fresca para exportaciones y para la industria. La Empresa Agroindustrial Ceballos, encargada del programa de recuperación del cultivo, está realizando estrategias para recuperar los niveles productivos, los cuales descendieron hace años debido, entre muchas causas, al deterioro de sus plantaciones por la escasez de insumos.

El lugar escogido, donde se concentra el programa de desarrollo del cultivo, pertenece a la Unidad Empresarial de Base (UEB) Producción de Piña de la Empresa Agroindustrial Ceballos, conocido como Raúl Martínez, perteneciente al Consejo Popular Alfredo Gutiérrez Lugones, al sur del municipio Ciego de Ávila, con un total de 2000 hectáreas. Está ubicado a los 21.65 ° de latitud Norte (° N) y a los 78.85 ° de longitud Oeste (° W), a 26.5 metros de altura sobre el nivel del mar (Figura 2.1).

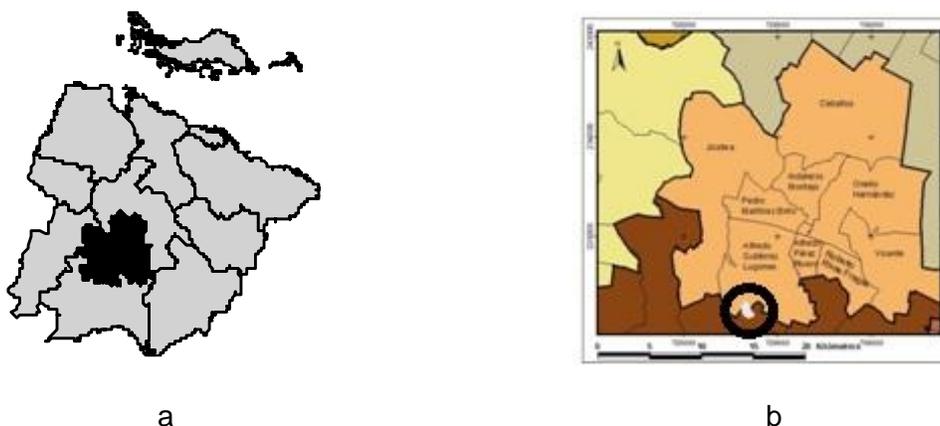


Figura 2.1. a) Provincia de Ciego de Ávila (municipio Ciego de Ávila sombreado). b) Distribución de los Consejos Populares del municipio Ciego de Ávila y localización de la UEB Producción de Piña.

Fuente: CITMA (2011a).

Esta zona la conforma una extensa llanura, con pendiente máxima orientada de Norte a Sur y los suelos predominantes son del tipo ferralíticos rojos, en los subtipos compactados e hidratados utilizados preferentemente para los cultivos varios. Es un área vulnerable a inundaciones por intensas lluvias, que pueden estar asociadas a fenómenos meteorológicos extremos, tanto de tipo Sinóptico (fundamentalmente organismos ciclónicos y Ondas Tropicales), como de Mesoescala (Tormentas Locales Severas).

Durante el período lluvioso (mayo-octubre, también considerado cálido) puede ser afectada por Ondas y Ciclones Tropicales, mientras que en el período poco lluvioso (noviembre-abril, invernal) ocurren afectaciones por Sures y Sistemas Frontales; la severidad de estos sistemas meteorológicos y su nivel de influencia son los responsables de las afectaciones por fuertes vientos, considerada en esa zona hasta el momento, de altamente potencial.

En la UEB Producción de Piña (Figura 2.2) es donde se ejecuta el programa de desarrollo de 2000 hectáreas (ha) de Piña de la variedad MD-2, lo que equivaldría a contar con 2660 ha en áreas físicas de piña al culminar la inversión en el año 2020. Hasta el momento se cuenta con 70 ha, 38 % del área prevista para el cierre de 2015. En la unidad laboran 130 trabajadores, componiéndose por 1 dirigente, 1 administrativo, 14 técnicos, 2 en servicio y 112 operarios.

A pesar de presentar problemas que han afectado la marcha de la inversión, se ha logrado una producción con rendimientos superiores a lo planificado, un incremento de las exportaciones y las ventas al turismo, calidad de la fruta, cuidado al medio ambiente, así como sentido de pertenencia de sus trabajadores, logrando resultados económicos satisfactorios e incremento de la productividad del trabajo.

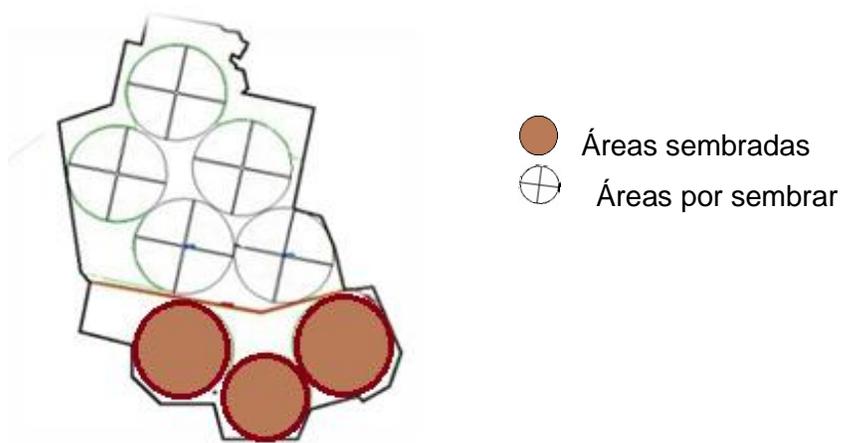


Figura 2.2. Distribución por áreas de la UEB Producción de Piña. Fuente: Elaboración propia.

Para la realización de la caracterización se tuvo en cuenta la información referida por:

- Estrategia Ambiental Provincial 2011-2015 (CITMA, 2011a).
- Estudios de Peligro, Vulnerabilidad y Riesgos de los municipios Ciego de Ávila (CITMA, 2011b) y Venezuela (CITMA, 2011c).
- Estrategia de Ciencia, Tecnología, Innovación y Medio Ambiente del municipio Venezuela (CITMA, 2013).
- Pronóstico climático mensual de la temperatura en Ciego de Ávila (Sorí *et al.*, 2014).
- Intercambio con los directivos de la UEB Producción de Piña de la Empresa Agroindustrial Ceballos (Anexo 1).

Condiciones térmicas

Para la caracterización térmica de la zona se utilizaron los datos de temperatura mínima, media y máxima de la estación 78346. Esta estación, perteneciente del Servicio Meteorológico Nacional, se encuentra en el municipio de Venezuela al sur del municipio cabecera de la provincia Ciego de Ávila, ubicada a los 21.78 ° N y a los 78.78 ° W. Los datos climáticos corresponden a un período base de 30 años (1981-2010), como establece la Organización Meteorológica Mundial, utilizadas en total 32714 observaciones, obtenidas de los archivos del Centro Meteorológico Provincial (Anexo 2).

Las temperaturas ascienden desde enero hasta julio y agosto, meses del período cálido del país (que se extiende desde mayo a octubre) cuando comienzan a descender hasta el mes de diciembre en correspondencia con el invierno del Hemisferio Norte. La temperatura media anual (Figura 2.3) es de 25.0 °C, 0.5 °C por debajo del promedio histórico.

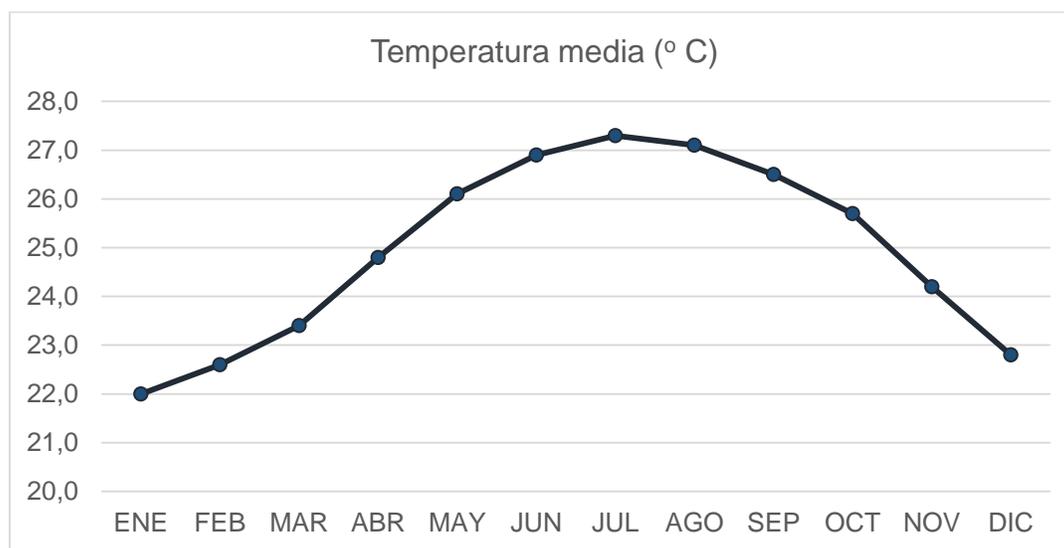


Figura. 2.3. Ciclo anual de la temperatura media.

La temperatura mínima (Figura 2.4) desde enero a abril tiene poca variación, 2.1 °C. Con la transición al período cálido esta aumenta de forma gradual, manteniéndose muy estable en los meses de junio a septiembre, por encima de 22.5 °C. Ya en octubre comienza a declinar, con el inicio del período invernal en noviembre.

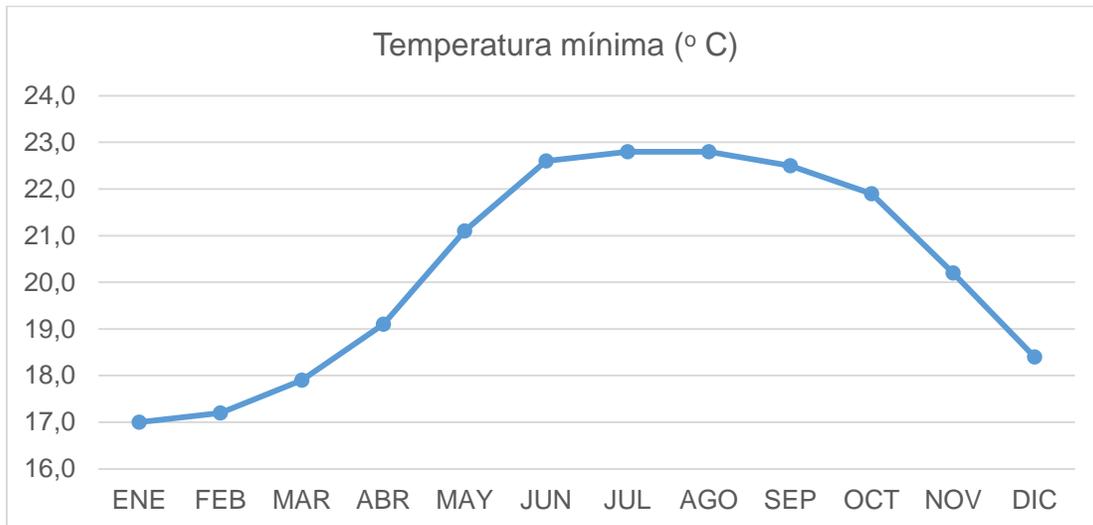


Figura. 2.4. Ciclo anual de la temperatura mínima.

La temperatura máxima (Figura 2.5) está todo el año por encima de 28 °C, llegando hasta 33.5 °C en el mes de julio. Se observa como en las anteriores, bien definidos los períodos especiales (cálido y frío), de mayo a octubre. La diferencia entre meses es muy marcada, superior a 0.5 °C, excepto de mayo a junio (+ 0.2 °C), por la posible ocurrencia de precipitaciones en junio y alta humedad relativa, así como de julio a agosto (- 0.2 °C), meses veraniegos en los cuales hay poca variación de la variable.

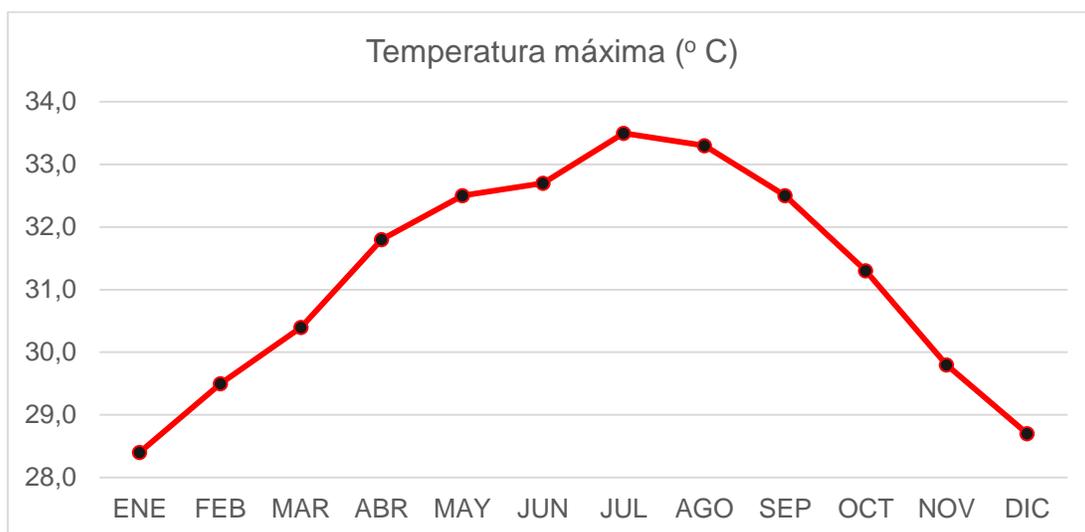


Figura. 2.5. Ciclo anual de la temperatura máxima.

En los 30 años de observación, en la temperatura mínima (Figura 2.6) se observa, exceptuando los años 1981 y 2002 por ser extremos, una irregularidad entre los valores pero que no excede de los 1.5 °C. Desde al año 2004 el rango de variación se redujo estando prácticamente idéntico, y aunque los valores comienzan a descender, producto quizás de condiciones sinópticas, eso no es lo que indica la tendencia. La variable seguirá teniendo un comportamiento quizás irregular entre los años, pero con una distinción a mantener la estacionalidad.

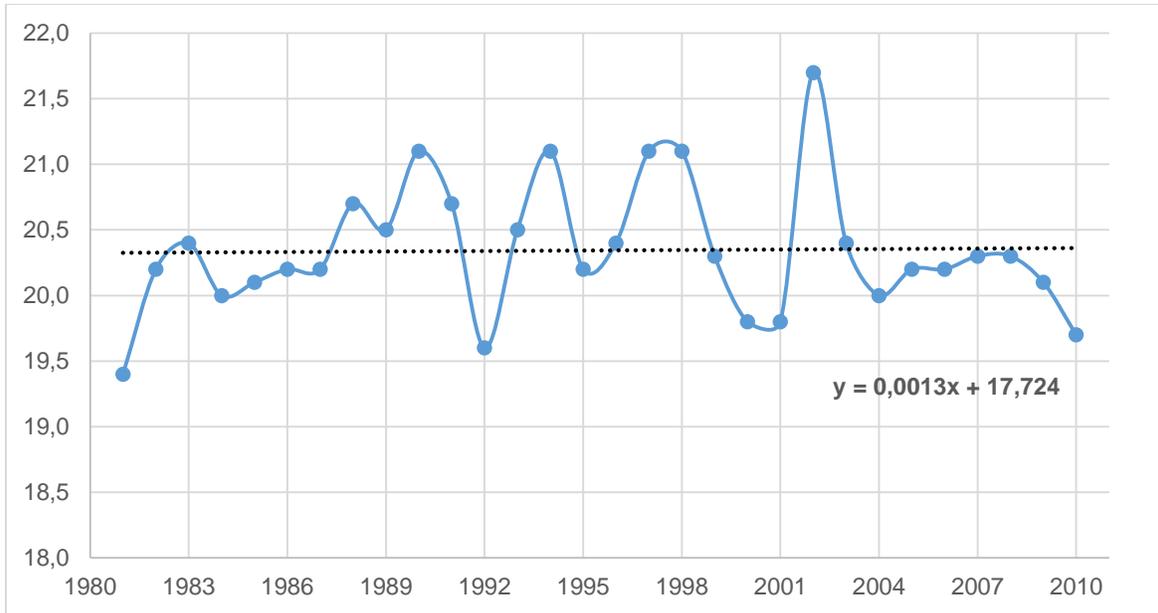


Figura 2.6. Tendencia de la temperatura mínima.

Aunque parezca en la Figura 2.7 que existe mucha variación entre los valores de temperatura máxima entre un año y otro, la variación es menor a 1.2 °C. Como también se apreció en la temperatura mínima, desde el 2004 existe poca variación de la variable, manteniéndose la máxima en este período en un rango de 25 °C.

El 2010 fue un año en el que las condiciones sinópticas hicieron que los valores de la variable estuvieran por debajo del promedio histórico. Según los Boletines Climáticos de ese año (CECLIM, 2016) las condiciones oceánicas y atmosféricas se correspondieron con un evento “El Niño – Oscilación del Sur” (ENOS) moderado, por lo que las temperaturas, sobre todo la extrema, estuvieron por debajo de lo normal en la mayor parte del país. La tendencia indica que se va a conservar la estacionalidad en la temperatura máxima en años siguientes, con un ínfimo descenso, manteniéndose cercana a los 24.9 °C.

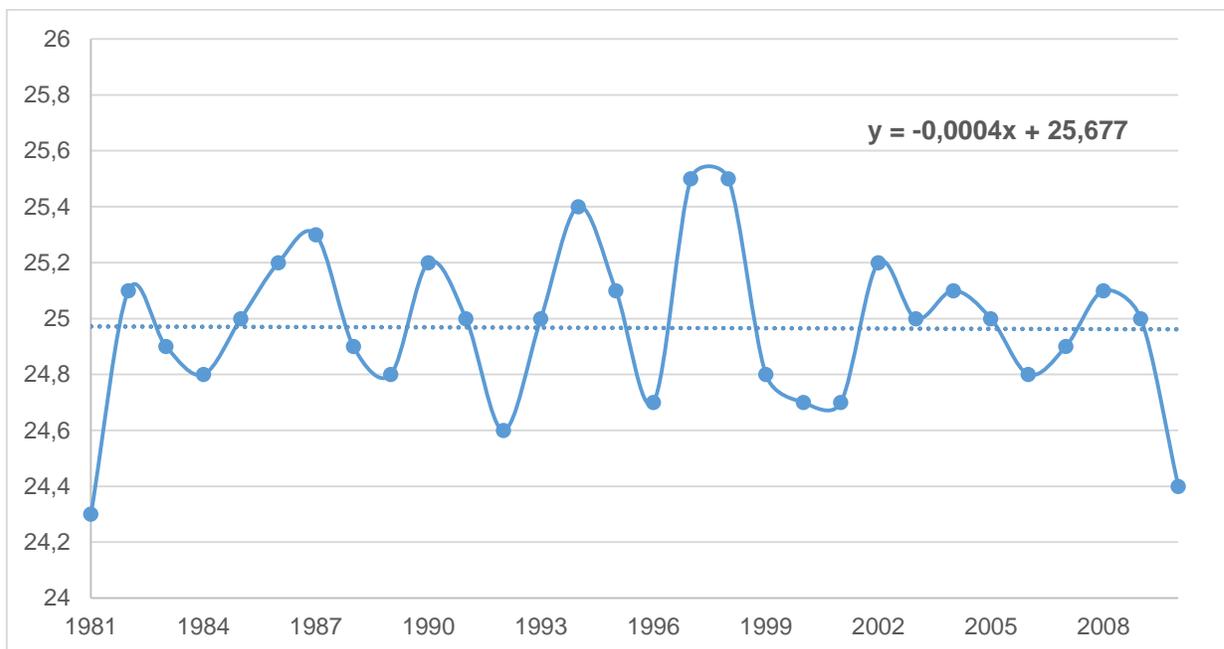


Figura 2.7. Tendencia de la temperatura máxima.

2.2 OBTENCIÓN DE LOS DATOS. CONTEXTUALIZACIÓN DE LA METODOLOGÍA PROPUESTA POR SOLER Y MARTÍN

La base de datos utilizada fue de la estación meteorológica 78346, del municipio Venezuela, disponiéndose de 30 años de registros diarios, de 1981 a 2010, según lo establecido por la Organización Meteorológica Mundial. Después de seleccionar los datos a utilizar, se comprobó la calidad de los mismos al verificar la existencia de estacionalidad, valores lógicos y datos faltantes. De las 10956 observaciones que debían existir para cada temperatura en el período, en total 32868, sólo se trabajó con 32714, por faltante de datos. Los tres calendarios propuestos (para temperatura mínima, media y máxima) se analizaron de la misma forma, como se explica a continuación:

A partir de la metodología brindada por Soler & Martín (2002) para calendarios climáticos, se realizó el calendario térmico, utilizando como técnica estadística la media móvil. Se elaboró un calendario para cada valor de temperatura, es decir, uno para temperatura media, uno para la temperatura mínima y otro para la temperatura máxima. Se calculó para cada fecha del año el promedio de los 30 valores observados en ella, utilizando Microsoft Excel 2013.

Se hallaron medias móviles centradas de cada día con diferentes intervalos, de 3 en 3, de 5 en 5, de 7 en 7, de 9 en 9 y de 11 en 11. Las medias móviles de las primeras y de las

últimas fechas del calendario se calcularon suponiendo que este es cíclico, de modo que el valor que le siguió al del 31 de diciembre fue el del 1 de enero. En el caso singular del 29 de febrero, solo existían 7 observaciones en este período, por lo que se decidió eliminarlo del calendario y no contar con este día.

Para observar las diferentes suavizaciones que se produjeron en las medias móviles, se empleó el software estadístico R (Versión 3.0.0), pues los gráficos disponibles en este programa son de gran calidad y de una versatilidad impresionante. R es un software libre, flexible y dinámico, con una amplia variedad de técnicas estadísticas que se incorporan rápidamente y hacen posible su aplicación (Mirabal *et al.*, 2010). Como software estadístico, ofrece una mayor gama de herramientas y análisis, con gran fiabilidad y enorme potencial para mejorar y crear nuevos módulos a medida que avanzan los conocimientos estadísticos. R, según la bibliografía consultada, es uno de los mejores ejemplos de las ventajas del software libre.

La representación gráfica se hizo de forma anual, pero para una mejor visualización de las irregularidades se efectuó un análisis gráfico de forma bimestral, sin alterar los períodos especiales, seleccionando para esto los meses enero-febrero, marzo-abril, mayo-junio, julio-agosto, septiembre-octubre y noviembre-diciembre. El procedimiento de hallar valores de probabilidad a partir de los cuartiles no se realizó, pues no es objetivo del trabajo encontrar una clasificación relativa de los períodos ni del lugar, sino simplemente elegir la mejor media móvil a partir de su suavizado.

2.3 EMPLEO DE OTRAS TÉCNICAS ESTADÍSTICAS. SELECCIÓN DE LA MÁS EFECTIVA

Según la bibliografía consultada, trabajar con las medidas de tendencia central, hace que los análisis de los datos obtenidos permitan la descripción, el resumen y la presentación de estos de una manera fácil de interpretarlos. Es por eso que se escogió la media aritmética, la mediana y la moda como técnicas para la confección de un calendario climático, así como la media tijereada o recortada por las propiedades que posee.

Igual que el procedimiento de media móvil, trabajando en Microsoft Excel 2013, en el período de 30 años, se calcularon la media aritmética, la mediana, la moda y la media tijereada para cada valor diario, en cada una de las temperaturas. Para saber cuál de las técnicas arrojaba un mejor resultado, se calculó el porcentaje de veces en 5 años que cada una se encontraba en un rango de ± 2 con respecto a la observación real.

Los 5 años escogidos fueron desde el 1ro de enero de 2011 hasta el 31 de diciembre de 2015, y se seleccionó el rango de ± 2 por ser el valor estimado por el Sistema Meteorológico Nacional para augurar un pronóstico acertado. En este período se utilizaron de la estación meteorológica 78346, correspondientes a los 5 años, 1746 datos con cada valor de temperatura, para un total de 5292 datos; en diciembre 2014 y marzo de 2015 no se encontraron observaciones disponibles en la base de datos.

Al comparar el período 2011-2015, no se tiene en cuenta que existe una variación climática en la variable analizada, por lo que puede proporcionar un resultado que no sea el esperado, o simplemente, que esté lejos de lo que realmente el método puede ofrecer en condiciones normales (Tabla 2.1). Es por eso que se decide comparar los métodos con meses que hayan tenido un comportamiento normal de la variable, según la anomalía estandarizada de la temperatura, que es el valor que resulta de la razón entre la anomalía y la desviación estándar, referidas a una misma norma (en este caso 1971-2000), para cada estación. Para esto se recurrió a los Boletines Climáticos Mensuales elaborados por el Centro del Clima del Instituto de Meteorología (CECLIM, 2016) desde el enero 2011 hasta junio de 2016.

Tabla 2.1. Anomalía estandarizada de la temperatura (1971-2000). Fuente: CECLIM (2016).

Valor de anomalía	Categoría
Mayor de 2.0 °C	En extremo por encima de la norma
1.5 a 2.0 °C	Muy por encima de la norma
1.0 a 1.5 °C	Por encima de la norma
0.5 a 1.0 °C	Ligeramente por encima de norma
-0.5 a 0.5 °C	Normal
-0.5 a -1.0 °C	Ligeramente por debajo de la norma
-1.0 a -1.5 °C	Por debajo de la norma
-1.5 a -2.0 °C	Muy por debajo de la norma
Menor de -2.0	En extremo por debajo de la norma

Fueron seleccionados los meses de distintos años según cumplieran con la condición de mostrar un comportamiento climático normal. A lo largo del período, un mismo mes de distinto año repetía ese comportamiento dadas las condiciones sinópticas imperantes del momento, pero otros no, por lo que fue seleccionado un solo mes para el análisis, el de mejor comportamiento, representados en la Tabla 2.2.

Tabla 2.2. Meses con un comportamiento normal de la variable temperatura en la zona de estudio.

Meses	Años
Enero	2012
Febrero	2011
Marzo	2012
Abril	2012
Mayo	2011
Junio	2011
Julio	2011
Agosto	2011
Septiembre	2011
Octubre	2012
Noviembre	2014
Diciembre	2011

Se compararon los calendarios (de las tres temperaturas) de los distintos métodos con 365 datos (un año). Para evaluar la efectividad de cada método, se consideró que estos estuvieran en un rango de ± 2 °C en comparación con la observación real. Al ser resultados muy similares, pues se acortó el rango, y se comparó entonces con ± 1 °C.

2.4 TRABAJO CON EL MODELO DE CIRCULACIÓN REGIONAL PRECIS CARIBE

Se utilizaron datos diarios del Modelo de Circulación Regional PRECIS Caribe, con salidas del MCG Echam4, bajo condiciones de Escenarios de Cambio Climático B2. Se escogieron estas salidas pues el desarrollo esperado en el territorio en los próximos años no comprende una evolución industrial con elevadas emisiones de gases de efecto

invernadero, ni un desarrollo económico superior con un crecimiento poblacional, como para que también se hubiera analizado otras salidas, como el A2.

2.4.1 TRATAMIENTO DE LOS DATOS EN EL PERÍODO BASE

Para la determinación de los puntos cercanos a la zona de estudio y que correspondieran a las salidas del modelo se empleó el Nuevo Atlas Nacional de Cuba (Instituto de Geografía, 1989). Se determinó un punto con las coordenadas geográficas 21.75 ° N y 78.75 ° W (punto medio de los cuatro escogidos del modelo (21.5 ° N - 78.5 ° W, 22 ° N - 78.5 ° W, 21.5 ° N - 79 ° W, 22 ° N - 79 ° W)), y que se ubica a 14 km de la zona de estudio.

El tratamiento de los datos del PRECIS Caribe fue realizado en el programa Microsoft Excel 2013 y en el paquete estadístico STATGRAPHICS.Centurion.v15.1.0.2. Primeramente, se hizo la verificación de la exactitud del modelo (Figura 2.4)¹, que según León (2005) se puede idealizar como una comparación simple del pronóstico del modelo y la verdad atmosférica:

$$\textit{Exactitud} = \textit{Pronóstico del Modelo} - \textit{Atmósfera real}$$

Se utilizaron los años del período base de 1991 a 2010, pues en años anteriores había faltante de datos, aunque en este período también existió un mes sin datos, enero de 1998.

Se realizó el procedimiento de Comparación de Varias Muestras para comprobar la aproximación de los resultados del modelo utilizado y los datos obtenidos por la estación meteorológica, verificando la existencia de diferencias estadísticamente significativas entre las muestras. Este procedimiento está diseñado para comparar dos o más muestras independientes de datos variables, se hacen pruebas para determinar si hay o no diferencias significativas entre las medias, varianzas y/o medianas de las poblaciones de las que se tomaron las muestras.

En el análisis de cada temperatura se trazó la misma hipótesis nula

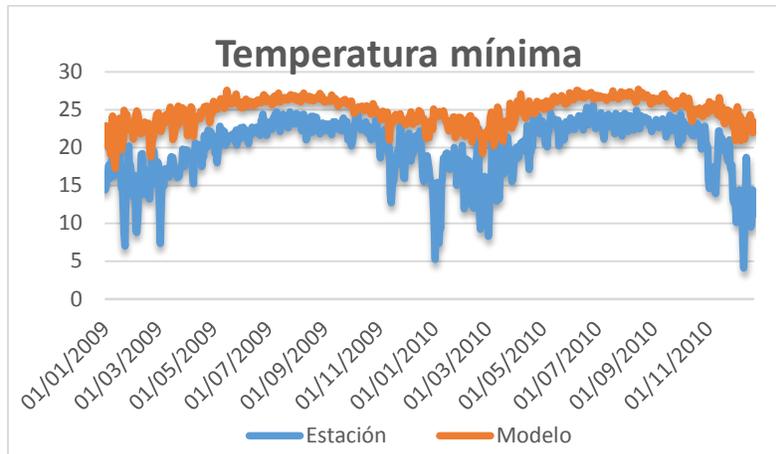
$$H_0: \mu_1 = \mu_2 = \dots = \mu_n \text{ (todas las medias iguales)}$$

contra la hipótesis alternativa

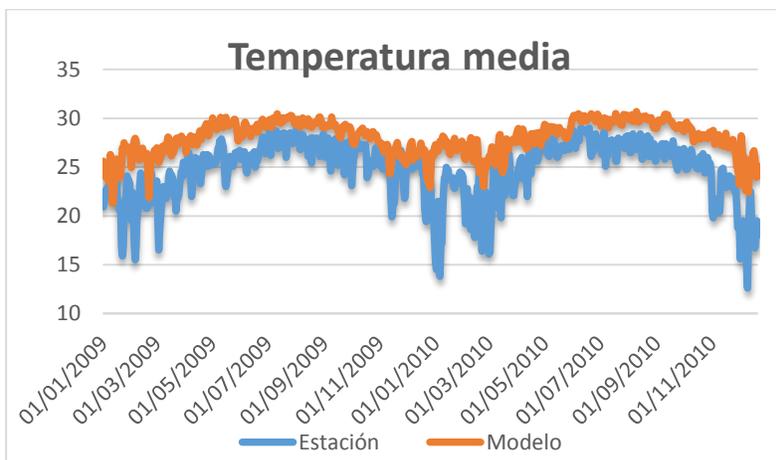
$$H_1: \text{no todos los } \mu_j \text{ son iguales,}$$

utilizando para la determinación de la hipótesis el programa STATGRAPHICS.

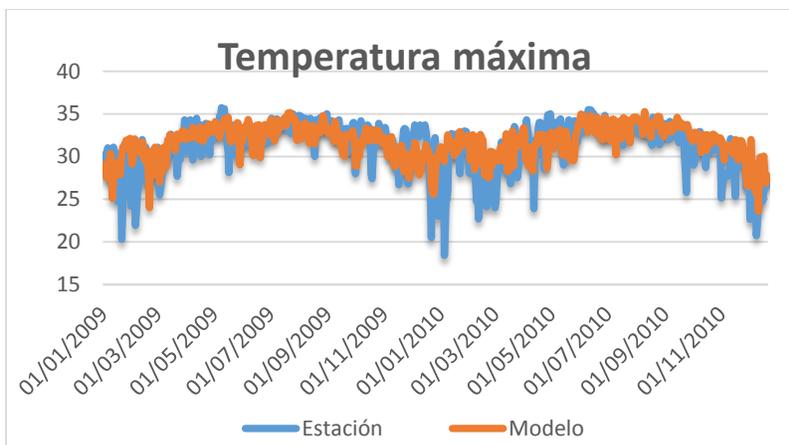
¹ Se decidió escoger pocos años para que se definiera lo mejor posible la diferencia entre los datos del modelo y la estación, fue el período 2009-2010 pues son en los que se termina el período base.



a



b



c

Figura 2.4. Comparación entre el modelo PRECIS CARIBE y los datos de la estación meteorológica 78346. Escenario de Cambio Climático B2. a) Temperatura mínima b) Temperatura media c) Temperatura máxima.

Para una primera comparación de las muestras se utilizó el gráfico de dispersión. Para saber si los datos que se contrastan provienen de distribuciones normales, se acudió al análisis del sesgo estandarizado y a la curtosis estandarizada, tomando el criterio que se deben encontrar en el rango de - 2 a + 2.

La varianza de los datos, manifestada a través de la tabla ANOVA, se descompuso en dos componentes: un componente *entre grupos* y otro *dentro de grupos*. El cociente entre estos componentes se determinó mediante la razón-F para probar la existencia de una diferencia estadísticamente significativa entre las medias de las dos muestras.

Se realizó las Pruebas de Múltiples Rangos para determinar cuáles medias eran significativamente diferentes de otras, empleando el procedimiento de diferencia mínima significativa (LSD) de Fisher, con un intervalo de confianza para cada par de medias al nivel de confianza seleccionado (95%) usando:

$$M = t_{\alpha/2, n-q}$$

donde t representa el valor de la distribución t de Student con $n - q$ grados de libertad dejando un área de $\alpha/2$ en la cola superior de la curva. Este procedimiento es llamado el procedimiento de la Diferencia Menos Significativa, ya que la magnitud de los límites indica la diferencia más pequeña entre dos medias cualesquiera que puedan ser declaradas para representar una diferencia estadísticamente significativa. Con este método hay un riesgo del 5 % al decir que cada par de medias es significativamente diferente, cuando la diferencia real es igual a 0.

En el caso de que no se cumplieran las suposiciones del análisis de la varianza, se aplicó la prueba de Kruskal-Wallis para el contraste de k medianas, trazándose como hipótesis nula:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_k$$

de que las medianas entro de cada una de las columnas es la misma. Este método lo primero que hace es combinar los datos de todas las columnas y ordenarlos de menor a mayor. Después, calcula el rango promedio para los datos de cada columna.

Al existir una diferencia estadísticamente significativa entre las medianas, se seleccionó el Gráfico de Caja y Bigotes para determinar cuáles medianas eran significativamente diferentes de otras. Este se construyó de la siguiente manera:

- Se dibujó una caja que va del *cuartil inferior* de la muestra hasta el *cuartil superior*. Este es el intervalo cubierto por 50 % de los datos al ser ordenados de menor a mayor.
- Se trazó una línea vertical en la mediana (valor medio).
- Al requerirse, se colocó un signo de más en el lugar de la media muestral.
- Se esbozaron bigotes desde los bordes de la caja hasta los valores más pequeño y más alto. Los puntos exteriores, que son puntos abajo o arriba de la caja más de 1.5 veces el rango intercuartil, se indicaron con símbolos de punto. Los puntos muy alejados son aquellos que quedaron a más de 3.0 veces el rango intercuartílico por arriba o por abajo de la caja, y se mostraron como pequeños cuadrados con un signo más en su interior.

2.4.2 EXACTITUD DE LOS DATOS

Para comparar la exactitud entre los datos del modelo y los corregidos con respecto a los de la estación, se tuvieron en cuenta los aspectos descritos por León (2005) para la verificación de los modelos meteorológicos y se realizaron los cálculos del error cuadrático medio (*mean square error* - MSE), del error absoluto medio (*mean absolute error* - MAE) y del error del sesgo (BE) para el período poco lluvioso (cálido) y el poco lluvioso (invernal) y así respetar la estacionalidad y valorar mejor los resultados.

En cuanto al error cuadrático medio (MSE) y error absoluto medio (MAE), son dos medidas comunes de exactitud que miden la distancia "promedio" entre las predicciones y las correspondientes observaciones y ambas tienen valor cero para pronósticos perfectos.

El error cuadrático medio (error total) es la raíz cuadrada del promedio del cuadrado de las diferencias ajustadas individuales entre el pronóstico (f_n) y la observación (o_n), donde N es el número total de comparaciones pronosticadas:

$$RMSE = \left[\frac{1}{N} \sum_{n=1}^N (f_n - o_n)^2 \right]^{\frac{1}{2}}$$

El cuadrado de esta diferencia fuerza el RMSE para tener en cuenta tanto los errores positivos y negativos, así el RMSE hace la medida del error total del modelo.

El error absoluto medio (magnitud del error total) es el promedio del valor absoluto de la diferencia entre el pronóstico (f_n) y la observación (o_n), donde N es el número total de comparaciones pronosticadas. Tomar el valor absoluto de esta diferencia fuerza a MAE a

tener en cuenta, tanto errores positivos como negativos, así también hace la medida del error total del modelo:

$$MAE = \frac{1}{N} \sum_{n=1}^N |f_n - o_n|$$

La utilización del MAE también fomenta los pronósticos conservativos, pero este problema es menor que para el caso del MSE, que es particularmente sensible a los grandes errores. Sin embargo, para muestras grandes, en las que tanto los pronósticos como las observaciones tienen distribuciones simétricas (la media y la mediana son similares), el MAE y el RMSE deberían ser similares.

El error del sesgo (error sistemático) es una medida de la fiabilidad del modelo. Para un conjunto de N pronósticos de este tipo, la medida de fiabilidad es el error medio, definido como el promedio de la diferencia entre el pronóstico (f_n) y la observación (o_n), donde N es el número total de pronósticos comparados:

$$BE = \frac{1}{N} \sum_{n=1}^N (f_n - o_n)$$

Debido a que BE carece del término en valor absoluto o al cuadrado, se permite la cancelación de errores individuales positivos y negativos, poniendo en evidencia solamente el exceso, es decir el error sistemático.

BE representa la desviación promedio entre los pronósticos y las observaciones. Si BE es positivo, los valores pronosticados tienden a superar a los observados, mientras que si el sesgo es negativo sucede lo contrario.

2.4.3 AJUSTE DEL MODELO

Después de comprobar la existencia de diferencias estadísticamente significativas entre los datos, se decide realizarle al modelo un ajuste para que tenga un comportamiento similar al de la zona y que los resultados que se obtengan a partir de este sean acordes al territorio en estudio.

Para ajustar los valores del modelo con los reales de la estación en el período base, se les realizó a las variables una corrección de forma bimestral. Esta se efectuó determinando el promedio de la diferencia entre los valores en el bimestre. En cuanto a los coeficientes de la temperatura mínima (Tabla 2.3) y de la temperatura media (Tabla 2.4) fueron obtenidos

tomando los valores de la variable (en °C) del modelo para los años de interés. El coeficiente fue restado a los valores de temperatura del modelo.

Tabla 2.3. Coeficientes bimestrales de la temperatura mínima

Bimestre	Coeficiente
enero-febrero	5.3
marzo-abril	5.2
mayo-junio	3.9
julio-agosto	3.5
septiembre-Octubre	3.4
noviembre-diciembre	3.8

Tabla 2.4. Coeficientes bimestrales de la temperatura media

Bimestre	Coeficiente
enero-febrero	3.3
marzo-abril	3.3
mayo-junio	3.0
julio-agosto	3.0
septiembre-Octubre	2.1
noviembre-diciembre	2.1

En las temperaturas en que los datos modelados fueron corregidos, se compararon los nuevos valores con los de la estación para identificar si esta vez existían entonces diferencias estadísticamente significativas entre estos, utilizando el programa STATGRAPHICS la opción de Comparación de Varias Muestras.

En el Capítulo 3 se explica la comparación conjunta entre los datos del modelo, los de la estación y los corregidos para simplificar la explicación del trabajo realizado en el STATGRAPHICS, utilizando las mismas operaciones manifestadas anteriormente.

✓ Temperatura mínima

De las 7200 observaciones que debían existir en el período de 1991-2010, solo habían 7163 por faltante de datos. En la tabla 2.5 se muestra varios estadísticos para cada una de las 3 columnas de datos y los cuales fueron tomados para realizar los análisis pertinentes.

Tabla 2.5. Resumen estadístico para la temperatura mínima.

	Recuento	Promedio	Desviación Estándar	Coefficiente de Variación	Mínimo	Máximo	Rango
Estación	7163	20.3832	3.05747	15.0%	4.1	25.7	21.6
Modelo	7163	24.56	1.75428	7.14285%	16.8	28.8	12.0
Corrección	7163	20.3322	2.56237	12.6025%	4.0	25.4	21.4
Total	21489	21.7585	3.20229	14.7174%	4.0	28.8	24.8

✓ Temperatura media

Debían existir 7200 observaciones en este período, solo se encontraban 7166, pues existen faltante de datos. La tabla 2.6 muestra varios estadísticos para cada una de las 3 columnas de datos y los cuales fueron tomados para realizar los análisis pertinentes.

Tabla 2.6. Resumen estadístico para la temperatura media.

	Recuento	Promedio	Desviación Estándar	Coefficiente de Variación	Mínimo	Máximo	Rango
Estación	7166	25.0259	2.36089	9.43378%	12.6	29.5	16.9
Modelo	7166	27.6857	1.66025	5.99678%	20.0	31.1	11.1
Corrección	7166	24.9674	2.20152	8.81758%	7.5	28.9	21.4
Total	21498	25.893	2.44936	9.45954%	7.5	31.1	23.6

✓ Temperatura máxima

Al igual que en las temperaturas anteriores existen faltante de datos en los registros. De las 7200 observaciones que debían existir, aparecen solo 7168. En la Tabla 2.7 aparecen los estadísticos de las dos columnas tratadas en este caso.

Tabla 2.7. Resumen estadístico para la temperatura media.

	Recuento	Promedio	Desviación Estándar	Coefficiente de Variación	Mínimo	Máximo	Rango
Estación	7168	31.2882	2.48383	7.93855%	18.4	37.5	19.1
Modelo	7168	31.2794	1.90655	6.09521%	21.0	35.3	14.3
Total	14336	31.2838	2.21401	7.07718%	18.4	37.5	19.1

2.4.4 TRATAMIENTO DE LOS DATOS EN EL PERÍODO 2016-2020

Para estimar la posible existencia de desviaciones en la variable y observar los cambios que pueda experimentar el calendario propuesto en cada temperatura durante los próximos años, se escogió trabajar el período de 2016 a 2020, utilizando las salidas diarias del Modelo de Circulación Regional PRECIS Caribe para los años de interés. Al utilizar los datos del modelo, en caso de que en el análisis del período base hubieran presentado diferencias estadísticamente significativas con las observaciones de la estación, les fue aplicado a los bimestres de cada año el factor de corrección hallado en el tratamiento de datos del período base (Tablas 2.2 y 2.3), para que hubiera un comportamiento similar al existente en la zona.

Se comprobó en primera instancia si el calendario propuesto tenía diferencias estadísticamente significativas con los datos del modelo (ya corregidos), utilizando en el programa STATGRAPHICS el procedimiento de Comparación de Varias Muestras. Como el PRECIS Caribe modela solo 360 datos al año, pues también se eligió esa cantidad del calendario térmico propuesto para cada valor de temperatura, eliminando los últimos días del año.

En las Tablas 2.8, 2.9 y 2.10, aparece un resumen estadístico de los datos que fueron tratados. Se realizó varias pruebas estadísticas para comparar las muestras. Con la prueba-F de la tabla ANOVA se determinó la no existencia de diferencias significativas entre las medias.

Tabla 2.8. Datos tratados en el programa STATGRAPHICS para la temperatura mínima.

Muestra		Cantidad de valores	Rango
1	Calendario propuesto	360	16.7 a 23.3
2	PRECIS 2016	360	11.3 a 24.7
3	PRECIS 2017	360	13.9 a 24.4
4	PRECIS 2018	360	13.3 a 23.7
5	PRECIS 2019	360	15.1 a 24.9
6	PRECIS 2020	360	14.2 a 24.3

Tabla 2.9. Datos tratados en el programa STATGRAPHICS para la temperatura media

Muestra		Cantidad de valores	Promedio
1	Calendario propuesto	360	25.2853
2	PRECIS 2016	360	25.1722
3	PRECIS 2017	360	25.2822
4	PRECIS 2018	360	25.28
5	PRECIS 2020	360	25.3169

Tabla 2.10. Datos tratados en el programa STATGRAPHICS para la temperatura máxima

Muestra		Cantidad de valores	Promedio
1	Calendario propuesto	360	31.5028
2	PRECIS 2016	360	31.3042
3	PRECIS 2017	360	31.59
4	PRECIS 2018	360	31.7228
5	PRECIS 2020	360	31.6656

Las posibles diferencias y la efectividad del calendario con los años futuros fueron realizadas en Microsoft Excel 2013. La efectividad se analizó teniendo en cuenta si el calendario térmico propuesto tenía una diferencia de ± 2 °C en comparación con los años 2016 a 2020, tanto para temperatura mínima, como media y máxima.

CONCLUSIONES DEL CAPÍTULO

- Los pasos descritos para la confección de calendarios térmicos a partir de diferentes técnicas, así como en el análisis de los modelos climáticos, constituyen un aporte metodológico que otorgan rigor científico y teórico al resultado.
- Para el tratamiento estadístico de los datos se utilizan distintos programas, a decisión del investigador, que posibilita encontrar sesgos, diferencias estadísticas, mejor visualización gráfica, sin perder el objetivo central de la investigación.

CAPÍTULO 3: DISCUSIÓN DE LOS RESULTADOS

En este capítulo se presentan los resultados relacionados con la aplicación contextualizada de la metodología propuesta por Soler & Martín (2002) para la confección del calendario térmico. Se establecen comparaciones entre meses de comportamiento normal de la variable temperatura para evidenciar la efectividad del calendario propuesto. Además, se comprueba la exactitud del Modelo de Circulación Regional PRECIS CARIBE y se compara el calendario con las salidas del modelo para analizar variaciones de la variable en los próximos 5 años.

3.1 CALENDARIO TÉRMICO SEGÚN LA METODOLOGÍA POR SOLER & MARTÍN (2002)

Como se trabajó con las tres temperaturas (mínima, media y máxima), se realizó un calendario para cada valor, analizándolos de la misma forma.

3.1.1 CALENDARIO DE TEMPERATURA MÍNIMA

Al analizar las suavizaciones que presenta cada intervalo (Figura 3.1), si se realiza la interpretación gráfica de forma anual, dadas las marcadas diferencias entre meses que se exhiben en los distintos períodos (cálido y frío), es poco perceptible las diferencias entre ellos. Por eso es que se efectúa el análisis gráfico de forma bimestral para apreciar la mejor suavización entre los intervalos.

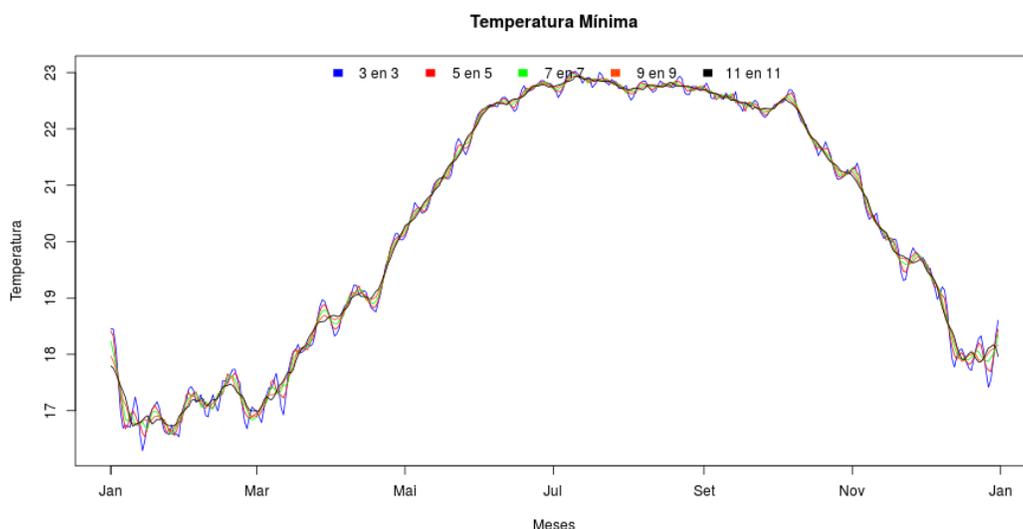


Figura 3.1. Representación anual de temperatura mínima en los distintos intervalos de media móvil.

Al observar la Figura 3.2 se aprecia que el intervalo con más alteraciones es de 3 en 3 dadas las claras variaciones entre un día y otro, sobre todo en el mes de enero. En el

segundo mes las irregularidades son menos, pero el intervalo de 3 en 3 es que presenta más cambios. El mejor suavizado se observa en el intervalo de 11 en 11 en todo el bimestre.

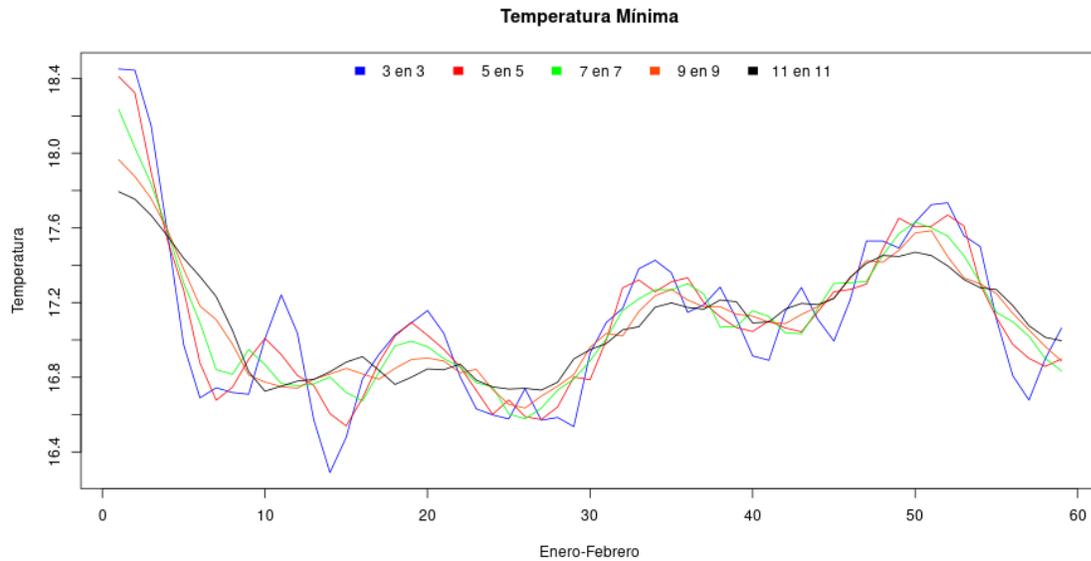


Figura 3.2. Representación de temperatura mínima en los distintos intervalos de media móvil (Bimestre Enero-Febrero)

En el bimestre marzo-abril (Figura 3.3) las mayores variaciones son en la media móvil cada 3 días, sobre todo en el primer mes. Durante este período la media móvil con mejor suavizado es la de 11 días, los demás intervalos en ocasiones presentan pequeños cambios, pero no es como el de 11 que tiene una curva muy suave.

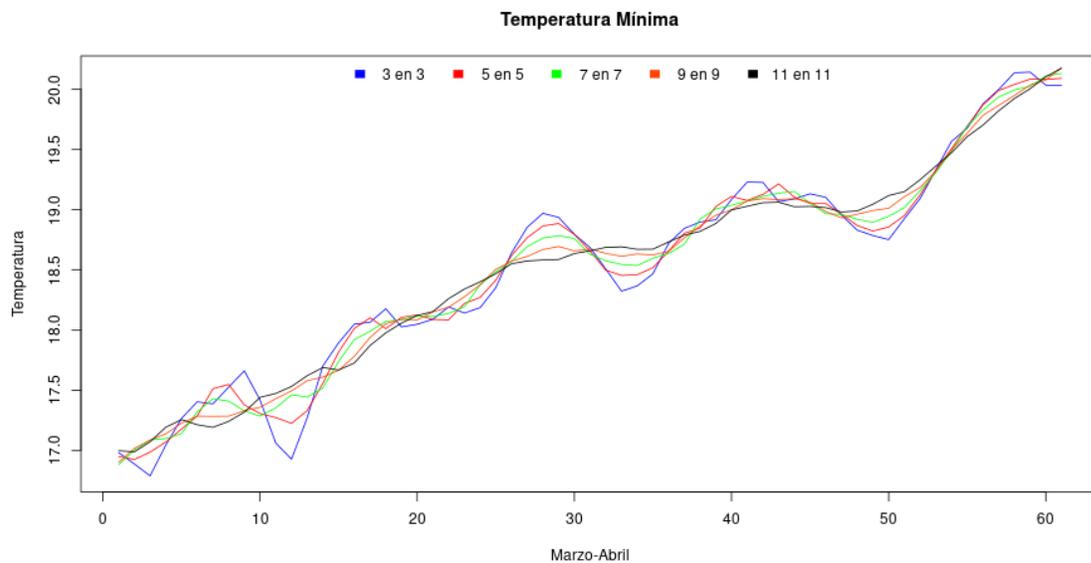


Figura 3.3. Representación de temperatura mínima en los distintos intervalos de media móvil (Bimestre Marzo-Abril)

En la Figura 3.4 se representa el bimestre mayo-junio, con un suavizado más definido que los bimestres anteriores. El intervalo de 3 en 3 es el que presenta más cambios, notándose más en el mes de mayo. Aun cuando los demás intervalos presentan mejores suavizaciones que en ocasiones anteriores, sigue siendo el de 11 días el mejor.

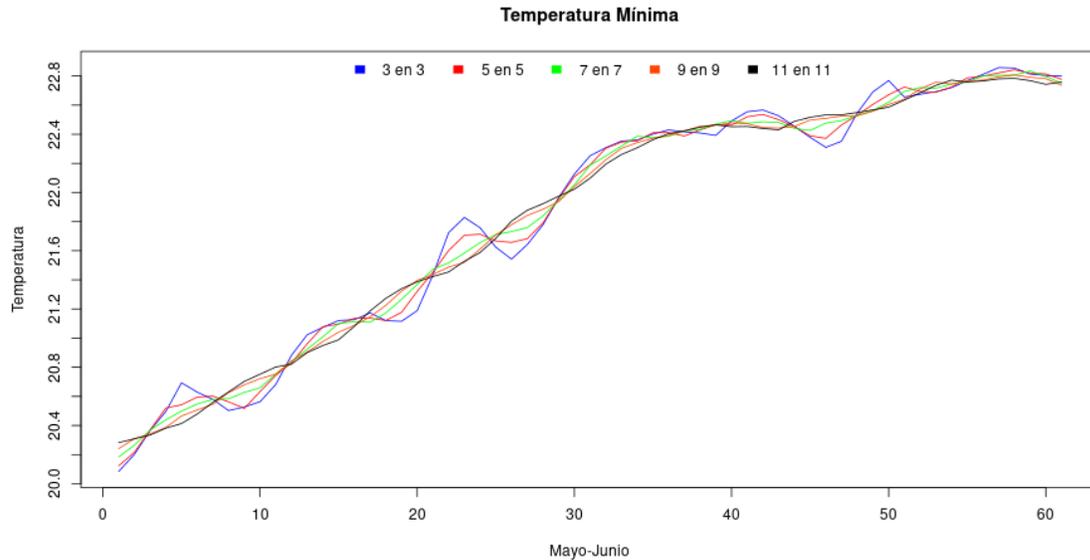


Figura 3.4. Representación de temperatura mínima en los distintos intervalos de media móvil (Bimestre Mayo-Junio)

El bimestre julio-agosto (Figura 3.5), período veraniego, presenta muchas variaciones por los altos valores de la variable en estos meses. A diferencia de la media móvil cada 11 días, las demás apenas presentan un suavizado, dadas las irregularidades imperantes, sobresaliendo el intervalo de 3 en 3 como el de cambios bruscos entre cada fecha.

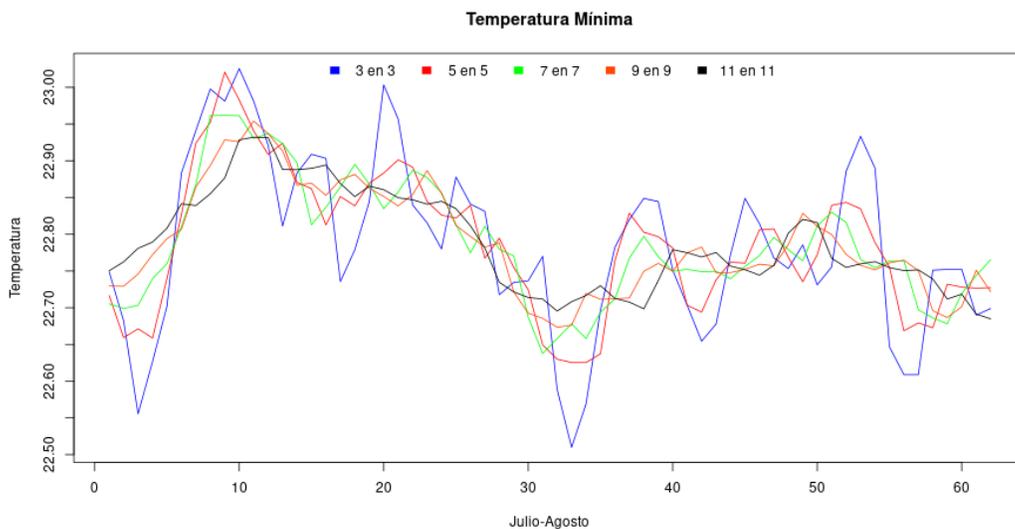


Figura 3.5. Representación de temperatura mínima en los distintos intervalos de media móvil (Bimestre Julio-Agosto)

En el quinto bimestre del año (Figura 3.6) se observa un mejor suavizado en cada uno de los intervalos, pero continúa el de 3 en 3 siendo el de mayor irregularidad. Aun cuando la temperatura comienza a descender en el mes de octubre, el intervalo de 11 en 11 es el de mejor suavizado, con pocas diferencias entre períodos continuos.

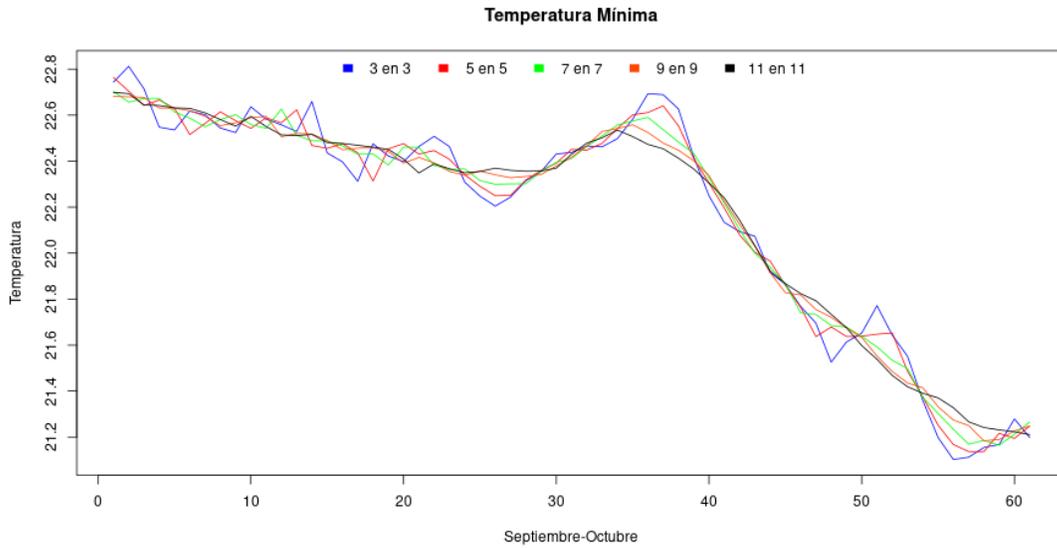


Figura 3.6. Representación de temperatura mínima en los distintos intervalos de media móvil (Bimestre Septiembre-Octubre)

En los dos últimos meses del año la temperatura mínima continúa descendiendo (Figura 3.7), siendo noviembre un mes sin cambios bruscos en cada uno de los intervalos, acentuándose más en los primeros días de diciembre y en el intervalo de 3 en 3. En este bimestre la media móvil de 11 días es la de mejor suavizado con respecto a las demás.

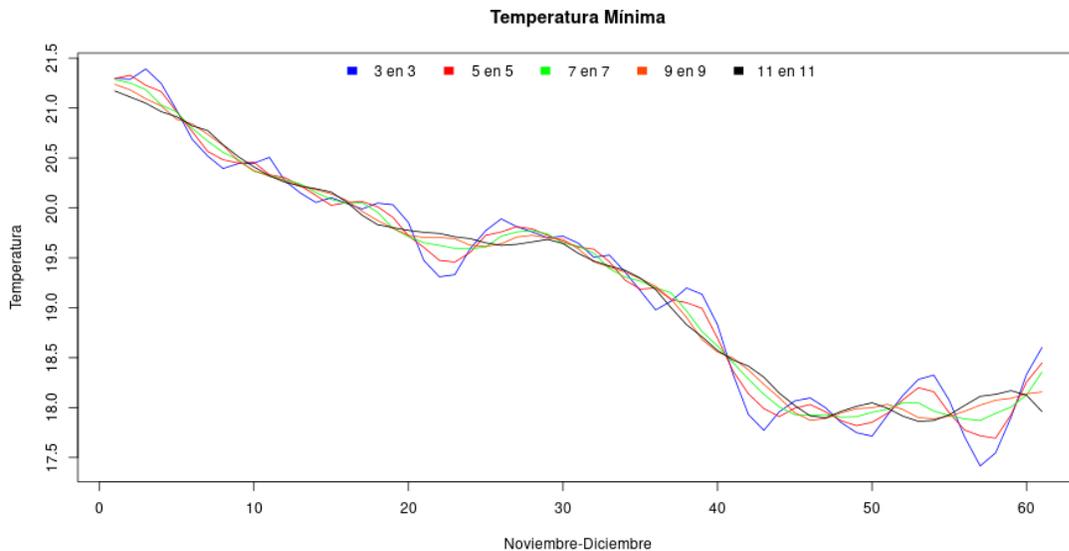


Figura 3.7. Representación de temperatura mínima en los distintos intervalos de media móvil (Bimestre Noviembre-Diciembre)

Al analizar las suavizaciones de forma bimestral se pudo apreciar mejor las diferencias existentes entre cada intervalo. La media móvil cada 11 días fue la que mejor suavización presentó en cada uno de los bimestres analizados, por lo que se escoge este intervalo como el mejor para la confección del calendario de temperatura mínima (Figura 3.8).

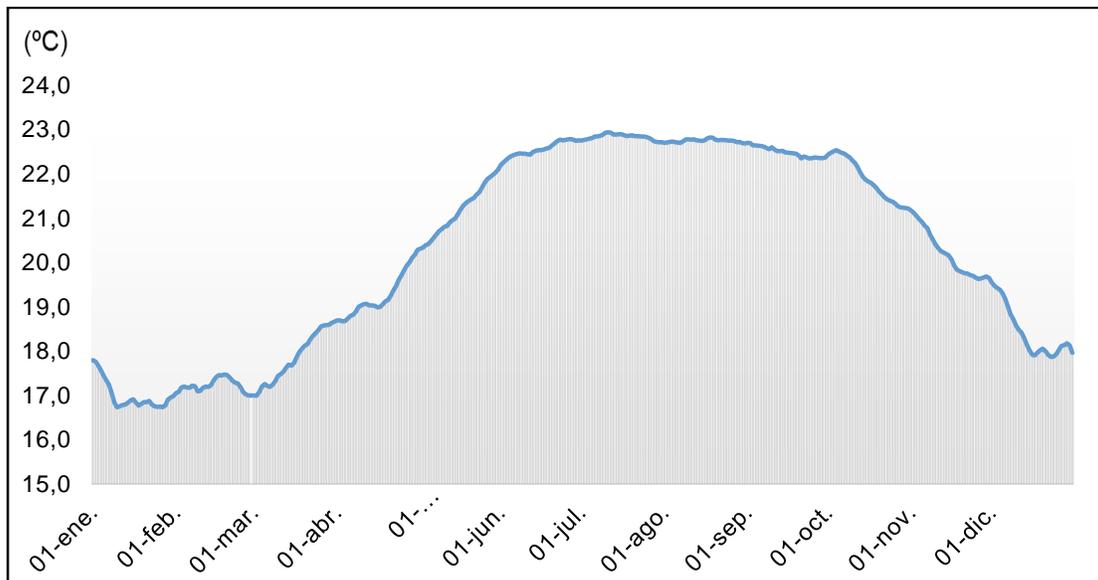


Figura 3.8. Calendario climático de temperatura mínima

3.1.2 CALENDARIO DE TEMPERATURA MEDIA

Si se analiza de forma anual los distintos intervalos de media móvil para la temperatura media (Figura 3.9) son poco apreciable los cambios entre ellos, aunque a simple vista ya se valora que las mayores irregularidades son en la media móvil cada 3 días.

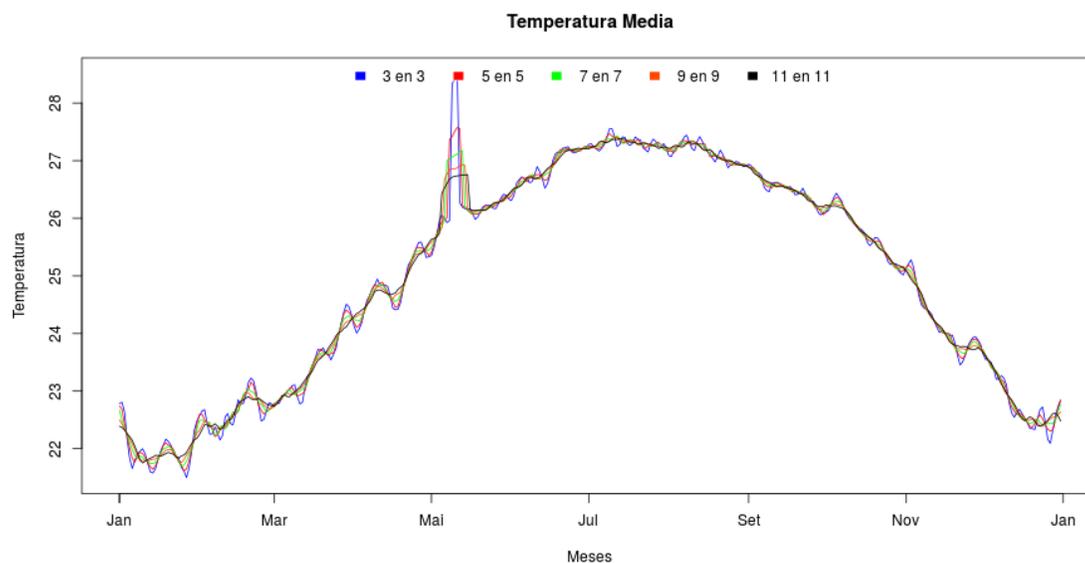


Figura 3.9. Representación anual de temperatura media en los distintos intervalos de media móvil

En el primer bimestre del año se evalúa marcadas diferencias entre los intervalos de media móvil para la temperatura media (Figura 3.10). La media móvil cada 3 días es la que presenta una considerable variación entre días consecutivos; en las demás también se puede aportar el mismo criterio, aunque cada 11 días el suavizado es sobresaliente, siendo esta la mejor media móvil para este período.

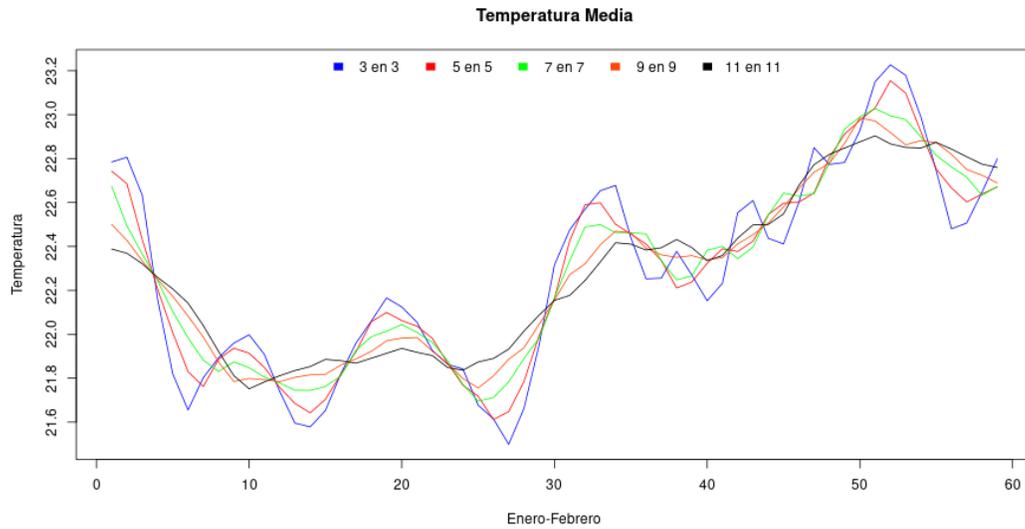


Figura 3.10. Representación de temperatura media en los distintos intervalos de media móvil (Bimestre Enero-Febrero)

En el segundo bimestre del año, mostrado en la Figura 3.11, se observa menos irregularidades con el aumento gradual de la temperatura media, aunque a simple vista se define que la media móvil cada 11 días es la de mejor suavizado, no así con la de 3, que se sobresale por presentar los cambios más notables entre valores continuos.

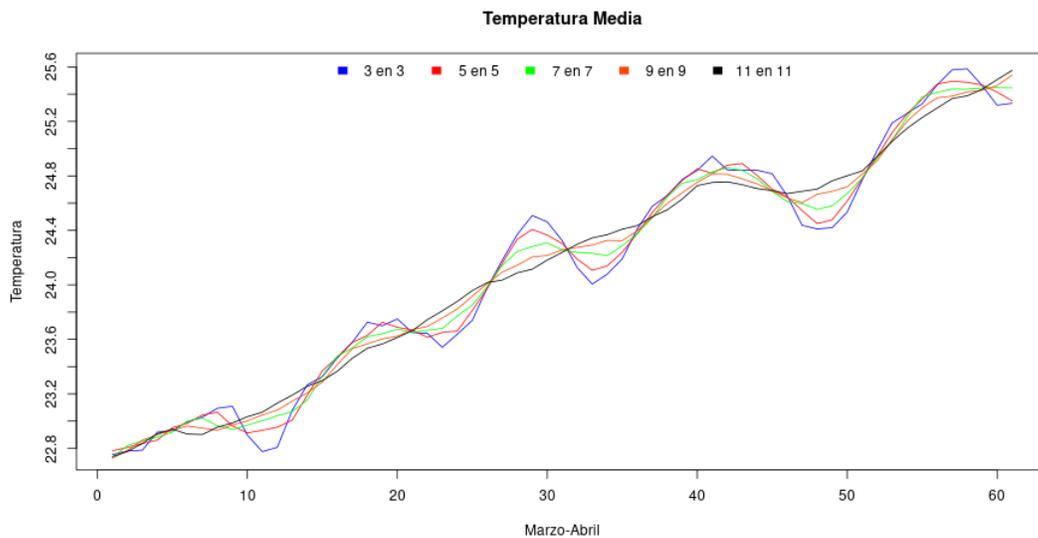


Figura 3.11. Representación de temperatura media en los distintos intervalos de media móvil (Bimestre Marzo-Abril)

A inicios del tercer bimestre del año (Figura 3.12) existe una gran irregularidad, la que se acondiciona a medida que se aumenta el intervalo de la media móvil. Ya después de la segunda mitad del primer mes y durante todo el segundo, los intervalos se muestran muy parejos, pero el de 11 en 11 continúa siendo el de mejor suavizado, así como el de 3 en 3 es todo lo contrario.

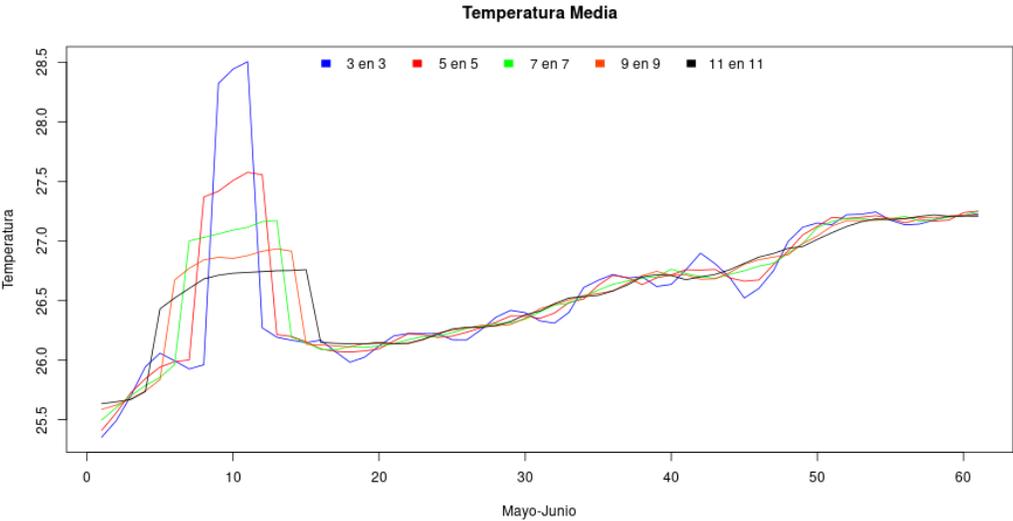


Figura 3.12. Representación de temperatura media en los distintos intervalos de media móvil (Bimestre Mayo-Junio)

Los meses de julio y agosto (Figura 3.13), además de ser los que presentan mayores valores de temperatura, también muestran grandes cambios con respecto a los valores diarios continuos, sobre todo en la media móvil cada 3 días. La que mejor elimina de la serie los componentes estacionales y accidentales es la media móvil de 11 días.

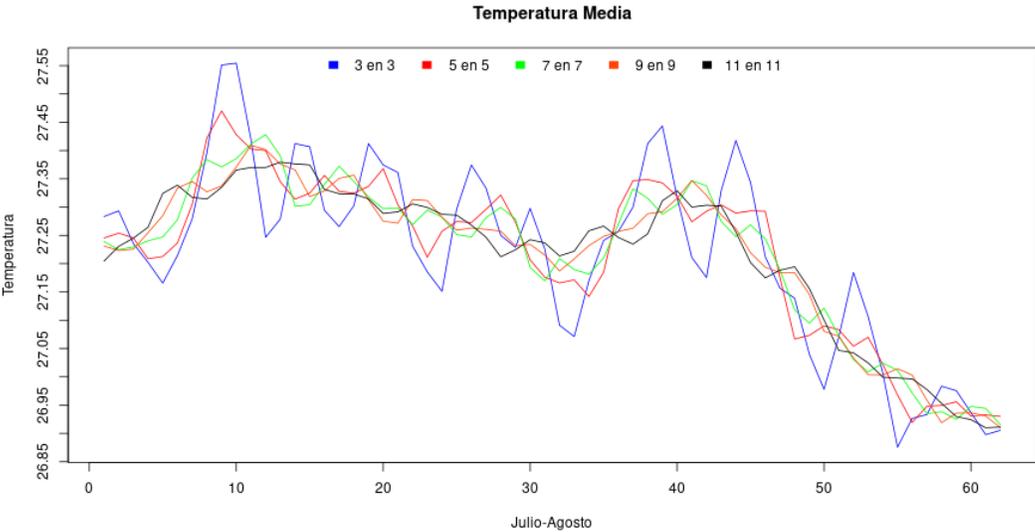


Figura 3.13. Representación de temperatura media en los distintos intervalos de media móvil (Bimestre Julio-Agosto)

En los meses de septiembre y octubre la temperatura media comienza a descender (Figura 3.14). En las curvas se notan escasos cambios acentuados, aunque en el intervalo de 3 en 3 es donde ocurren estas diferencias. El intervalo de 11 en 11, independientemente que sigue la tendencia de otros, es el que presenta mejor suavizado con respecto a los demás.

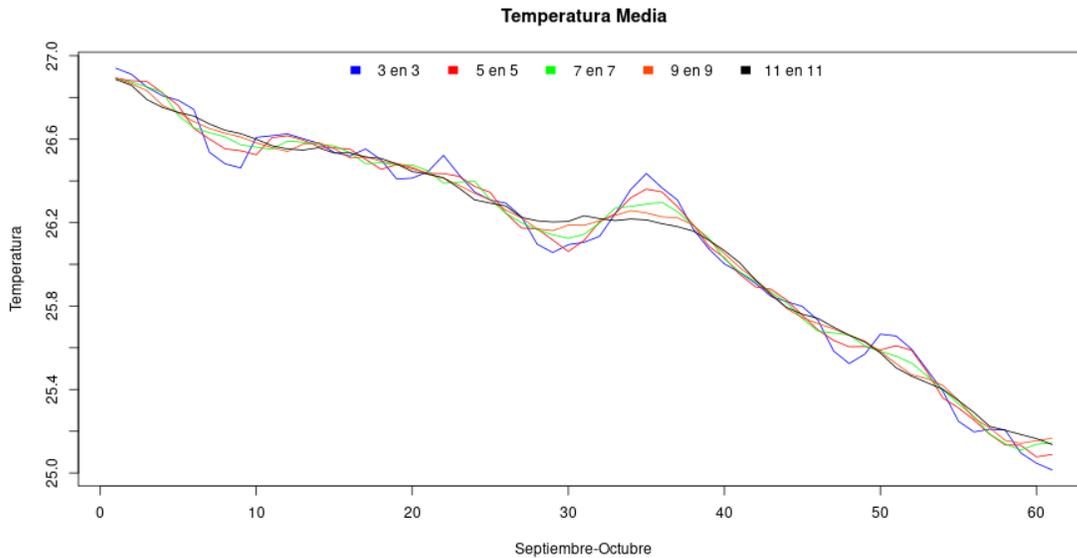


Figura 3.14. Representación de temperatura media en los distintos intervalos de media móvil (Bimestre Septiembre-Octubre)

En el último bimestre del año (Figura 3.15) la temperatura media continúa bajando hasta tener 2.5 °C de diferencia con respecto a la que tenía inicialmente. Los intervalos se muestran parejos y con pocas irregularidades, aunque se logra definir claramente el de mejor suavizado y viceversa, el de 11 en 11 y el de 3 en 3, respectivamente.

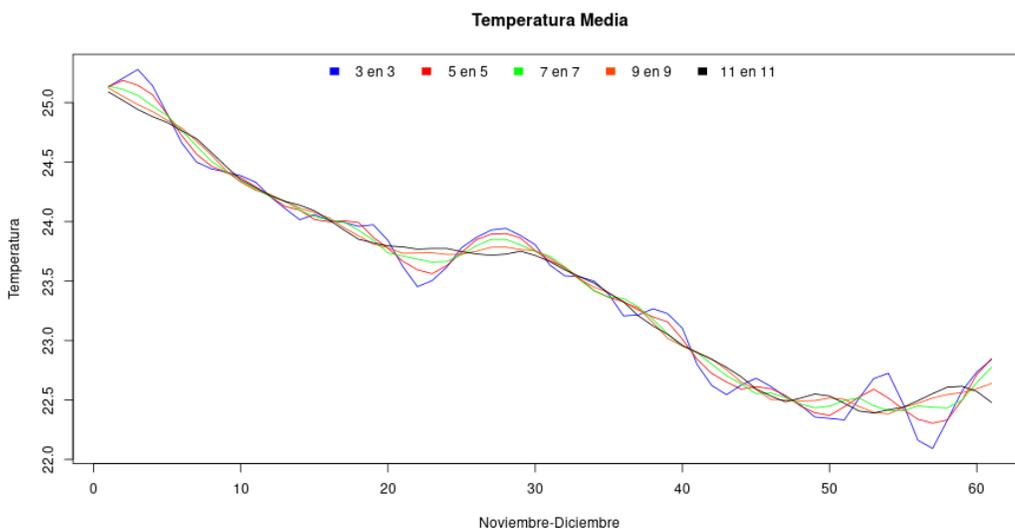


Figura 3.15. Representación de temperatura media en los distintos intervalos de media móvil (Bimestre Noviembre-Diciembre)

Al estudiar las suavizaciones que presenta la temperatura media, se valora que la media móvil cada 11 días es la que mejor suavización presenta en cada uno de los bimestres analizados, por lo que para elaborar el calendario térmico en la temperatura media se selecciona este intervalo como el mejor para su confección (Figura 3.16).

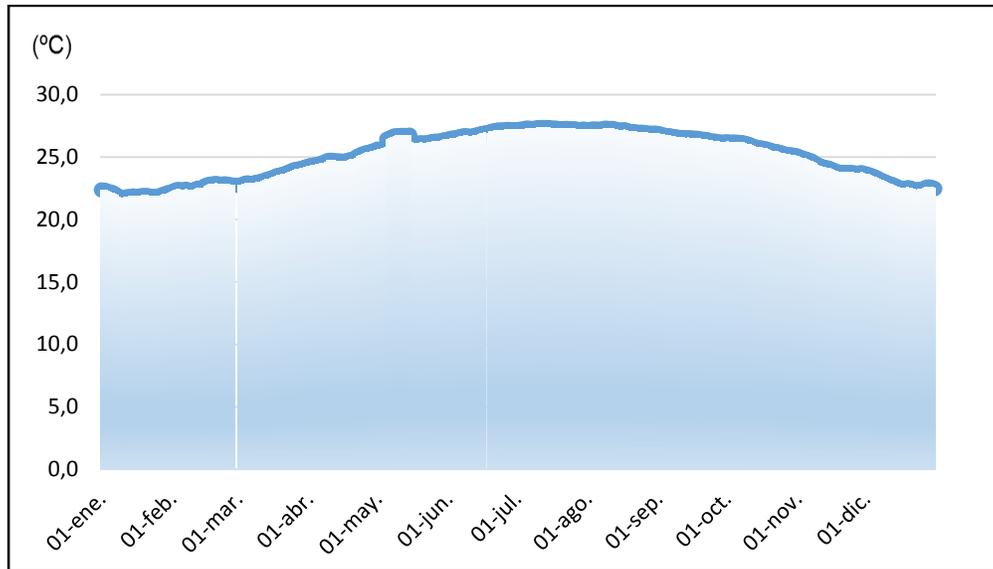


Figura 3.16. Calendario climático de temperatura media

3.1.3 CALENDARIO DE TEMPERATURA MÁXIMA

En la representación anual de los distintos intervalos de media móvil (Figura 3.17), se aprecia las diferencias entre los períodos cálidos y fríos, y a simple vista irregularidades en los intervalos, como el de 3 en 3 que sobresale como en las demás temperaturas. En el análisis de la temperatura máxima también se realiza observaciones bimestrales.

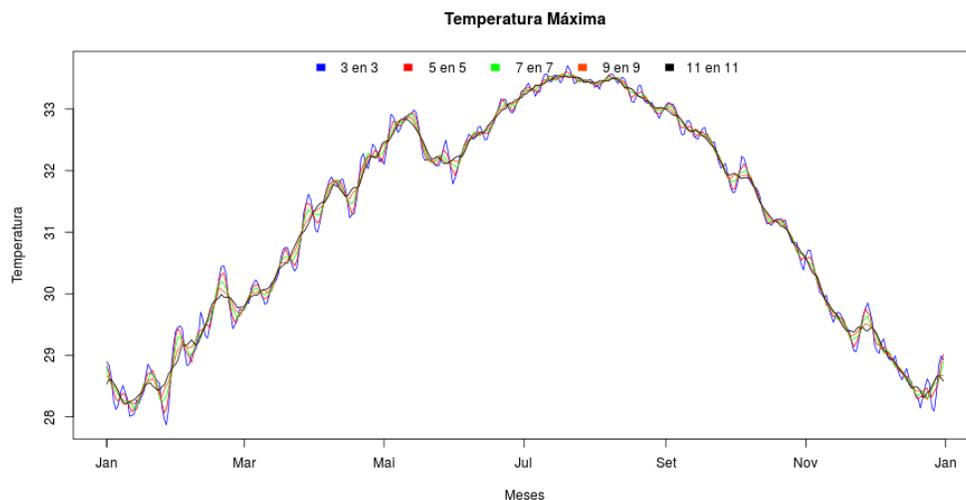


Figura 3.17. Representación anual de temperatura máxima en los distintos intervalos de media móvil

En el primer bimestre del año (Figura 3.18), desde el inicio, aunque en el segundo mes se acentúa más, hay irregularidades en los intervalos de 3 en 3, un poco menos en el de 5 en 5, ya en el de 7 en 7 baja con respecto a los anteriores, logrando una curva suave en el de 11 en 11.

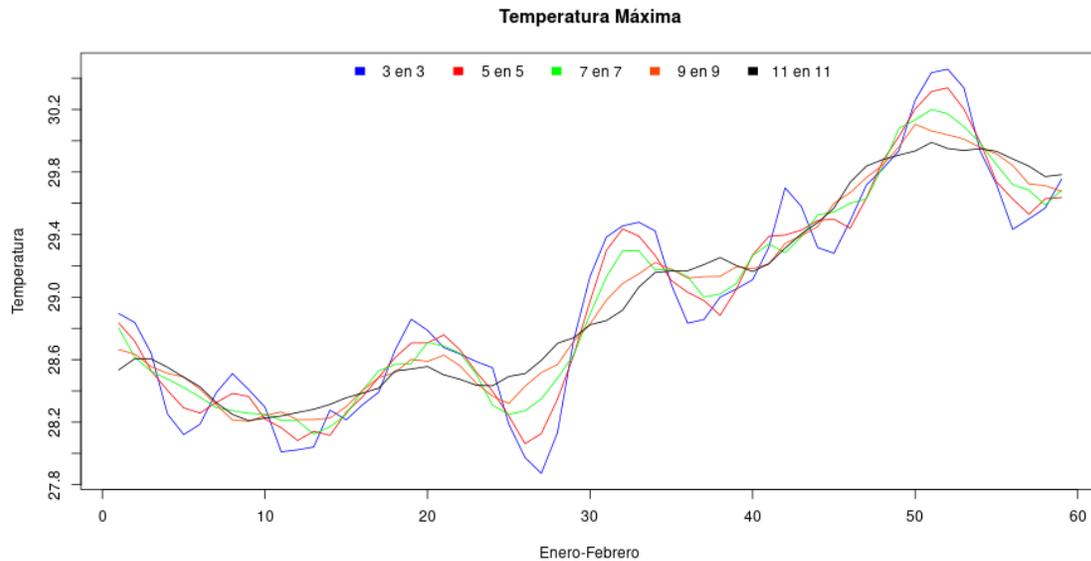


Figura 3.18. Representación de temperatura máxima en los distintos intervalos de media móvil (Bimestre Enero-Febrero)

La diferencia entre valores diarios hace que para el bimestre marzo-abril (Figura 3.19) manifieste alteraciones notables, apreciando las medias móviles de 3 días y de 11 días como las menos y más adecuadas, respectivamente, para la elaboración del calendario térmico en este bimestre. Las demás se mantienen con variaciones entre estas dos.

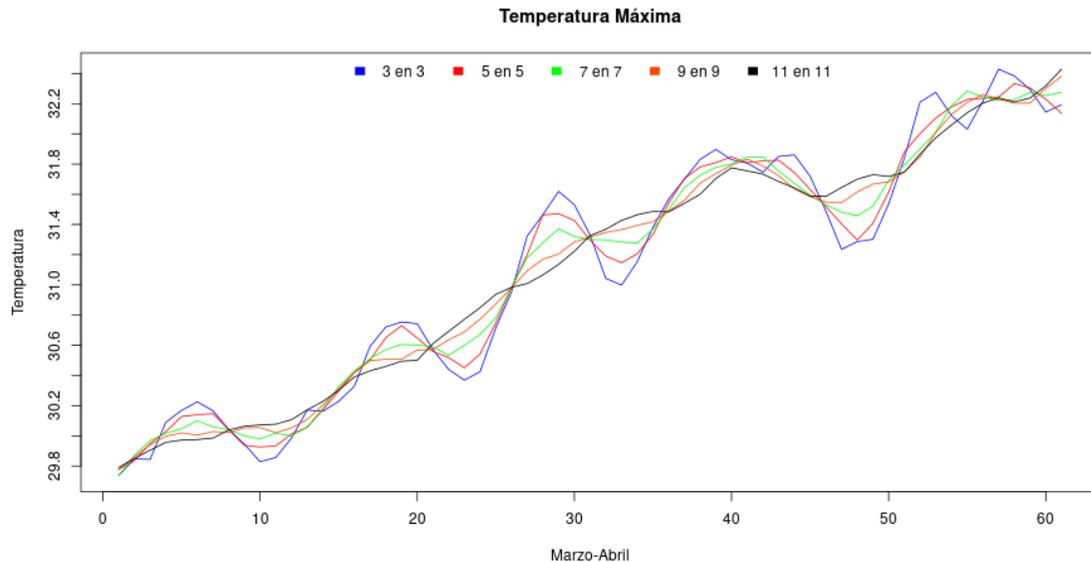


Figura 3.19. Representación de temperatura máxima en los distintos intervalos de media móvil (Bimestre Marzo-Abril)

En el bimestre de inicio del período cálido (Figura 3.20) se valora el ascenso gradual de la temperatura máxima, sobre todo en el mes de junio. En el segundo mes, aunque las diferencias diarias de valores son menos, no deja de ser la media móvil de 11 días la de mejor suavizado, y continúa mostrando la de 3 días grandes irregularidades.

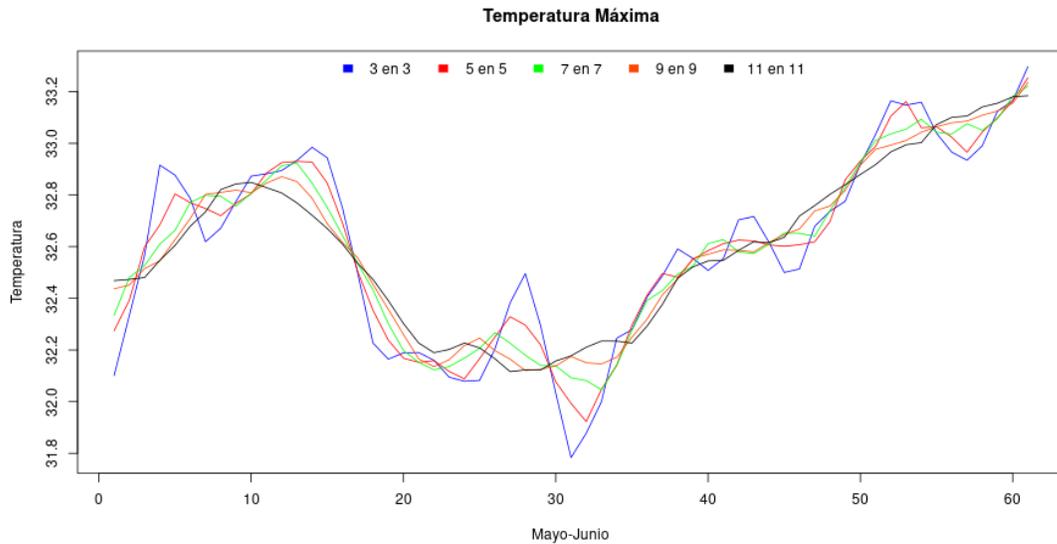


Figura 3.20. Representación de temperatura máxima en los distintos intervalos de media móvil (Bimestre Mayo-Junio)

A diferencia de las temperaturas mínima y media, el bimestre de julio-agosto (Figura 3.21) no presenta tan marcados contrastes entre sus intervalos, sin embargo, es indiscutible que el de 3 días es el de mayor irregularidad con respecto a los demás, continúa el de 11 días siendo en el que menos discrepancias se aprecia entre días consecutivos, si bien en el primer mes la temperatura se mantuvo en un rango bastante estable.

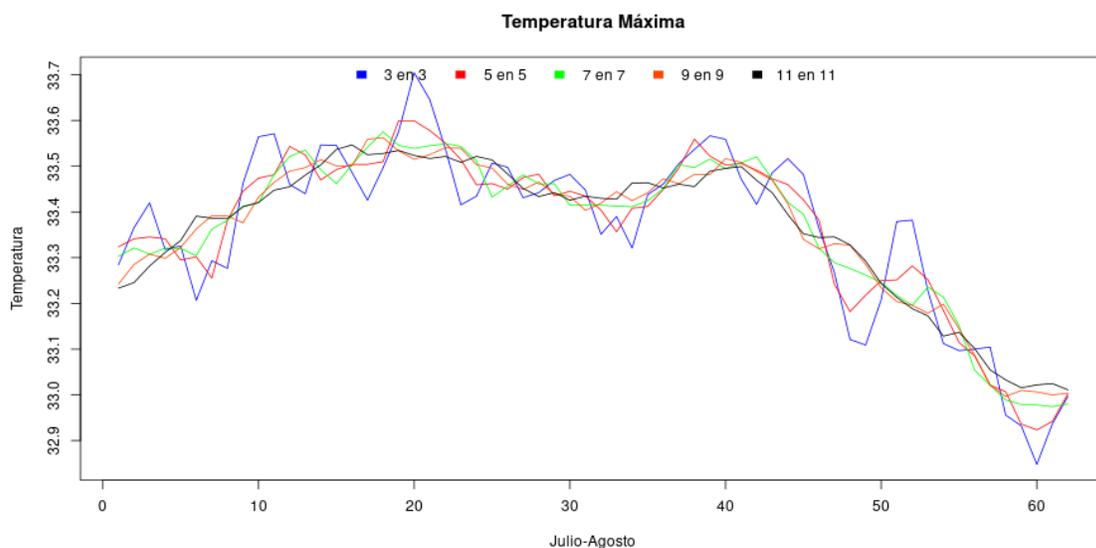


Figura 3.21. Representación de temperatura máxima en los distintos intervalos de media móvil (Bimestre Julio-Agosto)

Las curvas en el bimestre septiembre-octubre (Figura 3.22) son muy parejas entre sí a medida que los valores de temperatura van descendiendo, mostrándose un buen suavizado de las medias móviles a partir del segundo intervalo, pues el primero no escapa de ser el de mayor irregularidad, así como el último el de ser el de la mejor curva suavizada.

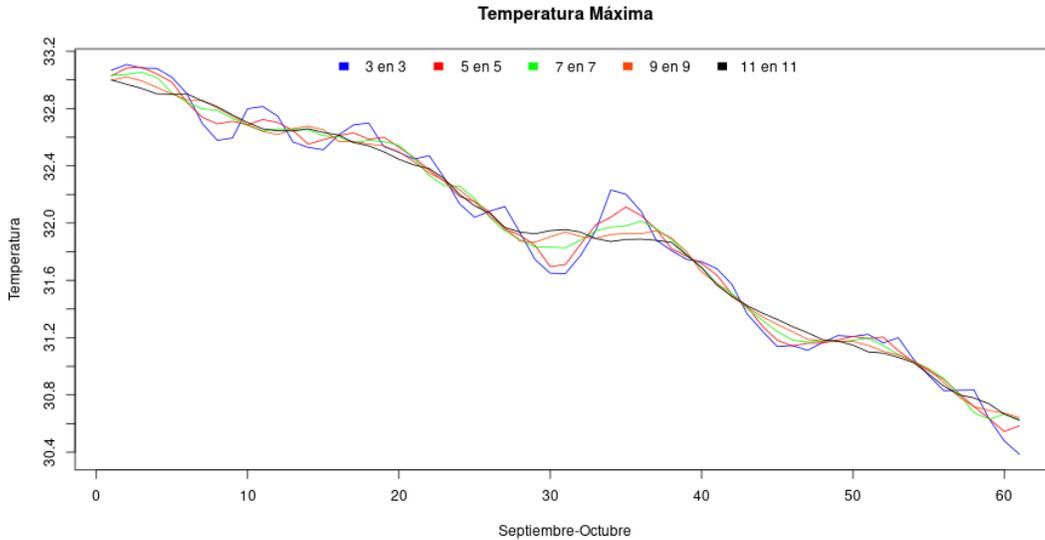


Figura 3.22. Representación de temperatura máxima en los distintos intervalos de media móvil (Bimestre Septiembre-Octubre)

En el último período del año (Figura 3.23), con el descenso de las temperaturas, las curvas van definiendo entre ellas este comportamiento sin grandes diferencias, aunque ya para finales de cada mes se distancian un poco, notándose la del intervalo de 3 días como la menos propicia para su uso por las diferencias que posee entre días consecutivos. Las demás, aunque son muy parejas, están por debajo de una curva muy suave, como la mostrada por el intervalo de 11 días.

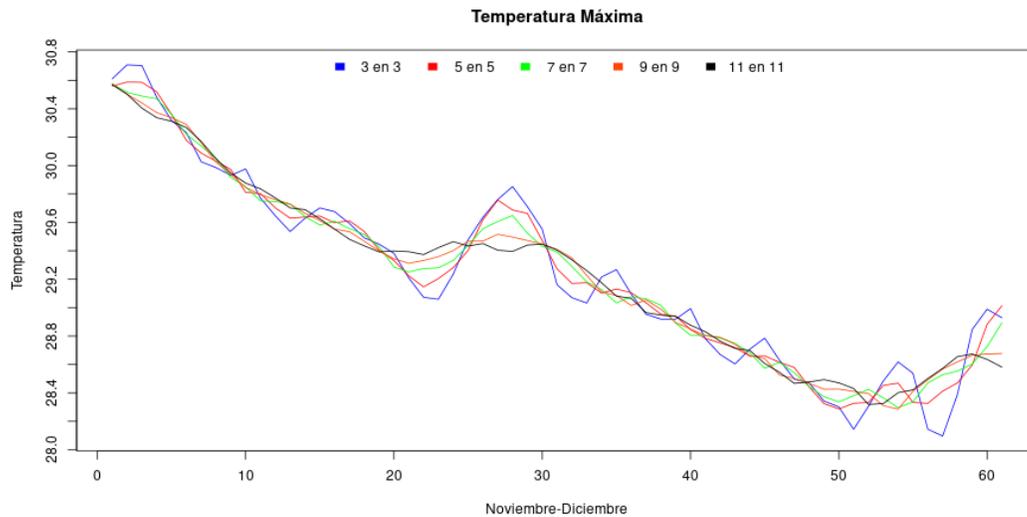


Figura 3.23. Representación de temperatura máxima en los distintos intervalos de media móvil (Bimestre Noviembre-Diciembre)

La temperatura máxima no es diferente a las demás con respecto a las curvas de mejor suavización, la de 11 días es la sobresaliente y que se selecciona para elaborar el calendario térmico en la temperatura máxima (Figura 3.24).

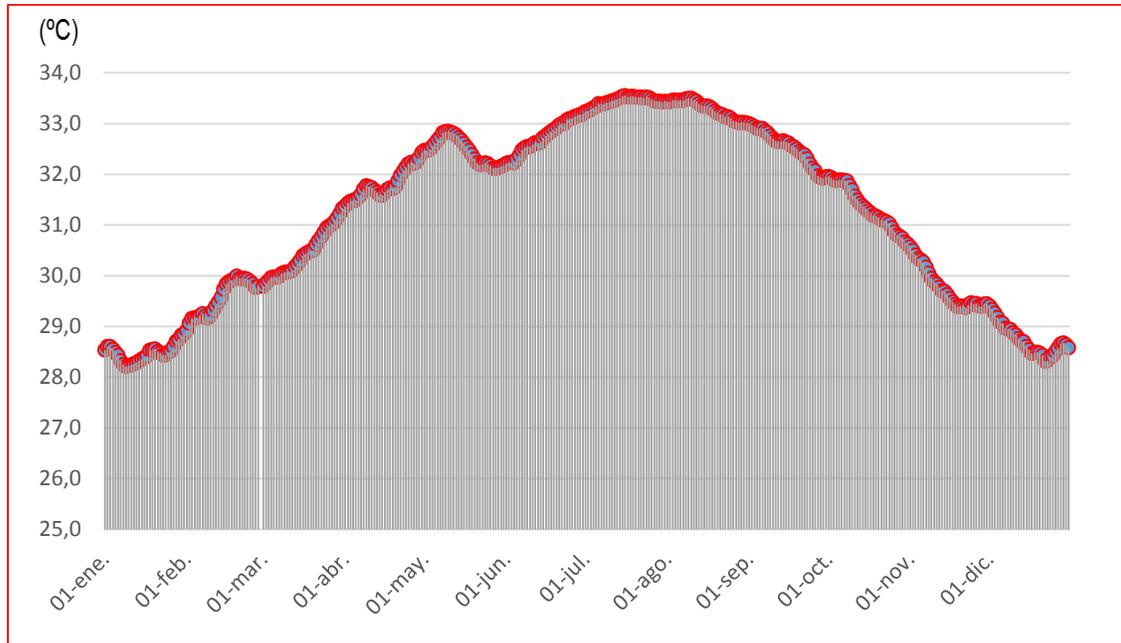


Figura 3.24. Calendario climático de temperatura máxima

En las tres temperaturas, mínima, media y máxima, prevalece la media móvil de 11 días como la de mejor suavizado con respecto a las demás, por lo que es ese intervalo el que se escoge para elaborar el calendario térmico en cada uno de ellas siguiendo esta metodología. Además, cuando se compara los resultados de este procedimiento con otros de distintas técnicas, se utiliza la media móvil cada 11 días.

3.2 CALENDARIO TÉRMICO UTILIZANDO OTRAS TÉCNICAS ESTADÍSTICAS

En este epígrafe se realiza un calendario térmico para cada medida de tendencia central seleccionada en cada una de las temperaturas (mínima, media y máxima). Cada medida, según sus características, arrojan valores diarios similares y muy próximos entre ellos en varias ocasiones, por lo que se estima el porcentaje que cada calendario presenta, con respecto al valor real (de 2011-2015), en un rango de ± 2 .

3.2.1 CALENDARIO DE TEMPERATURA MÍNIMA

En la tabla 3.1 se muestra por año la cantidad de veces que cada método estaba en un rango de ± 2 , así como el por ciento que cada uno representa. Al final se puede observar el total de los datos en los 5 años que presentaban esa condición y su por ciento. La

mediana es el método con mayor efectividad en los 5 años de comparación, con un 78.9 % de las veces. La media tijereada es la técnica que le sigue, así como la media móvil y la aritmética, con un 76 % de las veces. La moda se encuentra por debajo de las demás.

Tabla 3.1. Datos y porcentos de las técnicas estadísticas en la confección del calendario para la temperatura mínima

Años / %	Media móvil	Media aritmética	Mediana	Moda	Media tijereada
2011	291	293	293	267	296
%	79.7	80.3	80.3	73.2	81.1
2012	289	284	299	262	289
%	79.2	77.8	81.9	71.8	79.2
2013	268	267	281	265	274
%	73.4	73.2	77.0	72,6	75.1
2014	244	251	249	240	254
%	73.1	75.1	74.6	71.9	76.0
2015	248	246	269	259	259
%	74.0	73.4	80.3	77.3	77.3
Total 5 años	1340	1341	1391	1293	1372
% 5 años	76.0	76.0	78.9	73.3	77.8

3.2.2 CALENDARIO DE TEMPERATURA MEDIA

En la tabla 3.2 se muestra por año la cantidad de veces que cada método, en la temperatura media, estaba en un rango de ± 2 y el por ciento en cada uno. En general, todas las técnicas presentan una efectividad por encima del 80 %. Al evaluar los resultados, se observa que la mediana es el método con más garantías para que este logre estar en un rango de ± 2 con respecto a un valor real. La media móvil y la tijereada son las segundas más efectivas, con un 81.5 %, seguidos por la media aritmética y la moda.

Tabla 3.2. Datos y porcentajes de las técnicas estadísticas en la confección del calendario para la temperatura media

Años / %	Media móvil	Media aritmética	Mediana	Moda	Media tijereada
2011	342	339	335	317	336
%	93.7	92.9	91.8	86.8	92.1
2012	323	314	316	298	313
%	88.5	86.0	86.6	81.6	85.8
2013	308	308	324	300	312
%	84.4	84.4	88.8	82.2	85.5
2014	277	278	292	280	284
%	82.9	83.2	87.4	83.8	85.0
2015	252	252	266	263	257
%	75.2	75.2	79.4	78.5	76.72
Total 5 años	1502	1491	1533	1458	1502
% 5 años	85.1	84.5	86.9	82.7	85.1

3.2.3 CALENDARIO DE TEMPERATURA MÁXIMA

En cuanto a la temperatura máxima, también se analizaron las distintas técnicas para realizar un calendario térmico (Tabla 3.3). En esta ocasión, por un dato, que representa un 0.1 %, la media tijereada estuvo por encima de la mediana como método más efectivo, que estadísticamente no es significativa esta diferencia. Le siguieron la media móvil, muy cerca de las dos primeras, con solo 0.4 % de diferencia, la media aritmética y, un poco alejada, la moda.

Dados los resultados derivados en cada calendario de temperatura, se impone la mediana como el método más efectivo y el de mejores resultados para realizar un calendario y que este te ofrezca un valor cercano al real.

Tabla 3.3. Datos y porcentos de las técnicas estadísticas en la confección del calendario para la temperatura máxima

Años / %	Media móvil	Media aritmética	Mediana	Moda	Media tijereada
2011	309	301	307	290	306
%	84.7	82.5	84.1	79.5	83.8
2012	294	286	287	265	289
%	80.5	78.4	78.6	72.6	79.2
2013	303	302	311	289	307
%	83.0	82.7	85.2	79.2	84.1
2014	261	258	260	263	263
%	78.1	77.2	77.8	78.7	78.7
2015	254	249	262	244	263
%	75.8	74.3	78.2	72.8	78.5
Total 5 años	1421	1396	1427	1351	1428
% 5 años	80.6	79.1	80.9	76.6	81.0

3.3 COMPROBACIÓN DE LOS RESULTADOS CON PERÍODOS DE COMPORTAMIENTO NORMAL

Al comparar los calendarios de las distintas técnicas estadísticas (de cada temperatura), en los años posteriores al último con el que se conformó el período base, se facilita tener una idea de la efectividad de los métodos y escoger el mejor para conformar un calendario térmico lo más confiable posible. Sin embargo, al hacer estas comparaciones, no se tiene en cuenta que estos pueden estar marcados por diferencias climáticas, según las condiciones sinópticas imperantes, y se obtendría una efectividad alejada a la que realmente podría ofrecer cada una. Es por eso que se realiza una comparación de cada técnica con respecto a los meses que hayan tenido un comportamiento normal de la variable en la zona de estudio y así corroborar la mejor técnica a utilizarse.

Como fueron tan pocos datos para la comparación, y los resultados estar muy próximos e iguales en ocasiones, se decidió analizar también los métodos en un rango de ± 1 , y así poder ofrecer mejores consideraciones.

3.3.1 TEMPERATURA MÍNIMA

Cuando se compara las distintas técnicas con meses con un comportamiento climático normal, en un rango de ± 2 (Tabla 3.4), se aprecia que la mediana es el método con mejores resultados, con un 78.9 %, seguida por la media móvil y la tijerada, ambas con 78.1 %.

Tabla 3.4. Comparación de las técnicas estadísticas con meses que presentan un comportamiento normal en un rango de ± 2 . Temperatura mínima

Método	Cant. Veces en un año	%
Media móvil	285	78.1
Media aritmética	279	76.4
Mediana	288	78.9
Moda	256	70.1
Media tijereada	285	78.1

En el análisis de los datos que se encuentran en el rango de ± 1 (Tabla 3.5), excepto la moda todos sobrepasan el 50 %, la mediana es el de mejor efectividad con casi el 53 % de las veces con valores cercanos al real, aunque la media móvil difiere en solo un dato y la media tijereada y la media aritmética están muy próximas también.

Tabla 3.5. Comparación de las técnicas estadísticas con meses que presentan un comportamiento normal en un rango de ± 1 . Temperatura mínima

Método	Cant. Veces en un año	%
Media móvil	192	52.6
Media aritmética	184	50.4
Mediana	193	52.9
Moda	173	47.4
Media tijereada	186	51.0

3.3.2 TEMPERATURA MEDIA

Al examinar la Tabla 3.6, aunque la media móvil esté por encima de la mediana y otros métodos como la media aritmética y la tijerada, no hay mucha diferencia entre unas y otras, pues la más próxima, la mediana, está a 4 valores (1.1 %) de igualar ese resultado.

La moda difiere en casi 20 datos del primero, con un bajo valor de efectividad como se ha venido observando en otras ocasiones.

Tabla 3.6. Comparación de las técnicas estadísticas con meses que presentan un comportamiento normal en un rango de ± 2 . Temperatura media

Método	Cant. Veces en un año	%
Media móvil	331	90.7
Media aritmética	324	88.8
Mediana	327	89.6
Moda	312	85.5
Media tijereada	324	88.8

Al reducir más el intervalo (Tabla 3.7) la media tijereada es la que mejor acierto tiene con un 67.1 % de las veces con valores cercanos a un rango de ± 1 . Si bien es cierto que estos porcentajes nos dan una medida de efectividad, si se mira a los totales, el mejor resultado está a una diferencia de 4 de la mediana y la media aritmética, para un 1.1 % de ventaja, lo cual no viene siendo significativo, la moda si está por debajo en casi un 13 % y se aleja mucho de la efectividad que pudiera tener en comparación con el primero.

Tabla 3.7. Comparación de las técnicas estadísticas con meses que presentan un comportamiento normal en un rango de ± 1 . Temperatura media

Método	Cant. Veces en un año	%
Media móvil	237	64.9
Media aritmética	241	66.0
Mediana	241	66.0
Moda	198	54.2
Media tijereada	245	67.1

3.3.3 TEMPERATURA MÁXIMA

En la comparación de los resultados de la Tabla 3.8, la mediana posee un resultado igual al de la moda tijereada, 82.5 %, pues ambas estuvieron la misma cantidad de valores en un rango ± 2 en comparación con una observación real. Los valores más cercanos son la media móvil y la media aritmética, a 4 y 5 valores, respectivamente, de tener una cantidad análoga a los primeros.

Tabla 3.8. Comparación de las técnicas estadísticas con meses que presentan un comportamiento normal en un rango de ± 2 . Temperatura máxima

Método	Cantidad de veces en un año	%
Media móvil	297	81.4
Media aritmética	296	81.1
Mediana	301	82.5
Moda	281	77.0
Media tijereada	301	82.5

En la Tabla 3.9 la mediana aparece como mejor medida de tendencia central si se establece un rango de comparación de ± 1 , con un 57 % de efectividad, alejada 1.9 % de la media tijereada, que representa 7 valores, como el método más cercano. La media móvil y la aritmética presentan valores porcentuales de más de 2.5 de diferencia.

Tabla 3.9. Comparación de las técnicas estadísticas con meses que presentan un comportamiento normal en un rango de ± 1 . Temperatura máxima

Método	Cantidad de veces en un año	%
Media móvil	199	54.5
Media aritmética	194	53.2
Mediana	208	57.0
Moda	190	52.1
Media tijereada	201	55.1

Mientras más grande sea una base de datos, mejor serían el tratamiento estadístico, y por ende, los resultados. Al comparar con estos meses de comportamiento climático normal, solo se corroboraría la técnica que mejor se aproxima si se tuviese que utilizar un calendario térmico y no existiesen grandes cambios en los patrones climatológicos. En los casos analizados, en la comparación de los 5 años, la mediana fue el mejor método estadístico para la temperatura mínima y media; en las comprobaciones de meses con comportamiento normal, para un rango de ± 2 , fue más efectiva en la temperatura mínima y máxima, y en el ± 1 , igual. Por lo que, de las 6 exploraciones, en 4 fue superior, y en las otras 2, la diferencia fue mínima con respecto a la que ocupaba la mejor efectividad.

3.4 VERIFICACIÓN DEL PRECIS-CARIBE

Al realizar el análisis exploratorio de los datos del PRECIS CARIBE, las salidas diarias para el escenario B2 no se ajustaron a los valores reales medidos por la estación (los valores de temperatura mínima y media fueron superiores), por lo que se les realizó a las variables una corrección con la determinación de un coeficiente bimestral y el ajuste los valores del modelo a un comportamiento similar al de la zona. Estos cálculos permitieron verificar que al realizar la corrección los valores corregidos tienen en comparación con las observaciones de la estación un comportamiento medio mensual igual, para el período lluvioso, y similar para el período poco lluvioso (Tabla 3.10).

Tabla 3.10. Valores medios mensuales de temperatura en los períodos estacionales (1991-2010).

		Tmin	Tmax	Tm
Período Lluvioso (cálido)	Estación	22.3	32.6	26.6
	Modelo	25.8	32.3	28.9
	Corrección	22.3	32.6	26.6
Período Poco Lluvioso (invernal)	Estación	18.3	29.9	23.4
	Modelo	23.2	30.2	26.4
	Corrección	18.2	29.6	23.2

En la Tabla 3.11 se observa que al calcular el error cuadrático medio (RMSE) los valores corregidos tienen menos incertidumbre que los modelados, el error absoluto medio (MAE) es inapreciable en la corrección y según el nivel de fiabilidad (BE) siguen un comportamiento análogo a los datos históricos.

Tabla 3.11. Cálculo de RMSE, MAE y BE para las variables meteorológicas modeladas y corregidas en los períodos estacionales (1991-2010).

	Período lluvioso (cálido)						Período poco lluvioso (invernal)					
	RMSE		MAE		BE		RMSE		MAE		BE	
	Mod	Corr	Mod	Corr	Mod	Corr	Mod	Corr	Mod	Corr	Mod	Corr
Tmin	3.8	1.4	3.5	0.02	3.5	-0.02	6.1	3.0	5.0	0.01	5.0	0.01
Tmax	2.7	1.3	2.3	0.02	2.3	-0.02	4.4	2.5	3.3	0.01	3.3	0.01
Tmed	2.1	1.9	0.4	-0.04	-0.02	-0.02	3.7	3.0	0.6	0.01	0.6	0.005

3.5 COMPARACIÓN DEL CALENDARIO PROPUESTO CON EL PERÍODO 2016-2020

Para valorar la certidumbre que tiene el calendario propuesto en futuras observaciones, se compara con datos del período 2016-2020, tomados del PRECIS CARIBE y se mide la efectividad que tiene el calendario con datos futuros. Una vez que el período base fue ajustado, a las predicciones de los años 2016-2020 también les fue aplicado el coeficiente para el acomodo de las variables. El tratamiento de los datos fue en el programa estadístico STATGRAPHICS y los resultados para cada temperatura se muestran a continuación:

3.5.1 TEMPERATURA MÍNIMA

En Figura 3.25 aparecen 3 columnas en dependencia de los datos, con 21498 valores trazados. Se puede apreciar que la dispersión entre los datos corregidos y los de la estación son similares, no así con los del modelo que difiere de los dos anteriores.

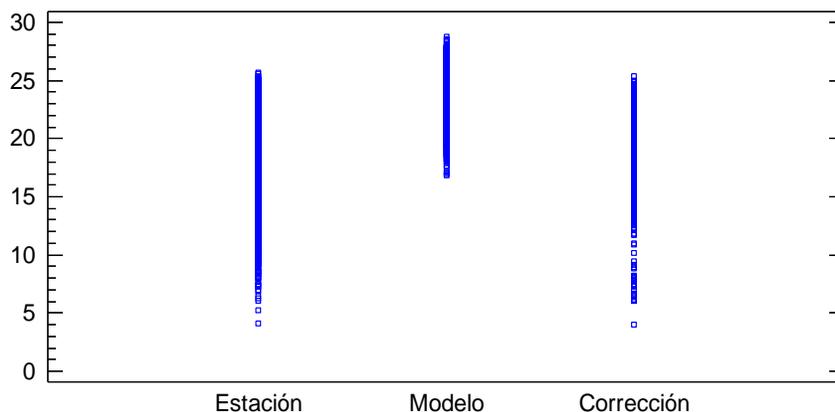


Figura 3.25. Diagrama de dispersión según muestra de la temperatura mínima

El sesgo estandarizado y la curtosis estandarizada (Tabla 3.12) se encuentran fuera del rango de -2 a +2 para las 3 columnas de datos, lo que indica algo de no normalidad significativa en los datos, por lo que se incumple con el supuesto de que los datos provienen de distribuciones normales.

Tabla 3.12. Sesgo estandarizado y curtosis estandarizada. Temperatura mínima.

	Sesgo Estandarizado	Curtosis Estandarizada
Estación	-43.4727	29.5457
Modelo	-27.3818	5.70688
Corrección	-34.3451	29.7903
Total	-49.3615	32.7993

Al observar la varianza de los datos (Tabla 3.13), el valor-P de la prueba-F es menor que 0.05, por lo que la hipótesis nula H_0 de medias iguales se rechaza, existiendo diferencias estadísticamente significativas entre las medias de las tres variables con un nivel del 95.0% de confianza. En la Figura 3.26 se aprecian los intervalos de los pares de media basados en el procedimiento de la diferencia mínima significativa (LSD) de Fisher, definiéndose que tienen una diferencia estadísticamente significativa la columna de la Estación y la Corrección con la de Modelo.

Tabla 3.13. ANOVA. Temperatura mínima.

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	84347.2	2	42173.6	6662.16	0.0000
Intra grupos	136026.	21488	6.33032		
Total (Corr.)	220373.	21490			

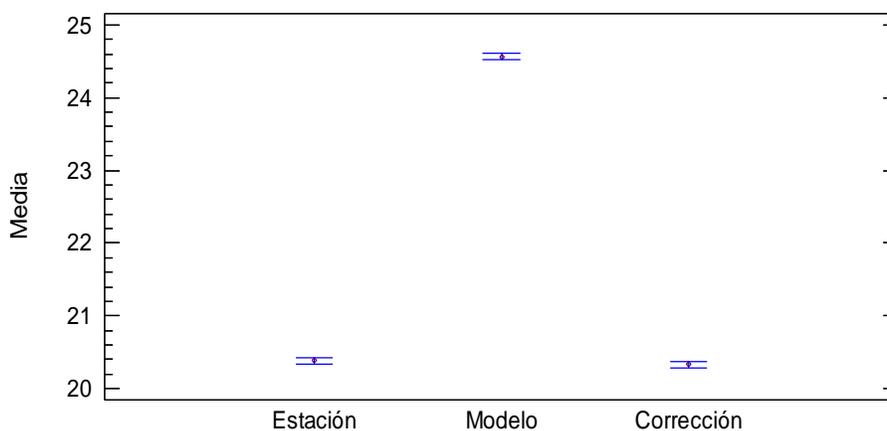


Figura 3.26. Diferencia mínima significativa (LSD) de Fisher. Temperatura mínima
Al realizar las Pruebas de Múltiples Rangos se identificaron 2 grupos homogéneos² (Tabla 3.14), entre ellos no existen diferencias estadísticamente significativas.

Tabla 3.14. Método: 95.0 porcentaje LSD. Temperatura mínima.

	Casos	Media	Grupos Homogéneos
Corrección	7163	20.3322	X
Estación	7163	20.3832	X
Modelo	7163	24.56	X

²Según la alineación de las X en columnas.

La Tabla 3.15 muestra las diferencias estimadas entre cada par de medias. El asterisco que se encuentra al lado de los dos pares indica que muestran diferencias estadísticamente significativas con un nivel del 95.0% de confianza. Entre la estación y los datos corregidos no existen diferencias estadísticamente significativas.

Tabla 3.15. Determinación de medias significativas. Temperatura mínima.

Contraste	Sig.	Diferencia	+/- Límites
Estación - Modelo	*	-4.17675	0.0823975
Estación - Corrección		0.0510446	0.0823975
Modelo - Corrección	*	4.22779	0.0823946

En la prueba de Kruskal-Wallis (Tabla 3.16), para comparar las medianas en lugar de las medias, el valor-P es menor que 0.05, por lo que existe una diferencia estadísticamente significativa entre las medianas con un nivel del 95.0% de confianza y se rechaza la hipótesis nula de que las medianas dentro de cada una de las columnas es la misma.

3.16. Prueba de Kruskal-Wallis. Temperatura mínima.

	Tamaño de Muestra	Rango Promedio
Estación	7163	8006.91
Modelo	7163	16732.9
Corrección	7163	7497.81

Estadístico = 10032.5 Valor-P = 0.0

En la Figura 3.27 se muestran 3 diagramas de caja y bigote, una para cada columna de datos. En este caso, hay 378 puntos alejados y 27 puntos muy alejados. La presencia de puntos muy alejados indica valores aberrantes. Como se está analizando la variable temperatura esto es bastante común, pues existen valores extremos que se van a comportar dadas las condiciones sinópticas imperantes de ese momento, como frentes fríos, cálidos, etc., y que después se pueden ver reflejados como en esta figura. El modelo PRECIS CARIBE no contempla valores extremos en sus datos modelados, por eso es que se aprecian pocos valores aberrantes.

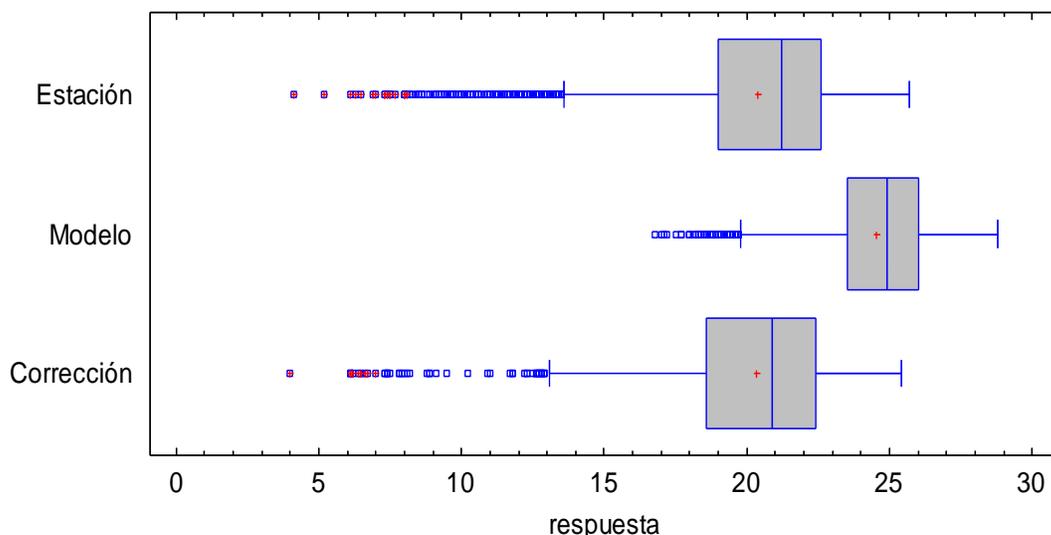


Figura 3.27. Gráfico de caja y bigotes de la temperatura mínima

Como se pudo apreciar, los datos de la estación no presentan diferencias estadísticamente significativas con los corregidos, y en los distintos análisis se consideran como valores con comportamiento análogo. Es por eso que para los años de 2016 a 2020, se les aplica a los datos del modelo los coeficientes de corrección bimestrales para que presenten un comportamiento similar al de la zona.

Al analizar la Tabla 3.17 para la relación entre los datos del calendario térmico y los del PRECIS CARIBE, obtenida por el programa STATGRAPHICS, el valor-P de la razón-F es mayor que 0.05, por lo que no existe una diferencia estadísticamente significativa entre las medias de las 6 variables, por lo que se trabajó con esos datos.

Tabla 3.17. ANOVA para datos del calendario térmico y PRECIS CARIBE

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	51.8193	5	10.3639	1.87106	0.0962
Intra grupos	11931.0	2154	5.53902		
Total (Corr.)	11982.9	2159			

En la Tabla 3.18 se compara la efectividad que tiene el calendario térmico propuesto (método de la mediana) con los valores modelados de los años 2016 a 2020.

Se considera buena efectividad a partir de los datos obtenidos. Los meses invernales son los que más difieren en los valores dentro del rango de ± 2 °C. Como es el caso de la temperatura mínima, esto es un resultado posible, pues en estos meses llegan sistemas frontales a esta zona y el valor real difiere en ocasiones del pronosticado.

Tabla 3.18. Efectividad del calendario térmico en los años de 2016-2020. Temperatura mínima.

Veces en el rango ± 2		
Años	Cantidad	%
2016	336	93.3
2017	341	94.7
2018	333	92.5
2019	347	96.4
2020	335	93.1
Total	1692	94.0

3.5.2 TEMPERATURA MEDIA

En la Figura 3.28 se muestra el diagrama de dispersión según la muestra. En las 3 columnas aparecen 21498 valores trazados. Entre los datos de la estación y los corregidos hay más homogeneidad, aunque los corregidos presentan valores más bajos que los registrados. Los datos del modelo son más concentrados en un rango menor con respecto al de las otras dos columnas.

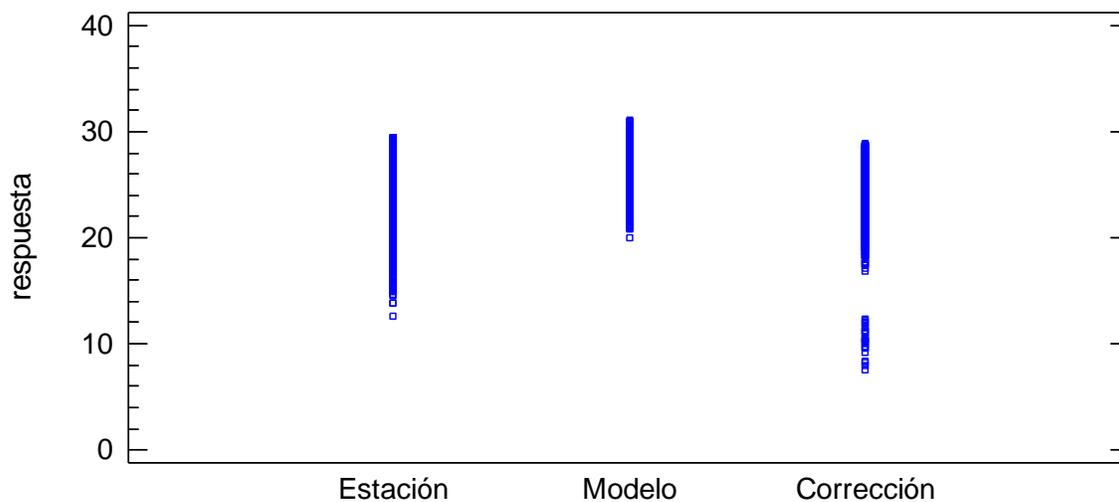


Figura 3.28. Diagrama de dispersión según muestra de la temperatura media

Los datos no provienen de distribuciones normales, pues al analizar el sesgo estandarizado y la curtosis estandarizada (Tabla 3.19), estos se encuentran fuera del rango de -2 a $+2$ para las 3 columnas, indicando algo de no normalidad significativa en los datos.

Tabla 3.19. Sesgo estandarizado y curtosis estandarizada. Temperatura media

	Sesgo Estandarizado	Curtosis Estandarizada
Estación	-34.7411	22.0935
Modelo	-24.6321	5.2793
Corrección	-53.5483	117.457
Total	-56.7988	72.7841

En la Tabla 3.20 se observa el análisis de varianza de los datos: el valor-P de la prueba-F es menor que 0.05, por lo que la hipótesis nula H_0 de medias iguales se rechaza, existiendo diferencias estadísticamente significativas entre las medias de las tres variables con un nivel del 95 % de confianza, además de comprobarse gráficamente en la Figura 3.29.

Tabla 3.20. ANOVA. Temperatura media

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	34560.4	2	17280.2	3934.24	0.0000
Intra grupos	94420.3	21497	4.39226		
Total (Corr.)	128981.	21499			

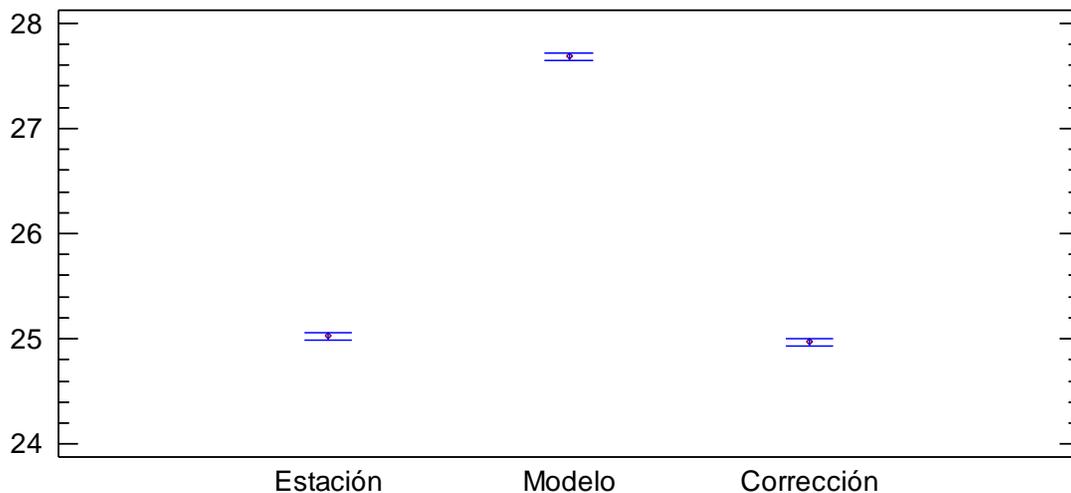


Figura 3.29. Diferencia mínima significativa (LSD) de Fisher. Temperatura media

Con las Pruebas de Múltiples Rangos se identificaron 2 grupos homogéneos³ (Tabla 3.21), entre ellos no existen diferencias estadísticamente significativas.

³Según la alineación de las X en columnas.

Tabla 3.21. Método: 95.0 porcentaje LSD. Temperatura media

	Casos	Media	Grupos Homogéneos
Corrección	7167	24.9674	X
Estación	7166	25.0259	X
Modelo	7167	27.6857	X

La Tabla 3.22 muestra las diferencias estimadas entre cada par de medias, representadas por el asterisco que se encuentra al lado de los dos pares, revelando diferencias estadísticamente significativas con un nivel del 95.0 % de confianza. Entre la estación y los datos corregidos no existen diferencias estadísticamente significativas.

Tabla 3.22. Determinación de medias significativas. Temperatura media

Contraste	Sig.	Diferencia	+/- Límites
Estación - Modelo	*	-2.65974	0.0686205
Estación - Corrección		0.0585497	0.0686205
Modelo - Corrección	*	2.71829	0.0686181

Para comparar las medianas en lugar de las medias, en la prueba de Kruskal-Wallis (Tabla 3.23), el valor-P es menor que 0.05, por lo que existe una diferencia estadísticamente significativa entre las medianas con un nivel del 95.0 % de confianza y se rechaza la hipótesis nula de que las medianas dentro de cada una de las columnas es la misma.

3.23. Prueba de Kruskal-Wallis. Temperatura media

	Tamaño de Muestra	Rango Promedio
Estación	7166	8490.47
Modelo	7166	15617.9
Corrección	7166	8142.85

Estadístico = 10032.5 Valor-P = 0.0

En los diagramas de caja y bigote, representados en la Figura 3.29, existen 262 puntos alejados y 31 puntos muy alejados. La presencia de puntos muy alejados indica valores aberrantes, conocidos como valores extremos, que no tienen representación climática. Los

datos del modelo, alejados estadísticamente de los demás, siguen presentando pocos valores apartados.

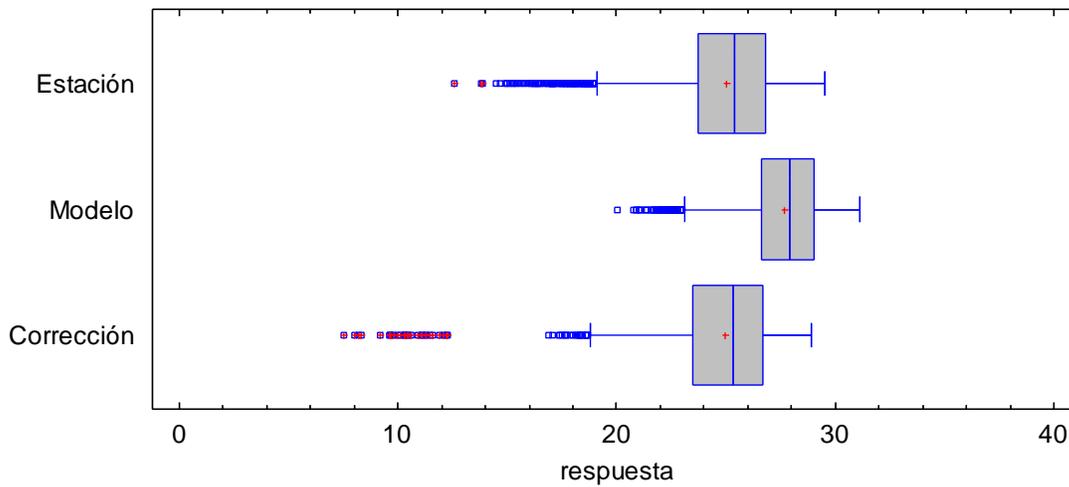


Figura 3.29. Gráfico de caja y bigotes de la temperatura media

También en la temperatura media, entre los datos de la estación y los corregidos no se exhiben diferencias estadísticamente significativas en el período base. Para los años de 2016 a 2020, a los datos del modelo se les aplica los coeficientes de corrección para que muestren un comportamiento conforme al del área de estudio.

Al analizar la Tabla 3.24 para la relación entre los datos del calendario térmico y los del PRECIS CARIBE, obtenida por el programa STATGRAPHICS, el valor-P de la razón-F es menor que 0.05, por lo que existe una diferencia estadísticamente significativa entre las medias de las 6 variables.

Tabla 3.24. ANOVA para datos del calendario térmico y PRECIS CARIBE

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	129.484	5	25.8967	6.14	0.0000
Intra grupos	18185.2	4314	4.21539		
Total (Corr.)	18314.7	4319			

En la Prueba de Múltiples Rangos en el STATGRAPHICS, tanto en la Tabla 3.25 como en la Figura 3.30, se aprecia dónde es que existe la diferencia estadísticamente significativa. Menos el año 2019, todos los demás son grupos homogéneos. El análisis y el coeficiente de corrección fueron realizados como mismo a los demás años, pero según un estudio

detallado para darle un por qué a este comportamiento, se delibera que el modelo en el 2019 experimenta un incremento de la temperatura media con respecto a los demás años, por eso es que aprecia este resultado. Es por eso que se decide no trabajar el año 2019 para ver la efectividad del calendario propuesto, dada la irregularidad en el comportamiento climático que se proyecta.

Tabla 3.25. Método: 95.0 porcentaje LSD. Datos del calendario térmico y PRECIS CARIBE.
Temperatura media

	Casos	Media	Grupos Homogéneos
PRECIS 2016	720	25.1722	X
PRECIS 2018	720	25.28	X
PRECIS 2017	720	25.2822	X
Calendario propuesto	720	25.2853	X
PRECIS 2020	720	25.3169	X
PRECIS 2019	720	25.7158	X

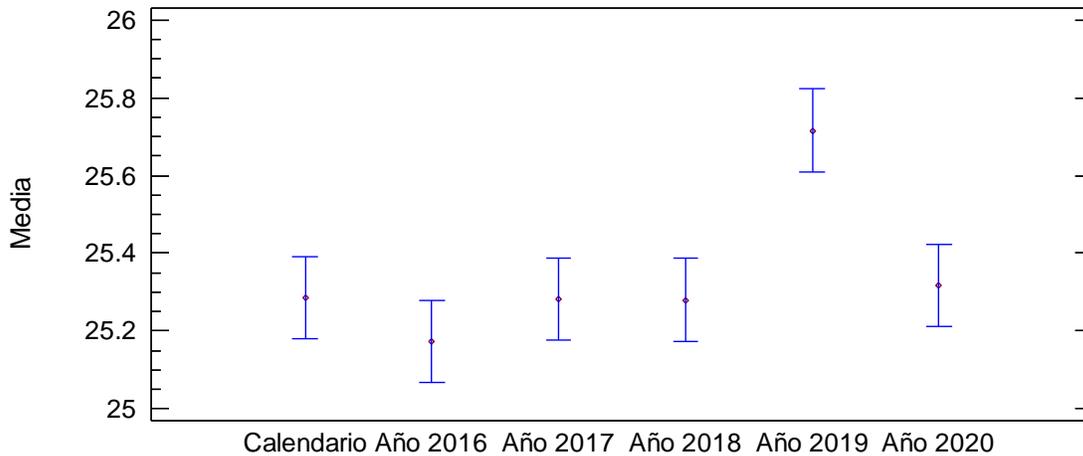


Figura 3.30. Diferencia mínima significativa (LSD) de Fisher. Calendario térmico y PRECIS CARIBE.

Al comparar la efectividad que tiene el calendario térmico propuesto con los valores modelados de los años 2016 a 2020, exceptuando el 2019 (Tabla 3.26), la certidumbre de los pronósticos es superior al 97 % en este período, resultando muy buena, teniendo en consideración que como estamos hablando de una variable meteorológica, siempre van a existir situaciones sinópticas que van a alterar el comportamiento de los territorios, y el

calendario desarrollado con el método de la mediana propone poca variación en años posteriores.

Tabla 3.26. Efectividad del calendario térmico en los años de 2016-2020. Temperatura media

Veces en el rango ± 2		
Años	Cantidad	%
2016	339	94.2
2017	352	97.8
2018	355	98.6
2020	354	98.3
Total	1400	97.2

3.5.3 TEMPERATURA MÁXIMA

Al analizar los datos del período base de la estación con los modelados por el PRECIS CARIBE, según la dispersión de los datos (Figura 3.31) las dos columnas tienen bastante aproximación, exceptuando los puntos extremos, que el modelo no es capaz de representarlos.

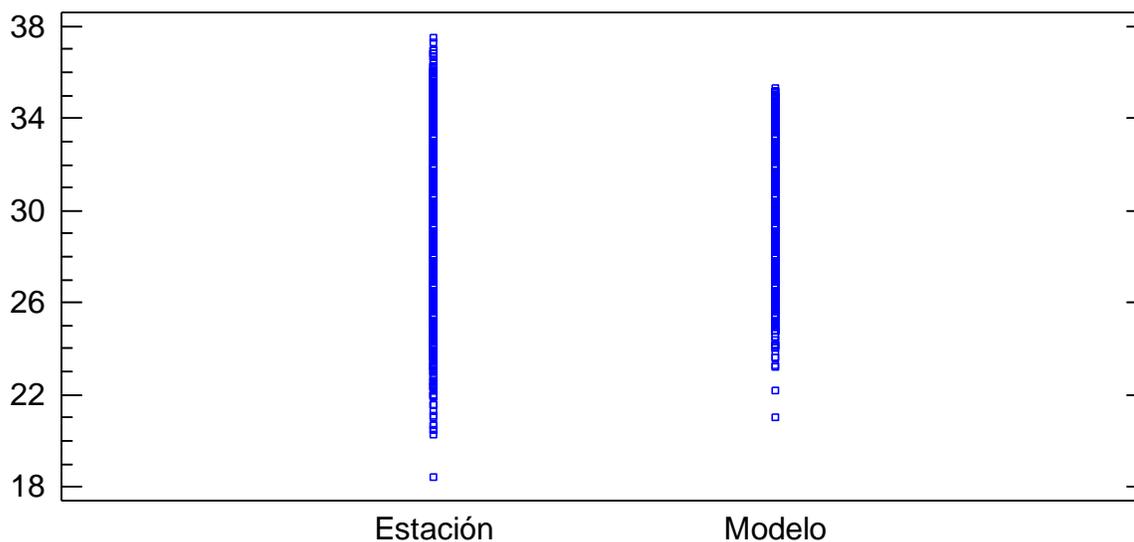


Figura 3.31. Diagrama de dispersión según muestra de la temperatura máxima

En la Tabla 3.27 el valor-P de la razón-F es mayor que 0.05, por lo que no existe una diferencia estadísticamente significativa entre las medias de los 2 grupos analizados.

Tabla 3.27. Tabla ANOVA. Temperatura máxima

Fuente	Suma de Cuadrados	GI	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	0.273351	1	0.273351	0.06	0.8133
Intra grupos	70267.6	14334	4.90216		
Total (Corr.)	70267.9	14335			

Si se aplica también el procedimiento de diferencia mínima significativa (LSD) de Fisher, tanto de forma tabular (Tabla 3.28) como gráficamente (Figura 3.32), se aprecia que no hay diferencias estadísticamente significativas entre el par de medias de los datos. Es por eso que para trabajar con la temperatura máxima se utilizaron los datos del modelo sin hacerle ninguna corrección.

Tabla 3.28. Método: 95.0 porcentaje LSD. Temperatura máxima

	Casos	Media	Grupos Homogéneos
Modelo	7168	31.2794	X
Estación	7168	31.2882	X

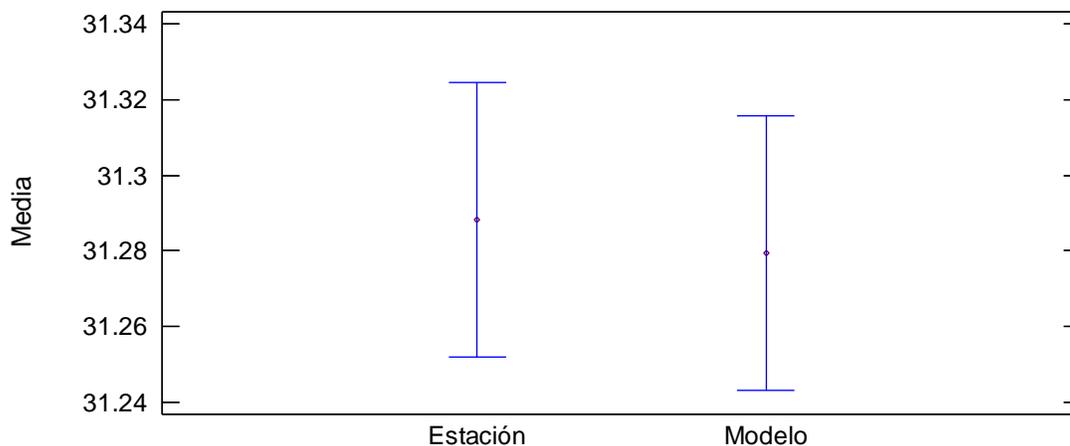


Figura 3.32. Diferencia mínima significativa (LSD) de Fisher. Temperatura máxima

Al analizar los datos del calendario térmico propuesto con los del PRECIS CARIBE, en los años 2016-2020, se aplica el procedimiento de comparación múltiple para determinar si las medias de los datos son significativamente diferentes de otras (Tabla 3.29), apreciando que existe diferencias estadísticamente significativas entre varios años, sin embargo, el

calendario, no presenta discrepancias con los años del período, a no ser con el 2019, que sí difiere de todos los demás.

Tabla 3.29. Método: 95.0 porcentaje LSD. Datos del calendario térmico y PRECIS CARIBE. Temperatura máxima

	Casos	Media	Grupos Homogéneos
Año 2016	360	31.2806	X
Calendario	360	31.5028	XX
Año 2017	360	31.5617	XX
Año 2020	360	31.6428	X
Año 2018	360	31.6972	X
Año 2019	360	32.1828	X

En la prueba de Kruskal-Wallis, donde se evalúa la hipótesis nula de que las medianas dentro de cada una de las 6 columnas es la misma, el Valor-P es igual a 0.0000101184, menor que 0.05, por lo que existe una diferencia estadísticamente significativa entre las medianas de cada una de las columnas, rechazándose la hipótesis.

En la Figura 3.33 se muestra los diagramas de caja y bigote, uno para cada columna de datos. En este caso, hay 14 puntos alejados y 1 punto muy alejado. Según la alineación de la línea central dentro de cada caja, se puede observar la irregularidad que tiene la localización de las medianas en cada gráfico.

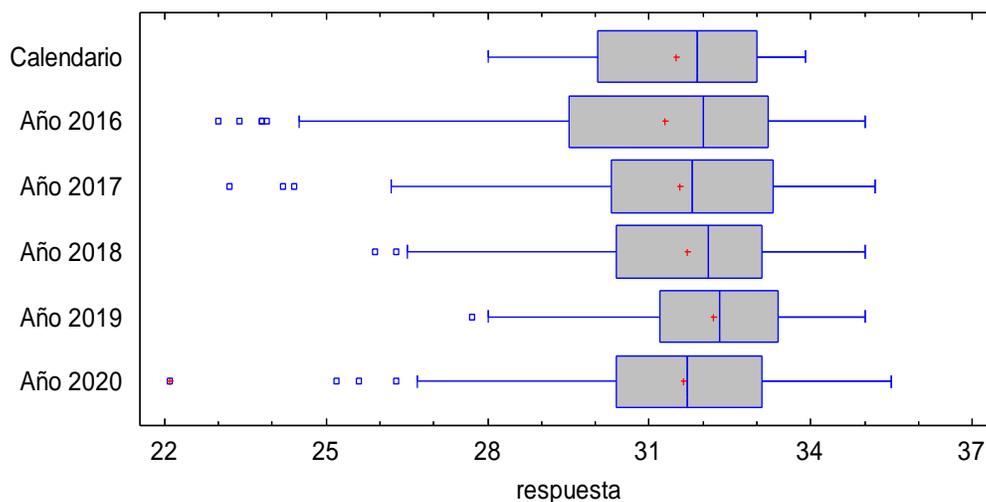


Figura 3.33. Gráfico de caja y bigotes de la temperatura máxima

Al igual que en la temperatura media, el año 2019 no es tratado en el análisis de efectividad del calendario térmico en este período, pues se prevé un año con anomalías, según las salidas del modelo. Según la Tabla 3.30, en los años analizados se aprecia una efectividad en el pronóstico de un 84.4 %. Aunque durante todo el año existen valores que se alejan del rango de ± 2 °C del calendario térmico, los mayores casos de incompatibilidad se observan en el período invernal, sobre todo en los meses de enero y febrero, meses en el que los pronósticos se ven afectados por la entrada de sistemas sinópticos al territorio.

Tabla 3.30. Efectividad del calendario térmico en los años de 2016-2020. Temperatura máxima

Veces en el rango ± 2		
Años	Cantidad	%
2016	309	85.8
2017	313	86.9
2018	286	79.4
2020	308	85.6
Total	1216	84.4

CONCLUSIONES DEL CAPÍTULO

- Al contextualizar la metodología propuesta por Soler y Martín el mejor suavizado se observa en la media móvil cada 11 días en las tres temperaturas.
- Tanto en el período 2011-2015, como en la comparación de los calendarios en meses de un comportamiento normal de la variable, la mediana es el método de mejores resultados y el más efectivo para realizar un calendario cercano a una observación real.
- Una vez compatibilizado el Modelo de Circulación Regional PRECIS CARIBE con la información climática local este es capaz de expresar las probables condiciones futuras de la zona.
- Resulta efectivo y de obligatoriedad la realización de las correcciones, análisis y procesamientos estadísticos de las datas de las variables climáticas, en función de lograr el mayor ajuste y homogeneidad entre los modelos empleados y las series locales.

CONCLUSIONES

La realización de esta investigación permite arribar a las siguientes conclusiones:

- El análisis de los antecedentes teóricos y metodológicos permite identificar las técnicas estadísticas para la elaboración de calendarios climáticos, aplicadas en la conformación de los calendarios térmicos.
- La selección de las metodologías y técnicas estadísticas más apropiadas exponen el método de la media móvil y las medidas de tendencia central como lo más conveniente para la elaboración de un calendario térmico.
- La comparación entre meses de comportamiento normal de la variable temperatura evidencia a la mediana como la técnica estadística con mejor efectividad.
- La comparación entre el calendario térmico propuesto y los datos del Modelo de Circulación Regional PRECIS CARIBE estima pocas variaciones en la variable temperatura en el período de 2016 a 2020.

RECOMENDACIONES

- Aplicar el procedimiento empleado para la elaboración de calendarios térmicos en otras zonas de interés económicos.
- Utilizar las técnicas estadísticas estudiadas para evaluar los cambios del clima a partir de los diferentes servicios meteorológicos que se ofrecen en el Sistema Nacional, logrando una mayor confiabilidad en los cálculos proporcionados.

BIBLIOGRAFÍA

- Barón, L. (1998). *Bioestadística: Métodos y Aplicaciones*. Universidad de Málaga. España.
- Batanero, C., & Díaz, C. (2008). *Análisis de datos con STATGRAPHICS*. Departamento de Didáctica de la Matemática, Facultad de Ciencias de la Educación, Universidad de Granada.
- Best, J. (2001). *Damned Lies and Statistics: Untangling Numbers from the Media, Politicians, and Activists*. *University of California Press*.
- Bloch, D. (1966). A note on the estimation of the location parameters of the Cauchy distribution. *Journal of the American Statistical Association*, (61)
- CECLIM. (2016). Boletín de Vigilancia del Clima. Enero 2010-Junio 2016. *Centro del Clima. Instituto de Meteorología. Ministerio de Ciencia, Tecnología y Medio Ambiente. República de Cuba*.
- Chan, C. (2009). *Examen profesional para obtener el grado de Licenciado en Enseñanza de las Matemáticas*. Facultad De Matemáticas, Universidad Autónoma de Yucatán.
- CITMA. (2011 a). *Estrategia Ambiental Provincial 2011/2015*. Delegación Territorial del Ministerio de Ciencia, Tecnología y Medio Ambiente, Ciego de Ávila.
- CITMA. (2011 b). *Estudios de Peligro, Vulnerabilidad y Riesgos. Inundaciones por intensas lluvias. Afectaciones por fuertes vientos. Municipio Ciego de Ávila*. Delegación Territorial del Ministerio de Ciencia, Tecnología y Medio Ambiente, Ciego de Ávila.
- CITMA. (2011 c). *Estudios de Peligro, Vulnerabilidad y Riesgos. Inundaciones por intensas lluvias. Afectaciones por fuertes vientos. Municipio Venezuela*. Delegación Territorial del Ministerio de Ciencia, Tecnología y Medio Ambiente, Ciego de Ávila.
- CITMA. (2013). *Estrategia de Ciencia, Tecnología, Innovación y Medio Ambiente del municipio Venezuela*. Centro de Creación de Capacidades del CITMA Venezuela.
- Conde, A., & Gay, C. (2008). Guía para la generación de escenarios de Cambio Climático a escala regional. *Centro de Ciencias de la Atmósfera. Primera versión*.
- Dante, A. (2002). *Inferencia Estadística Básica*. Universidad Nacional de Educación a Distancia, España.
- De Aragón, H. (2013). Estadística para describir y predecir. *Milenio*.

- García, R. (2002). *Curso Básico de STATGRAPHICS Plus*. Universidad Complutense de Madrid.
- Gu, X., & Jiang, J. (2005). A complex autoregressive model and application to monthly temperature forecasts. *Annales Geophysicae*, 23, 3229 - 3235.
- Gutiérrez, J. M., Cano, R., S., C. A., & M., S. C. (2004). Redes Probabilísticas y Neuronales en las Redes Atmosféricas. *Universidad de Cantabria, Instituto Nacional de Meteorología, España*.
- Hernández, A., López, A., Córdova, O. L., & Alonso, J. D. (2009). Climatología de la Temperatura en Ciego de Ávila. Análisis comparativo en dos de sus zonas. Período 1970 – 2008. *V Congreso Cubano de Meteorología. La Habana. Sociedad Meteorológica de Cuba*.
- Illescas, M. (2015). *¿Cuál es la media móvil ideal?* . Recuperado de <http://compraraccionesdebolsa.com>
- Instituto de Geografía. (1989). *Nuevo Atlas Nacional de Cuba*. Editora Cartográfica.
- IPCC. (2000). *Informe especial del IPCC. Escenarios de emisiones*. Grupo Intergubernamental de Expertos sobre el Cambio Climático. .
- IPCC. (2001). *Cambio Climático: Informe de Síntesis. Contribución de los Grupos de Trabajo I, II y III al Tercer Informe de Evaluación del Grupo Intergubernamental de Expertos sobre el Cambio Climático*. Grupo Intergubernamental de Expertos sobre el Cambio Climático.
- IPCC. (2007). *Cambio climático: Informe de Síntesis. Contribución de los grupos de Trabajo I, II y III al Cuarto Informe de Evaluación del Grupo Intergubernamental de Expertos sobre Cambio Climático*. Suiza: Grupo Intergubernamental de Expertos sobre el Cambio Climático.
- Jones, R., Noguera, M., Hasell, D., Wilson, S., Jenkins, G., & Mitchell, J. (2004). *Generating High Resolution Climate Scenarios using PRECIS Met Office*. Hadley Center. United Kingdom.
- Lecha, L., Paz, L., & Lapinel, B. (1994). El clima de Cuba. *Editorial Academia*, 65 – 82.
- León, G. (2005). Verificación de los modelos meteorológicos. *Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), Bogotá*.

- Martín, J. (2003). El tiempo y el clima. *Rubes Editorial*, 69-92.
- Martín, J., & Esteban, P. (1998). *Metodología para la construcción de un calendario de nubosidad. El caso de Barcelona. Tecnología geográfica para el siglo XXI*. Asociación de Geógrafos Españoles y Universidad Autónoma de Barcelona.
- Martín, M. (2009). *Calendarios pluviométricos para la provincia de Ciego de Ávila, Cuba*. (Tesis presentada en opción del Título Académico de Máster en Meteorología), Instituto Superior de Tecnologías y Ciencias Aplicadas, La Habana.
- Martín, M., & Matos, F. (2009). Calendario pluviométrico para la estación meteorológica de Júcaro. *Revista del Aficionado a la Meteorología*.
- Matos, F., & Batista, L. (2005). Calendario pluviométrico para la zona litoral de Cayo Coco. *Memorias de la II Conferencia Internacional de Manejo Integrado de Zonas Costeras CARICOSTAS 2005*.
- Mayén, S. (2009). *Comprensión de las medidas de tendencia central por estudiantes mexicanos de Educación Secundaria y Bachillerato*. (Tesis doctoral). Universidad de Granada, España.
- Mirabal, M., Robaina, M., & Uranga, R. (2010). R: una herramienta poco difundida y muy útil para la investigación clínica. *Revista Cubana de Investigaciones Biomédicas*, 29. Recuperado de http://www.bvs.sld.cu/revistas/ibi/vol29_2_10/ibi13210.htm
- Novaes, D. (2011). *Concepções de professores da Educação Básica sobre variabilidade estatística*. (Tesis doctoral). Pontifícia Universidade Católica de São Paulo, Brasil.
- Palma, B. E., Conde, C., Morales, E. R., & Colorado, G. (2009). Escenarios climáticos.
- Pérez, S. P., Sierra, E. M., Peschiera, J. M., & Escalante, D. (2008). Calendario pluviométrico en el margen oriental del área agrícola argentina. *Revista de Climatología*, 8, 41-49.
- Prensa Latina. (2013). *Descubren calendario más antiguo del mundo*. Recuperado de <http://www.prensa-latina.cu>
- Real Academia Española. (2014). *Calendario*. Madrid, España: Diccionario de la lengua española.
- Rothenberg, T., Fisher, F., & Tilanus, C. (1966). A note on estimation from a Cauchy sample. *Journal of the American Statistical Association*, 59, 460-463.

- Ruiz, J. (2007). Escenarios de Cambio Climático, algunos modelos y resultados de lluvia para Colombia bajo el escenario A1B. *Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM. Subdirección de Meteorología. Bogotá, D.C.*
- Ruiz, J. (2007). Escenarios de Cambio Climático, algunos modelos y resultados de lluvia para Colombia bajo el escenario A1B. *Instituto de Hidrología, Meteorología y Estudios Ambientales – IDEAM. Subdirección de Meteorología. Bogotá, D.C.*
- Salazar, W. (2015). *Idoneidad de las tareas sobre media aritmética en textos de primer grado de Educación Secundaria*. (Tesis para optar el grado académico de Magister en Enseñanza de las Matemáticas), Pontificia Universidad Católica del Perú, Perú.
- Salvador, M. D. (2003). *Anova un factor y Kruskal-Wallis*. Universidad de Castilla-La Mancha.
- Sánchez, J. (2012). Capítulo 4: Series temporales. En *Introducción a la Estadística Empresarial*.
- Sayritupac, J. (2013). *Significados de las medidas de tendencia central. Un estudio con alumnos universitarios de Carreras de Humanidades*. (Tesis para optar el grado académico de Magister en Enseñanza de las Matemáticas), Pontificia Universidad Católica del Perú, Perú.
- SINAVEF. (2009). El cambio climático y su influencia en las plagas agrícolas. *Sistema Nacional de Vigilancia Epidemiológica Fitosanitaria*.
- Soler, X., & Martín, J. (2002). Los calendarios climáticos. Una propuesta metodológica. *Memorias del III Congreso de la Asociación Española de Climatología*, 577-585.
- Sorí, R., Hernández, A., López, A., Benedico, O., Córdova, O. L., & Ávila, M. (2014). Pronóstico climático mensual de la temperatura en Ciego de Ávila. Cuba. *Ciencias de la Tierra y el Espacio*, 15(1), 23-35.
- Strauss, S., & Bichler, E. (1988). The development of children's concepts of the arithmetic average. *Journal for Research in Mathematics Education*, 19(1), 64-80.
- Terrádez, M., & Juan, A. (2002). *Series temporales. Proyecto e-Math. Financiado por la Secretaría de Estado de Educación y Universidades (MECD), España*. Recupeardo de <http://www.uoc.edu>
- UNC. (2017). *Bioestadística. Análisis de varianza*. (Universidad Nacional de Córdoba, Editor) Aulas virtuales. Recupeardo de <http://www.fca.proed.unc.edu.ar>

ANEXOS

ANEXO 1: ENTREVISTA A PRODUCTORES Y DIRECTIVOS DE LA UNIDAD EMPRESARIAL DE BASE (UEB) PRODUCCIÓN DE PIÑA DE LA EMPRESA AGROINDUSTRIAL CEBALLOS

Objetivo: Conocer las características de la entidad y el uso de servicios meteorológicos por parte de los trabajadores de la UEB.

Preguntas:

1. ¿Cuál es la planificación del cultivo de la piña por parte de la empresa?
2. ¿Cuál es la situación laboral que presenta la UEB? ¿Es personal capacitado?
3. ¿Cuál es el plan de inversión para los próximos años dentro de los campos de cultivo y en sus alrededores?
4. ¿Utilizan algún servicio meteorológico para el desarrollo fenológico y fitosanitario en el ciclo de cultivo de la piña? ¿Cuál?
5. ¿A qué variables meteorológicas recurren con más frecuencia para la toma de decisiones? ¿Por qué?
6. ¿Son acertados los pronósticos realizados por el Centro Meteorológico Provincial? ¿Con que frecuencia aciertan?
7. Si se tuviera un calendario térmico para el tratamiento de los cultivos en las distintas fases, ¿qué nivel de importancia tendría?

ANEXO 2: GUÍA PARA EL ANÁLISIS DOCUMENTAL

Objetivo: Analizar las bases de datos de la estación meteorológica 78346 del municipio Venezuela.

Indicadores:

1. Períodos recogidos en la base de datos.
2. Años de interés para el estudio, que tengan significación climática y según esté regulado por la Organización Meteorológica Mundial.
3. Variables con que se desea trabajar.
4. Análisis exploratorio de datos para verificar la existencia y/o faltantes de valores en las observaciones.
5. Comportamiento de la variable de interés.
6. Boletines climáticos realizados a partir de la base de datos.
7. Métodos estadísticos aplicados para el control de la variable.