

**SUPERIOR POLITÉCNICA DEL LITORAL**

**Facultad de Ciencias Naturales y Matemáticas**



**PROYECTO INTEGRADOR**

**“CONTROL DE CALIDAD DATOS METEREOLÓGICOS A LA  
ESTACIÓN AUTOMÁTICA FACULTAD DE CIENCIAS  
NATURALES UNIVERSIDAD DE GUAYAQUIL”**

Previo la obtención del Título de:

**INGENIERO EN ESTADÍSTICA INFORMÁTICA**

Presentado por:

**Carlos Isidro Reina Campuzano**

**GUAYAQUIL - ECUADOR**

**Año: 2019**

## **AGRADECIMIENTOS**

A Dios por darme vida, salud y la oportunidad de poder prepararme profesionalmente, mostrándome siempre el camino del éxito y dándome la perseverancia necesaria para alcanzar este objetivo.

A mis padres

Carl Kenneth Reina Murillo

María Campuzano Gurumendi

por siempre apoyarme e impulsar mi desarrollo personal y profesional.

A mi abuela materna y segunda madre

Abigail Gurumendi

por siempre incentivar y motivarme.

A mi maestra y tutora

PhD. Sandra García

por compartirme sus conocimientos y brindarme la mejor orientación.

Al Ing. Gabriel Mantilla

por brindarme el soporte necesario para la elaboración de este estudio.

Al INAMHI por la confianza depositada en mí y por facilitar toda la información necesaria para el desarrollo de este trabajo.

## DEDICATORIA

A mis padres

Carl Kenneth Reina Murillo

María Campuzano Gurumendi

por haber forjado desde siempre mi carácter inculcando en mí siempre el trabajo duro y la perseverancia, por siempre creer en mí y ser los pilares fundamentales en mi vida, fuente de la máxima expresión de amor conocida y apoyo incondicional.

A mi abuela materna y segunda madre Abigail Gurumendi por siempre incentivarme y apoyarme.

## DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, me corresponden conforme al reglamento de propiedad intelectual de la institución; Yo Carlos Isidro Reina Campuzano doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

*Carlos Reina C.*

---

Carlos Isidro Reina Campuzano

## EVALUADORES



**PhD. Sandra Lorena García Bustos**

PROFESOR DE LA MATERIA



**PhD. Sandra Lorena García Bustos**

PROFESOR TUTOR

## RESUMEN

Este trabajo trata sobre el control de calidad de los datos de varias variables atmosféricas, las mismas que se mencionan a continuación: Precipitación, Temperatura del Aire, Humedad Relativa, Presión Atmosférica, Radiación Solar, Temperatura del Suelo, Dirección del Viento, Velocidad del Viento.

Los datos fueron obtenidos por la estación meteorológica automática del INAMHI ubicada en Facultad de Ciencias Naturales de la Universidad de Guayaquil.

Para el análisis de la calidad de los datos se hizo uso de métodos estadísticos básicos y avanzados.

Entre los métodos estadísticos básicos que se utilizó se encuentran análisis de mínimos, máximos y promedio; mientras que los métodos estadísticos avanzados de los que se ha hecho uso son modelos de aproximación de máximos y mínimos, gráficos de control, pruebas de bondad de ajuste para los máximos y mínimos.

Como resultado se obtuvo la eliminación de valores sospechosos lo que permitió obtener datos confiables y verificados.

**Palabras Clave:** Control de calidad, datos meteorológicos, temperatura del aire, humedad relativa, presión atmosférica, dirección del viento, velocidad del viento, valores sospechosos.

## **ABSTRACT**

*This work is about the quality control of the data of several atmospheric variables, the same ones mentioned below: air temperature, relative humidity, atmospheric pressure, wind direction, wind speed.*

*The data were obtained by the automatic meteorological station of INAMHI located in the Faculty of Natural Sciences of the University of Guayaquil.*

*For the analysis of data quality, basic and advanced statistical methods were used.*

*Among the basic statistical methods that were used are minimum, maximum and average analyzes; while the advanced statistical methods used are maximum and minimum approximation models, control charts, goodness of fit tests for the maximum and minimum.*

*As a result, the elimination of suspicious values was obtained, which allowed obtaining reliable and verified data.*

*Keywords: Quality control, weather data, air temperature, relative humidity, atmospheric pressure, wind direction, wind speed, suspicious values.*

# ÍNDICE GENERAL

ÍNDICE GENERAL .....	8
ABREVIATURAS.....	10
SIMBOLOGÍA.....	11
ÍNDICE DE FIGURAS .....	12
ÍNDICE DE TABLAS .....	13
CAPÍTULO 1 .....	15
1. INTRODUCCIÓN.....	15
1.1 Descripción del problema.....	15
1.2 Justificación del problema.....	16
1.3 Alcance del Proyecto.....	16
1.4 Objetivos .....	17
1.4.1 Objetivo General.....	17
Objetivos Específicos.....	17
1.5 Marco Teórico .....	18
1.5.1 Conceptos relacionados al estudio de la Meteorología .....	18
1.5.2 Control Estadístico de Procesos.....	22
1.5.3 Otros estudios relacionados .....	25
CAPÍTULO 2 .....	26
2. METODOLOGÍA:.....	26
2.1 Enfoque de la investigación .....	26
2.1.1 Tipo de investigación .....	26
2.1.2 Obtención de Datos.....	27
2.2 Herramientas Estadísticas .....	28
2.2.1 Estadísticos .....	28



2.2.2	Estadísticos de Orden .....	31
2.2.3	Valores Extremos .....	32
2.2.4	Fases de un Gráfico de Control .....	40
2.2.5	Gráfico de Control Multivariado $T^2$ de Hotelling .....	41
2.2.6	Longitud Promedio de Corrida (ARL).....	43
CAPÍTULO 3 .....		45
3.	RESULTADOS Y ANÁLISIS:.....	45
3.1	Análisis Especializado.....	45
3.1.1	Eliminación de valores sospechosos.....	45
3.1.2	Fase 1 (Corrida 1).....	47
3.1.3	Fase 1 (Corrida 2).....	53
3.2	Análisis Descriptivo .....	59
3.3	Comportamiento de las Variables en el Tiempo.....	62
3.3.1	Análisis de las Correlaciones de las Variables .....	66
3.4	Análisis de Rendimiento para Gráficos de Control.....	72
3.6	Discusión y Análisis Final.....	79
CAPÍTULO 4 .....		82
4.	CONCLUSIONES Y RECOMENDACIONES.....	82
4.1	Conclusiones.....	82
4.2	Recomendaciones.....	83
5.	Bibliografía.....	84

## ABREVIATURAS

INAMHI	=	Instituto Nacional de Hidrología y Meteorología
SENAGUA	=	Secretaría del Agua
EMAs	=	Estaciones Meteorológicas Automáticas
EMCs	=	Estaciones Meteorológicas Convencionales
UG	=	Universidad de Guayaquil
CCNN	=	Ciencias Naturales
RR	=	Precipitación
TA	=	Temperatura del Aire
RH	=	Humedad Relativa
PA	=	Presión Atmosférica
SR_GLOBAL	=	Radiación Solar
TG	=	Temperatura del Suelo
WIND_DIR	=	Dirección del Viento
WIND_SP	=	Velocidad del Viento
CEP	=	Control Estadístico de Procesos <i>(Statistical Process Control)</i>
TVE	=	Teoría de Valores Extremos
GEV	=	Distribución Valores Extremos Generalizada <i>(Generalized Extreme Value)</i>
IC	=	Intervalo de Confianza <i>(Maximum Likelihood Estimation)</i>
ARL	=	Average Run Length <i>(Longitud Promedio de Corrida)</i>

## SIMBOLOGÍA

$\mu$	Parámetro de Localización
$\sigma$	Parámetro de Escala
$\xi$	Parámetro de Forma
$\alpha$	Probabilidad de ocurrencia del error Tipo I
$\beta$	Probabilidad de ocurrencia del error Tipo II

## ÍNDICE DE FIGURAS

Figura 1.1 Estación Meteorológica Automática .....	22
Figura 1.2 Gráfico de Control .....	24
Figura 2.1 Máximos.....	29
Figura 2.2 Mínimos.....	30
Figura 3.1 Gráficos de Control TA Máximas Fase 1 (Corrida 1).....	47
Figura 3.2 Gráficos de Control TA Mínimas Fase 1 (Corrida 1) .....	49
Figura 3.3 Gráficos de Control T2 Hotelling Fase 1 (Corrida 1) .....	51
Figura 3.4 Gráficos de Control TA Máximas Fase 1 (Corrida 2).....	53
Figura 3.5 Gráficos de Control TA Mínimas Fase 1 (Corrida 2) .....	55
Figura 3.6 Gráficos de Control T2 Hotelling Fase 1 (Corrida 2) .....	57
Figura 3.7 Histograma de Frecuencia del Promedio Horario.....	60
Figura 3.8 Variación en el Tiempo para Máximos .....	62
Figura 3.9 Variación en el Tiempo para Mínimos .....	63
Figura 3.10 Variación en el Tiempo Promedio Original .....	64
Figura 3.11 Variación en el Tiempo Promedio Depurado.....	65
Figura 3.12 Correlaciones entre los Promedios de las Máximas Mensuales ..	66
Figura 3.13 Correlaciones entre los Promedios de las Mínimas Mensuales ....	68
Figura 3.14 Correlaciones entre los Promedios de las Medias Mensuales, RR acumulado.....	70

## ÍNDICE DE TABLAS

Tabla 2.1 Propiedades Distribución de Gumbel .....	35
Tabla 2.2 Propiedades Distribución de Fréchet.....	36
Tabla 2.3 Propiedades de la distribución de Weibull.....	37
Tabla 2.4 Propiedades Distribución GEV .....	39
Tabla 3.1 Parámetros del Gráfico de Control para las Temperaturas Máximas del Aire Fase 1 (Corrida 1).....	48
Tabla 3.2 Parámetros del Gráfico de Control para las Temperaturas Mínimas del Aire Fase 1 (Corrida 1).....	50
Tabla 3.3 Bandas de Confianza para T2 de Hotelling Análisis Multivariado del Promedio Fase 1 (Corrida 1).....	52
Tabla 3.4 Parámetros del Gráfico de Control para las Temperaturas Máximas del Aire Fase 1 (Corrida 2).....	54
Tabla 3.5 Parámetros del Gráfico de Control para las Temperaturas Mínimas del Aire Fase 1 (Corrida 2).....	56
Tabla 3.6 Bandas de Confianza para T2 de Hotelling Análisis Multivariado del Promedio Fase 1 (Corrida 2).....	58
Tabla 3.7 Valores p Test de Shapiro .....	61
Tabla 3.8 Análisis de Rendimientos para los Gráficos de Control de Máximos (Enero-Abril).....	73
Tabla 3.9 Análisis de Rendimientos para los Gráficos de Control de Máximos (Mayo-Agosto).....	74
Tabla 3.10 Análisis de Rendimientos para los Gráficos de Control de Máximos (Septiembre-Diciembre) .....	75
Tabla 3.11 Análisis de Rendimientos para los Gráficos de Control de Mínimos (Enero-Abril).....	76
Tabla 3.12 Análisis de Rendimientos para los Gráficos de Control de Mínimos (Mayo-Agosto).....	77
Tabla 3.13 Análisis de Rendimientos para los Gráficos de Control de Mínimos (Septiembre-Diciembre) .....	78
Tabla 3.14 Valores que exceden los Límites de Control en la Corrida 1 .....	79

Tabla 3.15 Valores que exceden los Límites de Control en la Corrida 2 ..... 80

# CAPÍTULO 1

## 1. INTRODUCCIÓN

Guayaquil es una de las ciudades más grandes del Ecuador, siendo la segunda más poblada del país, es de clima tropical y está ubicada en las costas del Océano Pacífico, sólo cuenta con dos estaciones climáticas en el año que son: lluviosa (mal llamada invierno) comprendida entre los meses de Diciembre a Mayo, y seca (mal llamada verano) que está comprendida entre los meses de Junio a Noviembre, debido a que se encuentra en el Litoral Ecuatoriano sus dos estaciones climáticas poseen grandes diferencias en cuanto a sus características. La estación lluviosa está marcada por una fuerte radiación solar, altas temperaturas, humedad y precipitaciones; mientras que la estación seca es más agradable en cuanto a sus características ya que la humedad disminuye considerablemente y las temperaturas descienden, así mismo la ausencia de precipitaciones en esta estación la hacen mucho más cómoda para los habitantes de esta ciudad. (Guayaquil es mi Destino ®, 2019)

### 1.1 Descripción del problema.

El Instituto Nacional de Meteorología e Hidrología del Ecuador (INAMHI), cuenta con un gran número de Estaciones Meteorológicas Automáticas (EMAs), que están registrando información en tiempo real. Los datos recopilados por las estaciones meteorológicas automáticas registran información de varias variables atmosféricas, las mismas que se encuentran en frecuencia minutal, cada cinco minutos y horaria.

La red de las EMAs proporciona el registro continuo de las variables atmosféricas a medir y además complementa la información registrada por la red de Estaciones Meteorológicas Convencionales (EMCs) que mantiene la institución. Sin embargo toda la información registrada no puede ser depurada, no existe control de calidad de datos obtenidos en dichas estaciones y por lo tanto no es entregada al usuario final de un modo completo.

## 1.2 Justificación del problema.

La importancia de este estudio radica en la necesidad de implementar una metodología que estandarice los datos de las lecturas provenientes de las EMAs, para poder hacer el uso correcto de la información y entregar a los solicitantes una data útil con la que se pueda realizar análisis complejos. Cabe señalar que, una de las políticas internas del INAMHI es la de proporcionar información veraz, validada y confiable al cliente. Por lo cual, la información meteorológica provista de las EMAs, debe tener la calidad necesaria para cumplir con lo establecido.

La alianza estratégica entre ESPOL, Secretaría Nacional del Agua (SENAGUA) y el Instituto Nacional de Meteorología e Hidrología (INAMHI) de plantear el desarrollo de este proyecto de investigación, para depurar toda la base de datos hidrológica y meteorológica de EMAs que mantiene INAMHI es de suma importancia, ya que este trabajo se enfoca en poder implementar en base a herramientas y técnicas estadísticas procedimientos que ayuden a llevar a cabo el control de calidad de estos datos, de manera que puedan ser un producto validado y verificado, el cual luego pueda ser utilizado en diferentes análisis.

## 1.3 Alcance del Proyecto

El presente estudio se realiza en:

- **Espacio:** Las observaciones tomadas por la EMA ubicada en la Facultad de CCNN de la UG representan un radio de 30 kilómetros alrededor de la estación, cuya descripción se detalla a continuación:

Código: M1271

Provincia: Guayas

Propietario: INAMHI

Latitud: -2.146500

Longitud: -79.917300

Altitud: 21.00 metros

Tipo: Meteorológica

Estado: Operativa



- **Tiempo:** Periodo comprendido entre los años 2017-2019
- **Población:** Observaciones meteorológicas registradas en la EMA ubicada en la Facultad de CCNN.

Este estudio presentara el análisis de calidad y la validación de los datos de la EMA, ubicada en la Facultad de CCNN de la UG. Que de igual manera podrá ser replicado en cualquier estación meteorológica automática, utilizando la metodología implementada.

## **1.4 Objetivos**

### **1.4.1 Objetivo General**

Implementar una metodología estadística, realizando control de calidad a los datos, registrados en la estación meteorológica automática. La cual expondrá, umbrales estadísticos para controlar las variables atmosféricas que están siendo monitoreadas cada minuto.

### **Objetivos Específicos**

- Generar umbrales de confianza que permitan discriminar valores sospechosos, haciendo uso del mínimo y el máximo absolutos de cada hora, obtenidos de los estadísticos mínimos y máximos de cada variable, modelando la distribución del mínimo y la del máximo.
- Discriminar valores sospechosos con los promedios de cada hora, obtenidos de los estadísticos medios de cada variable, estudiando los gráficos de control multivariados.
- Obtener información validada, verificada y confiable.

## **1.5 Marco Teórico**

### **1.5.1 Conceptos relacionados al estudio de la Meteorología**

#### **1.5.1.1 Meteorología**

Es una ciencia que describe el comportamiento de la física atmosférica, los fenómenos que en ella se producen y cómo se producen los estados del tiempo.

La palabra meteorología proviene del griego “meteoros” cuyo significado es alto y “logos” que significa tratado.

Los Fenómenos atmosféricos están clasificados de la siguiente manera:

- Aéreos, por ejemplo de este tipo de fenómenos podemos mencionar el viento.
- Acuosos, por ejemplo garuas, lluvia o cualquier tipo de precipitaciones.
- Luminosos, donde el más conocido de estos fenómenos es el arcoíris.
- Eléctricos, ejemplo de este tipo de fenómenos son los relámpagos y los rayos.

Algunos de los factores más importantes dentro de la meteorología para poder realizar predicciones son: presión, temperatura y humedad. (Aprendiendo Meteorología, 2001)

### 1.5.1.2. Variables Atmosféricas

Las variables meteorológicas, registran secuencialmente el estado del tiempo y de la atmosfera, según el lugar y el momento donde se presentan (**National Weather Services, s.f.**).

Entre estas variables podemos mencionar: TA, RR, SR, PA, HR, TG, WIND\_SP y WIND\_DIRECTION.

- **Precipitación (RR)**, es la caída de agua en cualquiera de sus estados hacia la superficie terrestre. Esto puede ser lluvia, granizo, nieve, etc., se mide en milímetros (mm) (Wikipedia, 2019)
- **Temperatura del Aire (TA)** o también llamada temperatura atmosférica, esta medida en Celsius (°C) y representa el nivel de calor del aire en un lugar y momento específicos. (Pérez Porto & Gardey, 2016)
- **Humedad Relativa (RH)**, definida como la relación existente presión parcial de vapor de agua y presión de equilibrio de agua y una temperatura específica, se mide en porcentaje (%). (Steven M. Babin, 2016)
- **Presión Atmosférica (PA)**, es considerada una fuerza por cada unidad de superficie ejercida por el aire que compone la atmósfera sobre la superficie terrestre. En otras palabras la presión atmosférica es una fuerza que ejerce la atmósfera sobre todo objeto q se encuentra en su interior, la unidad de medida es el hectopascal (hPa). (Pérez Porto & Merino, Definicion.de, 2012)
- **Radiación Global (SR\_GLOBAL)**, La radiación global se define como la radiación solar recibida de un ángulo sólido de  $2\pi$  estereorradianes sobre una superficie horizontal. La radiación global incluye la recibida directamente del disco solar y también la radiación celeste difusa dispersada al atravesar la

atmósfera, se mide en vatios por metro cuadrado ( $W/m^2$ ).  
(Agencia Estatal de Meteorología)

- **Radiación Solar (SR)**, es un conjunto de energía electromagnética emitida por el sol en forma de radiación, se mide en vatios por metro cuadrado ( $W/m^2$ ). (Méndez Muñiz, Javier María; Cuervo García, Rafael; Bureau Veritas Formación, 2010)
- **Temperatura del Suelo (TG)**, esta medida en Celsius, simboliza el nivel de calor del suelo en un lugar y momento específicos, se mide en ( $^{\circ}C$ ).
- **Velocidad del Viento (WIND\_SPD)**, es la velocidad con la que el aire se mueve dentro de la atmósfera terrestre, se mide en (m/s) para lo que se utilizan instrumentos de medición denominados anemómetros. (Hamel, 2018)
- **Dirección del Viento (WIND\_DIR)**, esta variable atmosférica es calculada en grados ( $^{\circ}$ ) en dirección a las agujas del reloj desde el norte auténtico y se define como la dirección desde la cual sopla el viento.

### 1.5.1.3 Estación Meteorológica

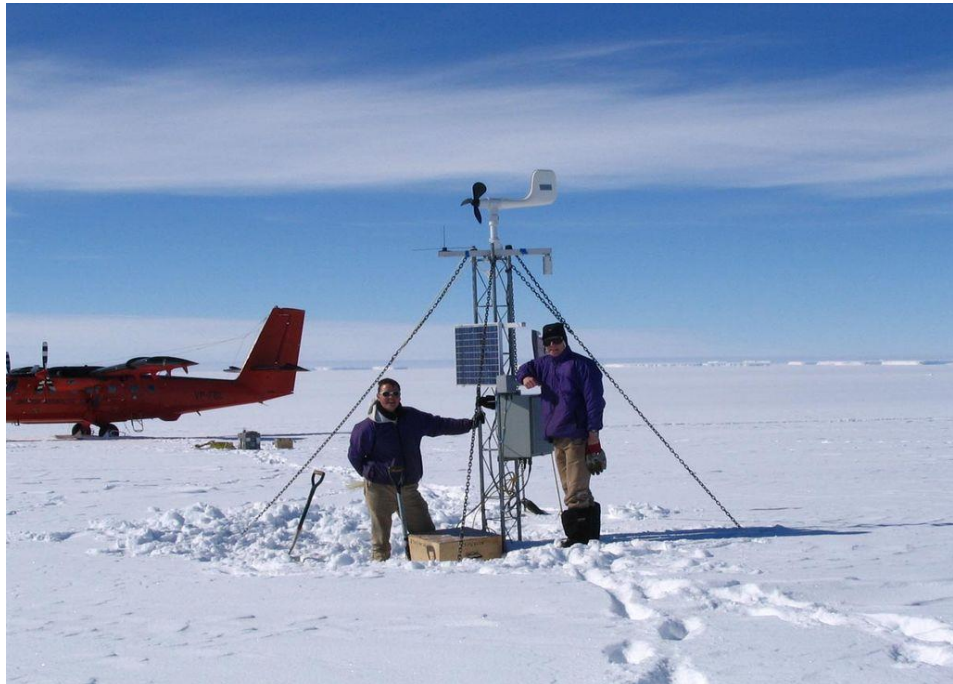
Una estación meteorológica es una instalación destinada a medir y registrar regularmente diversas variables meteorológicas. Estos datos se utilizan tanto para la elaboración de pronósticos climatológicos, predicciones en tiempo real, estado de la atmósfera, y eventos extremos que pueden ser estimados a partir de modelos estadísticos y numéricos como para estudios climáticos. (Wikipedia®, 2019)

#### **1.5.1.4 Estación Meteorológica Automática**

Para (Archive Internet) una estación meteorológica automática (AWS) es un tipo de estación meteorológica que ha sido automatizada de la estación meteorológica convencional, ya sea para permitir mediciones desde áreas remotas o para ahorrar mano de obra humana. El sistema puede informar de varias maneras diferentes. Puede ser en tiempo real a través de un enlace local a un sistema informático o de telecomunicaciones de sistemas de satélite, también utiliza la tecnología móvil GSM. Una alternativa es el almacenamiento de la información en los repositorios informáticos locales, como un dispositivo extraíble, para la recuperación en una etapa posterior. La mayoría de las estaciones meteorológicas automáticas tienen termómetro para medir la temperatura, anemómetro para medir la velocidad del viento, higrómetro para medir la humedad, barómetro para medir la presión. Algunos de ellos incluso tienen pluviómetro para medir la lluvia, ceilómetros para medir la altura de la nube, sensor meteorológico actual o sensor de visibilidad.

A pesar de los continuos avances en tecnología, las estaciones meteorológicas automáticas no pueden informar la clase y la cantidad de nubes. Además, se sabe que las mediciones de precipitación son algo problemáticas, especialmente para la nieve, ya que el medidor debe vaciarse entre lecturas. Para el clima actual, todos los fenómenos que no pasan dentro del pequeño rango de muestra del sensor, como parches de niebla, permanecen sin ser observados.

Las primeras estaciones meteorológicas automáticas iniciales debían ubicarse donde había disponibilidad de líneas eléctricas y de comunicación. En aplicaciones más modernas, la tecnología emplea paneles solares, turbinas eólicas y teléfonos móviles que han permitido tener AWS inalámbricos, que no están conectados a la red eléctrica o la red de telecomunicaciones.



**Figura 1.1 Estación Meteorológica Automática**

Fuente: De William M. Connolley, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=2888729>

### **1.5.2 Control Estadístico de Procesos**

El control estadístico de procesos CEP o SPC por sus siglas en inglés, tiene su origen en los años 20 de la mano de su creador Walter A. Shewhart quien lo utilizó en Bell Telephone Company con el fin de disminuir el porcentaje de defectos dentro de la empresa.

Walter A. Shewhart es considerado el padre el CEP debido a que fue el primero en desarrollar métodos estadísticos para realizar estudios sobre la calidad. En su libro “Economic Control of Quality of Manufactured Products”

(1931) marcó la pauta que seguirían otros discípulos distinguidos (Joseph Juran, W.E. Deming, etc.). (Ruiz-Falcó Rojas , 2006)

SPC no se trata realmente de estadísticas o control, se trata de competitividad.

Las organizaciones, cualquiera sea su naturaleza, compiten en tres temas: calidad, entrega y precio. No puede haber muchas personas en el mundo que sigan convencidas de que la reputación atribuida a una organización por la calidad de sus productos y servicios es clave para su éxito y el futuro de sus empleados.

Además, si la calidad es correcta, las posibilidades son la entrega y el precio. El rendimiento también será competitivo. (Oakland, 2003)

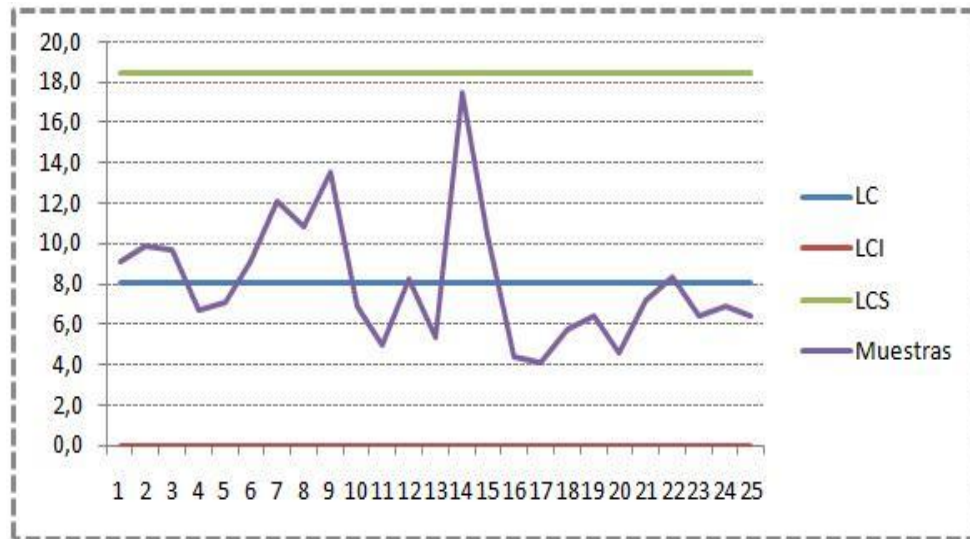
#### **1.5.2.1 Gráficos de Control**

Los gráficos de control también llamados cartas de control, son herramientas de calidad que se utilizan para monitorear y controlar el equilibrio de los procesos, a través del seguimiento de las propiedades de calidad, las cuales permiten determinar posibles causas variables en los datos.

Para poder realizar un gráfico de control se lo hace mediante un diagrama de líneas, y los umbrales o límites de control superior (LCS) e inferior (LCI) se obtienen en base a los datos, la mayor cantidad de datos, objeto de estudio, oscilarán entre estos límites.

Los umbrales obtenidos, sirven para verificar el correcto estado del proceso y determinar las observaciones que salen de los límites de control, con esto es posible controlar el proceso una vez que esos valores sean eliminados.

El gráfico de control es una herramienta grafica que permite monitorear el desarrollo del proceso en el tiempo, lo que brinda la posibilidad de observar si existen variaciones puntuales, o en su defecto si son continuas. Determinando de esta manera que existe algún desajuste dentro del proceso que deberá ser solucionado. (Cuatrecasas Arbós & González Babón, 2017)



**Figura 1.2 Gráfico de Control**

Fuente: [https://www.gestiondecalidadtotal.com/grafica\\_de\\_rangos.html](https://www.gestiondecalidadtotal.com/grafica_de_rangos.html)



### **1.5.3 Otros estudios relacionados**

(Araya López, 2010) presentó un estudio de calidad de datos meteorológicos de las variables, temperatura y humedad relativa, que se realizó en el Instituto Meteorológico Nacional (IMN) de Costa Rica, cuyo objetivo fue determinar la calidad de los datos que proporciona la red de estaciones meteorológicas.

(Veiga, Herrera , Skansi, & Podestá, 2014) desarrollaron un esquema de control de calidad de los datos mediante el cual se depura datos sospechosos provenientes del error humano, ya sea por observación, digitalización o digitación.

(Cruz-Saco Vilela, Natividad Alvarado, & Espinoza Paredes) desarrollaron un software para el tratamiento de grandes volúmenes de datos haciendo uso de pruebas estadísticas para ubicar datos sospechosos provenientes de datos errados.

(Varela Ledesma, Díaz Pereira, & Pomares, 2015) realizaron un estudio de calidad de datos de temperatura y humedad relativa, con datos obtenidos de las estaciones meteorológicas Sancti Spíritus y El Jíbaro, estableciendo una metodología que permita reconstruir series mensuales de temperatura seca y humedad relativa.

# CAPÍTULO 2

## 2. METODOLOGÍA:

Para el estudio que se desea implementar en el control de calidad inicialmente se realizó un análisis descriptivo de los datos proporcionados por la estación meteorológica, estimando el comportamiento estadístico de la zona.

Para las 16 variables atmosféricas que cuenta la estación meteorológica, se realizaron estadísticas como: el promedio, la mediana, el mínimo, el máximo y la desviación estándar. Es interesante, ya que brindan una idea de la información existente. Las temperaturas del aire se comportan como una distribución normal, según los histogramas de frecuencia realizados.

Para estimar si los datos pertenecen a una distribución normal, al vector de máximos, y mínimos absolutos por cada hora de TA, y al promedio por cada hora de RR, SR, PA, HR, TG WIND\_SP, WIND\_DIR, en el cual se ejecutó una prueba de normalidad de "Shapiro-Wilk", y así poder afirmar o rechazar datos gaussianos.

### 2.1 Enfoque de la investigación

El presente trabajo investigativo es de tipo meteorológico, cual fue realizado mediante un enfoque cuantitativo que considera herramientas estadísticas inferenciales.

#### 2.1.1 Tipo de investigación

Por el diseño y tipo de proyecto se determinó necesario un modelo documental y exploratorio.

### 2.1.2 Obtención de Datos.

Para poder desarrollar y obtener los resultados esperados en este proyecto se trabajó con datos proporcionados por el INAMHI, extraídos de la EMA. Dichos datos proporcionaron mediciones de varias variables atmosféricas que son medidas por la EMA. El periodo de tiempo en el que se encuentran recolectados estos datos, están en episodios minútales desde el 2017 hasta el 2019.

Se realizó la eliminación de los datos sospechosos, luego se obtuvo el máximo por hora, el cuál fue almacenado en un vector, con éste se ejecutó la función fevd con las dos familias de distribuciones Gumbel y GEV, a continuación se agruparon los datos por meses. Para cada mes se ejecutó el comando fevd del software de programación R, para luego comparar los resultados del modelo que se ajusta mejor a los datos. Luego se utilizó el comando lr.test, con la finalidad de escoger el mejor modelo con un nivel de significancia (valor P) mayor a 0.05. Si el parámetro alpha es menor que 0.05, entonces el modelo elegido es GEV, caso contrario se eligió el modelo Gumbel. Obteniendo los parámetros de locación y escala de la familia de distribuciones GEV.

Se generaron números aleatorios con el comando revd, que utiliza los parámetros de los números aleatorios y luego se calculó el percentil 99, representado para los valores máximos en contraste de los valores absolutos máximos. Mientras que el percentil 1 representa la banda de confianza mínima, en contraste de los valores absolutos mínimos. Este procedimiento se realizó de los datos agrupados por mes, es decir, se modeló para los meses de enero hasta diciembre.

Para el promedio se obtuvo la media para cada hora, se ordenó el vector por meses y de modo multivariado se usó RR, SR, PA, HR, TG, WIND\_SP y WIND\_DIRECTION en una tabla.

Además, se generó el límite superior, con el número de variables a usar cuyo tamaño es igual a 60, indicando cada cuantos minutos se promedió, es decir

cada hora, y el intervalo de confianza que es 0.9973. Indicando que habrá horas donde los datos salen del límite de control obtenido de la metodología  $T^2$  de Hotelling.

Se filtró la base de datos usando fecha y hora y se eliminaron los datos que están por fuera de los umbrales obtenidos, por cada minuto.

El máximo y el mínimo de la temperatura del aire sirvieron de modo univariado, mientras que las temperaturas del suelo estuvieron en el análisis multivariado, de modo general, al parecer se encuentran fuera de control o en su defecto es una variable estacionaria.

## **2.2 Herramientas Estadísticas**

### **2.2.1 Estadísticos**

Estadístico es cualquier función de los datos de la muestra.

Si los datos de la muestra se designan por  $X_1, X_2, \dots, X_n$  cualquier función  $T(X_1, X_2, \dots, X_n)$  es un estadístico. Por lo tanto solo existen estadísticos cuando la variable observada es cuantitativa. El poder de síntesis de los estadísticos es muy grande, ya que cada estadístico resume el conjunto de todos los datos en un único valor. En contrapartida la información suministrada por cada estadístico tiene que ser forzosamente menor que la información suministrada por toda la muestra. Dependiendo del tipo de información que proporcione un estadístico puede clasificarse en alguno de los siguientes tipos:

*Estadísticos de tendencia central o de localización.* Los más importantes son la media y la mediana muestrales.

*Estadísticos de Posición.* Los más importantes son los cuantiles muestrales y sus variantes: deciles, cuartiles y percentiles. (Luceño Vázquez & González Ortiz, 2015)

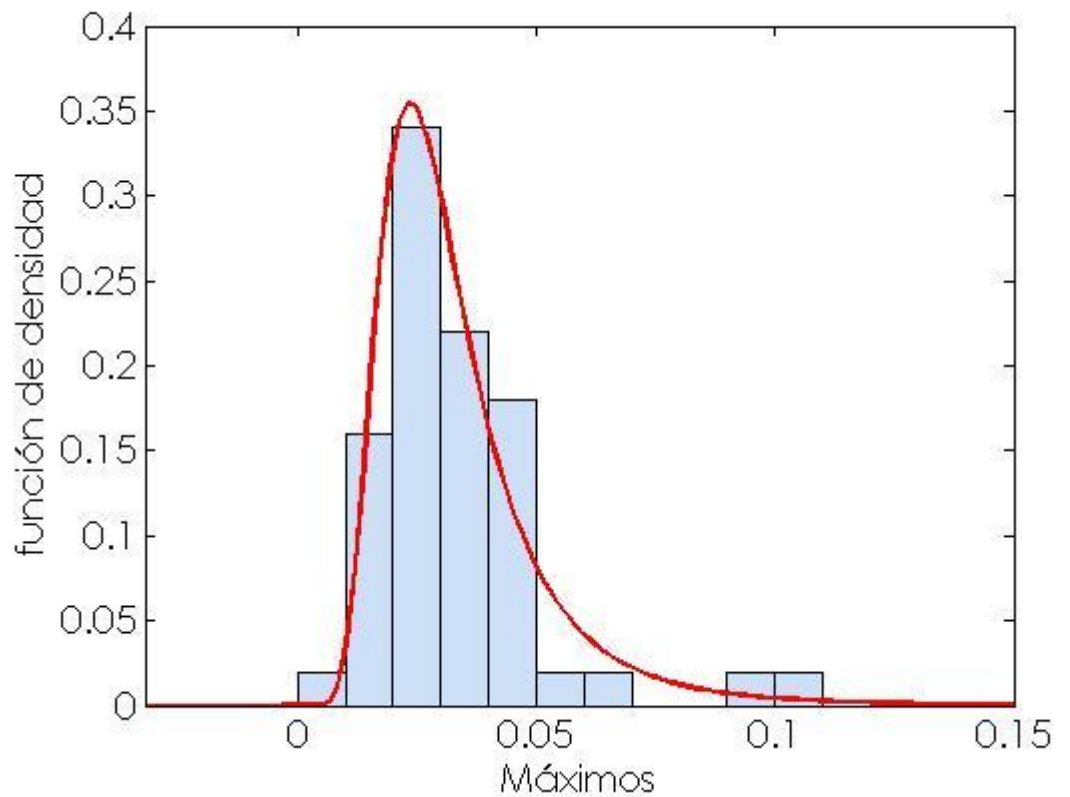
### 2.2.1.1 Promedio o Media

La media aritmética de cierto conjunto de números se encuentra sumando los números y dividiendo después entre la cantidad de datos. (Llinás Solano & Rojas Álvarez, 2006) En otras palabras, si  $x_1, \dots, x_n$  son números, entonces la media aritmética de este conjunto está dada por:

$$\text{Media Aritmética} = \frac{x_1 + \dots + x_n}{n}$$

### 2.2.1.2 Máximo

Según (IBM®) el máximo trata del valor mayor de una variable numérica.

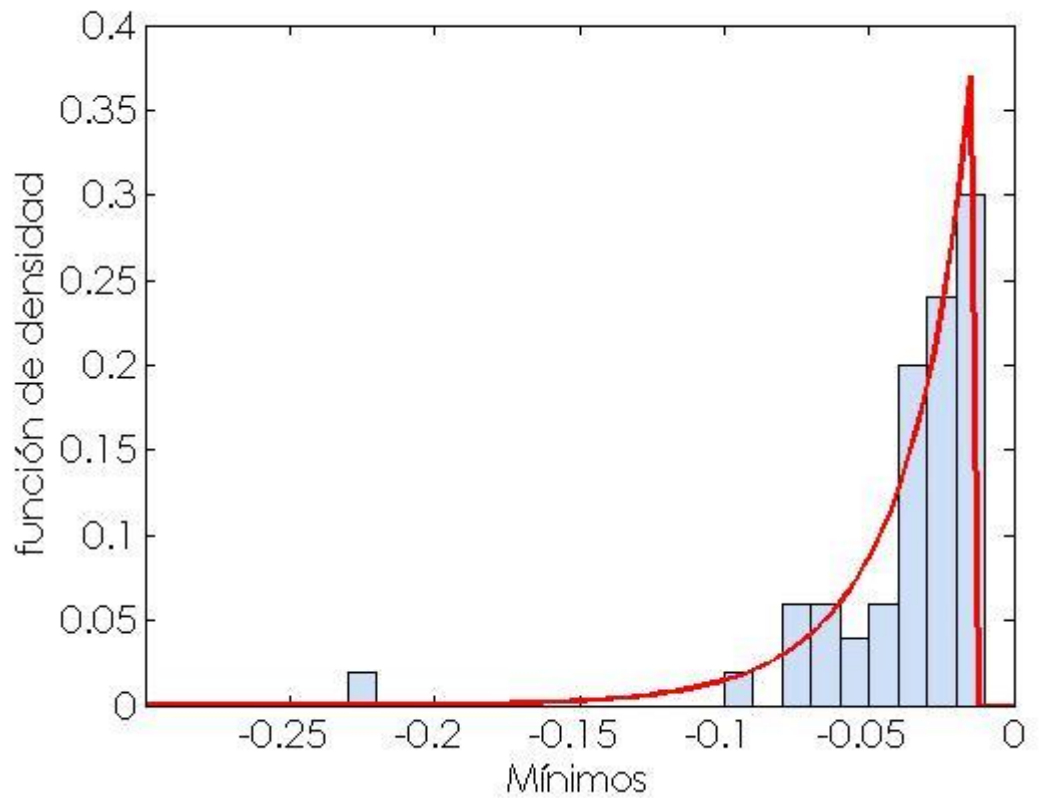


**Figura 2.1 Máximos**

Fuente: <https://quantdare.com/teoria-valores-extremos/>

### 2.2.1.3 Mínimo

Según (IBM®) el mínimo trata del valor menor de una variable numérica.



**Figura 2.2 Mínimos**

Fuente: <https://quantdare.com/teoria-valores-extremos/>

#### 2.2.1.4 Percentiles

Se conoce cómo percentiles a los 99 valares que dividen una serie (ordenada en forma creciente o decreciente), en 100 partes iguales. Un percentil equivale al 1% del total de observaciones de una serie estadística. La primera centésima de la serie es inferior al primer percentil ( $P_1$ ). El percentil 50 ( $P_{50}$ ), es igual a la mitad de las observaciones de la serie, y tiene el valor de la mediana. El percentil 85 ( $P_{85}$ ), corresponde al 85% de las observaciones de la serie, y así los demás percentiles de acuerdo al orden que se busca. (Martínez Vaca)

#### 2.2.2 Estadísticos de Orden

(Weckerly , Mendenhall III, & Scheaffer, 2010) Denota formalmente con  $Y_1, Y_2, \dots, Y_n$  a variables aleatorias continuas e independientes, con función de distribución  $F_{(y)}$  y función de densidad  $f_{(y)}$ . Denotamos las variables aleatorias ordenadas  $Y_{(i)}$  por  $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ , donde  $Y_{(1)} \leq Y_{(2)} \leq \dots \leq Y_{(n)}$ . (Debido a que las variables aleatorias son continuas, los signos de igualdad pueden ignorarse.) Usando esta notación,

$$Y_{(1)} = \min(Y_1, Y_2, \dots, Y_n)$$

es la mínima de las variables aleatorias  $Y_i$ , y

$$Y_{(n)} = \max(Y_1, Y_2, \dots, Y_n)$$

es la máxima de las variables aleatorias  $Y_i$ .

Las funciones de densidad de probabilidad para  $Y_{(1)}$  y  $Y_{(n)}$  se pueden determinar usando el método de las funciones de distribución. Debido a que las  $Y_{(i)}$  son independientes y  $P(Y_i \leq y) = F(y)$  para  $i = 1, 2, \dots, n$ , se deduce que la función de distribución de  $Y_{(n)}$  está dada por

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) \\ &= P(Y_1 \leq y)P(Y_2 \leq y)\dots P(Y_n \leq y) \end{aligned}$$

$$= [F(y)]^n. \quad (2.1)$$

Si con  $g_{(n)(y)}$  denotamos la función de densidad de  $Y_{(n)}$ , vemos que, al evaluar las derivadas de ambos lados,

$$g_n(y) = n[F(y)]^{n-1} f(y). \quad (2.2)$$

La función de densidad para  $Y_{(1)}$  se puede hallar de un modo similar. La función de distribución de  $Y_{(1)}$  es

$$F_{Y_{(1)}}(y) = P(Y_{(1)} \leq y) = 1 - P(Y_{(1)} > y).$$

Como  $Y_{(1)}$  es la mínima de  $Y_1, Y_2, \dots, Y_n$ , se deduce que el evento  $(Y_{(1)} > y)$  ocurre si y sólo si los eventos  $(Y_i > y)$  ocurren para  $i = 1, 2, \dots, n$ . Debido a que las  $Y_i$  son independientes y

$$P(Y_i > y) = 1 - F(y) \text{ para } i = 1, 2, \dots, n, \text{ vemos que}$$

$$\begin{aligned} F_{Y_{(1)}}(y) &= P(Y_{(1)} \leq y) = 1 - P(Y_{(1)} > y) \\ &= 1 - P(Y_1 > y, Y_2 > y, \dots, Y_n > y) \\ &= 1 - [P(Y_1 > y)P(Y_2 > y)P(Y_n > y)] \\ &= 1 - [1 - F(y)]^n. \end{aligned} \quad (2.3)$$

Por consiguiente, si  $g_{(1)}(y)$  denota la función de densidad de  $Y_{(1)}$ , al derivar en ambos lados de la última expresión obtenemos

$$g_{(1)}(y) = n[1 - F(y)]^{n-1} f(y). \quad (2.4)$$

### 2.2.3 Valores Extremos

Puesto que es necesario para este estudio analizar los máximos y mínimos de cada variable se debe tomar en consideración la Teoría de Valores Extremos.

La teoría de valores extremos hace referencia a las posibles desviaciones que puede tener una distribución en relación al valor esperado, su objetivo es, en base a una muestra de una variable aleatoria, determinar qué tan probable es que ocurran eventos aún más extremos.



### 2.2.3.1 Antecedentes de Valores extremos

La Teoría de Valores Extremos y su análisis tiene origen en Leonard Tippett (1928) quien lo utilizó con el fin de obtener mayor resistencia en las fibras de algodón que la empresa donde trabajaba producía, fue un empleado de British Cotton Industry Research Association. Tippett estableció q la resistencia del hilo estaba asociada críticamente con la resistencia de la fibra más frágil. Luego con colaboración de Ronald Fisher obtuvieron los tres límites asintóticos con los que se explicó la distribución de valores extremos. Finalmente Gumbel (1958) compiló todas las hipótesis en su libro de 1950 el cual fue titulado como Statistics of Extremes.

### 2.2.3.2 Teorema de Fisher-Tippett-Gnedenko

También llamado primer teorema de teoría de valores extremos trata de manera general la Teoría de Valores Extremos (TVE) que se refiera a la distribución asintótica para máximos en los estadísticos de orden, este máximo puede converger a tres tipos de distribuciones que son: Gumbel, Fréchet o Weibull (Tippett & Fisher , 1928)

Para poder realizar esta demostración intervino Boris V. Gnedenko (1943) a pesar de que previamente Fréchet (1927) y luego Fisher y Tippett (1928) respectivamente habían expuesto sus enunciados.

El Teorema se enuncia a continuación:

Sea una sucesión  $X_1, X_2, \dots, X_n$  de variables independientes e idénticamente distribuidas,

$M_n = \max\{X_1, X_2, \dots, X_n\}$ . Si existe una sucesión de pares de números reales  $(a_n, b_n)$  tal que para cada  $n$   $a_n > 0$  y  $\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(x)$  donde  $F$  es una función de distribución no-degenerada, entonces la distribución límite  $F$  es o bien una distribución de Gumbel, o una distribución de Fréchet o una distribución de Weibull. Estas distribuciones pueden considerarse casos particulares de Teoría de Valores Extremos.

### 2.2.3.3 Distribución de Gumbel

Desarrollada por E. J. Gumbel, brinda la posibilidad de determinar un modelo mediante el cual se pueda representar valores extremos (tanto máximos como mínimos) para eventos particulares (1934) y (1935), la cual viene dada de la siguiente manera:

- Función de distribución acumulada de Gumbel

$$F(x) = e^{-e^{\frac{(\mu-x)}{\sigma}}} \quad (2.6)$$

- Función de distribución de probabilidad de Gumbel

$$f(x) = \frac{1}{\sigma} * e^{\frac{(\mu-x)}{\sigma}} e^{-e^{\frac{(\mu-x)}{\sigma}}}, \quad x \in (-\infty, +\infty) \quad (2.7)$$

Gumbel es uno de los casos particulares de la distribución de valores extremos generalizada (GEV), la distribución estándar se da cuando  $\mu = 0$  y  $\beta = 1$ . En este caso la función de distribución acumulada y la función distribución de probabilidad de Gumbel serían las siguientes:

- Función de distribución acumulada de Gumbel

$$F(x) = e^{-e^{-x}} \quad (2.8)$$

- Función de distribución de probabilidad de Gumbel

$$f(x) = e^{-x} e^{-e^{-x}}, \quad x \in (-\infty, +\infty) \quad (2.9)$$

Propiedades	
<b>Media</b>	$\mu + \sigma\gamma,$ En el cual $\gamma =$ Constante de Euler-Mascheroni.
<b>Varianza</b>	$\frac{\sigma^2\pi^2}{6}$
<b>Moda</b>	$\mu$
<b>Mediana</b>	$\mu - \sigma\ln(-\ln(2))$

**Tabla 2.1 Propiedades Distribución de Gumbel**  
Fuente: Elaboración propia

#### 2.2.3.4 Distribución Fréchet

El segundo caso particular de la distribución generalizada de los valores extremos (GEV) es la que recibe su nombre de matemático francés Maurice Fréchet. Se usa principalmente en hidrología para estudiar máximo de precipitación y descarga de ríos (Oosterbaan, 1994), también es usada para especificar periodos de sequía (Burke, Perry, & Brown, 2010). La distribución de Fréchet cuando posee tres parámetros se enuncia de la manera siguiente (1927):

$\xi =$  Parámetro de Forma,  $\xi \in (0, \infty)$

$\mu =$  Parámetro de Localización,  $\mu \in (-\infty, +\infty)$

$s =$  Parámetro de Escala,  $\sigma \in (0, \infty)$

- Función de distribución acumulada de Fréchet

$$F(x) = e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}}, \text{ si } x > \mu \quad (2.10)$$

- Función de distribución de probabilidad de Fréchet

$$f(x) = \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{-1-\xi} e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}}, \text{ si } x > \mu \quad (2.11)$$

Para el caso en que  $\mu = 0$ , la función de distribución es la siguiente:

- Función de distribución acumulada de Fréchet

$$F(x) = e^{-\left(\frac{x}{\sigma}\right)^{-\xi}}, \text{ si } x > 0 \quad (2.12)$$

- Función de distribución de probabilidad de Fréchet

$$f(x) = \frac{\xi}{\sigma} \left(\frac{x}{\sigma}\right)^{-1-\xi} * e^{-\left(\frac{x}{\sigma}\right)^{-\xi}}, \text{ si } x > 0 \quad (2.13)$$

Propiedades	
Media	$\mu + \sigma \Gamma\left(1 - \frac{1}{\xi}\right)$ , siempre que $\xi > 1$
Varianza	$\sigma^2 \left[ \Gamma\left(1 - \frac{2}{\xi}\right) - \left(\Gamma\left(1 - \frac{1}{\xi}\right)\right)^2 \right]$ , siempre que $\xi > 2$
Moda	$\mu + \sigma \left(\frac{\xi}{1 + \xi}\right)^{\frac{1}{\xi}}$
Mediana	$\mu + \frac{\sigma}{\sqrt{\xi \ln(2)}}$

**Tabla 2.2 Propiedades Distribución de Fréchet**  
Fuente: Elaboración propia

### 2.2.3.5 Distribución Weibull

Finalmente el tercer caso particular de la distribución generalizada de los valores extremos (GEV), en un principio descubierta (Fréchet, 1927) y definida en detalle por Waloddi Weibull matemático sueco (1951), se define de la siguiente manera:

- Función de distribución acumulada de Weibull

$$F(x) = 1 - e^{-\left(\frac{x-\mu}{\sigma}\right)^\xi}, \text{ si } x \geq 0 \quad (2.14)$$

- Función de distribución de probabilidad de Weibull

$$f(x) = \frac{\xi}{\sigma} \left(\frac{x-\mu}{\sigma}\right)^{\xi-1} e^{-\left(\frac{x-\mu}{\sigma}\right)^\xi}, \text{ si } x \geq 0 \quad (2.15)$$

$\xi$  es el Parámetro de Forma y  $\sigma > 0$  es el Parámetro de Escala de la distribución.

Propiedades	
Media	$\mu + \sigma \Gamma\left(1 + \frac{1}{\xi}\right)$
Varianza	$\sigma^2 \left[ \Gamma\left(1 + \frac{2}{\xi}\right) - \left(\Gamma\left(1 + \frac{1}{\xi}\right)\right)^2 \right]$
Moda	$\mu + \sigma \left(\frac{\xi - 1}{\xi}\right)^{\frac{1}{\xi}}, \text{ si } \xi > 1$
Mediana	$\mu + \sigma (\ln(2))^{\frac{1}{\xi}}$

**Tabla 2.3 Propiedades de la distribución de Weibull**

Fuente: Elaboración propia

### 2.2.3.6 Distribución de Valores Extremos Generalizada (GEV)

A esta distribución se la conoce con varios nombres entre los cuales se encuentran Fisher-Tippett, tipo von Mises-Jenkinson o tipo von Mises. La cual en un principio se introdujo por Jenkinson (1955), dentro de esta distribución se explica de manera general las distribuciones de Gumbel, Fréchet y Weibull como si fuese una única distribución, la cual se define de la siguiente manera:

$$G(z) = e^{-\left(1+\xi\left(\frac{z-\mu}{\sigma}\right)\right)^{\frac{1}{\xi}}}, \quad \text{dado } 1 + \xi \left(\frac{z-\mu}{\sigma}\right) > 0 \quad (2.16)$$

Los parámetros que componen esta distribución son:

- Parámetro de Ubicación / Posición: rango y tendencia central ( $\mu \in \mathbb{R}$ )
- Parámetro de Escala: tendencia central y dispersión ( $\sigma > 0$ )
- Parámetro de Forma: dispersión y momentos de alto orden ( $\xi \in \mathbb{R}$ )

En base al valor que tome el parámetro de forma ( $\xi$ ) de la función generalizada se decide la familia a la que se aproxima su función de distribución ya sea Gumbel, Fréchet o Weibull, para esto el parámetro de forma puede tomar los siguientes valores:

- Si el parámetro de forma es que cero  $\xi > 0$ , entonces se aproxima a una distribución Fréchet.
- Si el parámetro de forma es menor que cero  $\xi < 0$ , entonces se aproxima a una distribución Weibull.
- Si el parámetro de forma es igual a cero  $\xi = 0$ , entonces se aproxima a una distribución Gumbel.

Propiedades	
<b>Parámetros</b>	$\mu \in \mathbf{R}$ — Ubicación / Posición $\sigma > 0$ — Escala $\xi \in \mathbf{R}$ — Forma
<b>Media</b>	$\begin{cases} \mu + \sigma \frac{\Gamma(g_1 - 1)}{\xi} & , si \xi \neq 0, \xi < 1 \\ \mu + \sigma \gamma & , si \xi = 0 \\ No existe & , si \xi \geq 1 \end{cases}$ <p>En el cual  <math>g_k = \Gamma(1 - k\xi)</math>  <math>\Gamma()</math> es la función Gamma,  <math>\gamma</math> es la cte. de Euler.</p>
<b>Varianza</b>	$\begin{cases} \sigma^2 \frac{(g_2 - g_1^2)}{\xi^2} & , si \xi \neq 0, \xi < \frac{1}{2} \\ \sigma^2 \frac{\pi^2}{6} & , si \xi = 0 \\ No existe & , si \xi \geq \frac{1}{2} \end{cases}$ <p>En el cual <math>g_k = \Gamma(1 - k\xi)</math>.</p>
<b>Moda</b>	$\begin{cases} \mu + \sigma \frac{(1 + \xi)^{-\xi} - 1}{\xi} & , si \xi \neq 0 \\ \mu & , si \xi = 0 \end{cases}$
<b>Mediana</b>	$\begin{cases} \mu + \sigma \frac{(\ln 2)^{-\xi} - 1}{\xi} & , si \xi \neq 0 \\ \mu - \sigma \ln(\ln(2)) & , si \xi = 0 \end{cases}$
<b>Cuantil p</b>	$\begin{cases} \mu + \sigma \frac{\ln\left(\frac{1}{p}\right)^{-\xi} - 1}{\xi} & , si \xi \neq 0 \\ \mu - \sigma \ln\left(\ln\left(\frac{1}{p}\right)\right) & , si \xi = 0 \end{cases}$

**Tabla 2.4 Propiedades Distribución GEV**

Fuente: Elaboración propia

#### **2.2.4 Fases de un Gráfico de Control**

En referencia a **(Montgomery, 2006)** el uso de gráficos de control estándar implica aplicaciones de Fase I y Fase II, con dos objetivos diferentes y distintos.

En la Fase I, se recopila y analiza un conjunto de datos del proceso de una vez en un análisis retrospectivo, construyendo límites de control de prueba para determinar si el proceso ha estado en control durante el período de tiempo en que se recopilaron los datos, y para ver si es confiable.

Se pueden establecer límites de control para monitorear la producción futura. Esto suele ser lo primero que se hace cuando los gráficos de control se aplican a cualquier proceso.

Los cuadros de control en la Fase I ayudan principalmente al personal operativo a llevar el proceso a un estado de control estadístico.

Es bastante típico en la fase I suponer que el proceso está inicialmente fuera de control, por lo que el objetivo del analista es llevar el proceso a un estado de control estadístico. Los límites de control se calculan en función de los subgrupos  $m$  y los datos trazados en los gráficos de control. Se investigan los puntos que están fuera de los límites de control, buscando posibles causas asignables. El personal de ingeniería y operación trabaja en las causas asignables identificadas en un esfuerzo por eliminarlas. Los puntos fuera de los límites de control se excluyen y se calcula un nuevo conjunto de límites de control revisados. Luego se recopilan nuevos datos y se comparan con estos límites revisados.

A veces, este tipo de análisis requerirá varios ciclos en los que se empleará el cuadro de control, se detectarán y corregirán las causas asignables, se calcularán los límites de control revisados y se actualizará y ampliará el plan de acción fuera de control. Finalmente, el proceso se estabiliza y se obtiene un conjunto limpio de datos que representa el rendimiento del proceso bajo control para su uso en la fase II.



La Fase II comienza después de que tenemos un conjunto "limpio" de datos de proceso recopilados en condiciones estables y representativas del rendimiento del proceso bajo control.

En la fase II, se utiliza el cuadro de control para monitorear el proceso comparando el estadístico de la muestra para cada muestra sucesiva a medida que se extrae del proceso a los límites de control.

### 2.2.5 Gráfico de Control Multivariado T<sup>2</sup> de Hotelling

Es importante para el presente estudio trabajar con la teoría matemática de Mahalanobis (1936), el cual introdujo una expresión para medir la distancia entre dos variables aleatorias ( $X, Y$ ) de dimensión  $p$  con la misma función de distribución de probabilidades y matriz de varianzas y covarianza  $\Sigma$ .

Esta expresión se definió de la siguiente forma:

$$d_m(X, Y) = \sqrt{(X - Y)^T \Sigma^{-1} (X - Y)}$$

Siendo  $X^T = (X_1, X_2, \dots, X_p)$  un conjunto de variables con un vector de medias  $\mu^T = (\mu_1, \mu_2, \dots, \mu_p)$  y matriz de varianzas y covarianzas  $\Sigma$  (de dimensión  $p \times p$ ) simétrica y definida positiva, el cuadrado de la distancia de Mahalanobis estandarizada de  $X$  a  $\mu$  estaría dado por:

$$d_m^2(X, Y) = (X - \mu)^T \Sigma^{-1} (X - \mu)$$

Para el caso en el que  $X$  provenga de una distribución normal multivariante con media  $\mu$  y matriz de varianzas y covarianzas  $\Sigma$ , se define al estadístico  $\chi_p^2$  con una distribución chi-cuadrada con  $p$  grados de libertad de la siguiente manera:

$$\chi_p^2 = n(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)$$

En donde  $\bar{X}$  es el vector de medias de la muestra.

Ya que por lo general no se conocen los parámetros del proceso hay que estimarlos, y para ello Harold Hotelling estableció el estadístico  $T^2$  que se utiliza para un conjunto de variables que siguen una distribución normal multivariada y que se define de la siguiente manera:

$$T^2 = n(\bar{X} - \hat{\mu})^T S^{-1}(\bar{X} - \hat{\mu})$$

Donde  $S$  y  $\hat{\mu}$  son los estimadores de la matriz de varianzas y covarianzas y del vector de medias respectivamente. Esta estimación se la realiza con  $m$  muestras de tamaño  $n$  y, si  $m$  tiende al infinito, la distribución del estadístico  $T^2$  converge a una distribución Chi-cuadrado con  $p$  grados de libertad (Seber, 1984).

El gráfico de control del estadístico  $T^2$  tiene como límite de control superior  $LCS = \chi_{\alpha, p}^2$  y como límite de control inferior  $LCI = 0$ , siendo  $\alpha$  el nivel de significancia y  $p$  el número de variables de calidad que serán controladas.

Cuando el proceso ha sufrido una desviación en al menos uno de los promedios de sus variables, entonces el vector  $\mu_1$  se aleja del vector de promedios del proceso bajo control  $\mu_0$  en una distancia  $d$  de Mahalanobis. Entonces, el estadístico  $T^2$  pasa a tener una distribución chi-cuadrado no centralizada con  $p$  grados de libertad y con parámetro de descentralizado definido por la siguiente ecuación:

$$\lambda = nd^2 = n(\mu_1 - \mu_0)^T \Sigma^{-1}(\mu_1 - \mu_0)$$

En donde  $n$  es el tamaño de la muestra y  $\Sigma$  es a matriz de varianzas y covarianzas de  $X$  cuando el proceso se encuentra bajo control. El ARL de este tipo de gráfico es calculado de la siguiente manera:

$$ARL = \frac{1}{1 - P(T^2 < LCS)}$$

En donde LCS es el límite de control superior. En el caso de que el estadístico caiga exactamente en el límite, entonces existe una señal de que el proceso pudiera estar fuera de control.

### 2.2.6 Longitud Promedio de Corrida (ARL)

Según **(Montgomery, 2006)** se denomina ARL (Average Run Length) al número promedio de puntos que se debe graficar en una carta de control antes de que un punto indique una condición fuera de control

$$ARL = \frac{1}{p}$$

donde p es la probabilidad de que cualquier punto graficado se ubique fuera los límites de control.

Para casos en los que la información es poca, se recomienda el uso de un  $p=0,0027$  con este valor se alcanza aproximadamente límites de 3 sigmas.

Se clasifica de la siguiente manera:

#### **ARL<sub>0</sub> (ARL Bajo Control)**

Se define como el número de puntos en una carta de control hasta que el gráfico muestre una condición fuera de control.

Se emplea bajo la restricción de que el proceso se encuentra bajo control.

$$ARL = \frac{1}{\alpha}$$

Donde  $\alpha$  = probabilidad de obtener falso positivo (probabilidad de ocurrencia del error tipo I)

#### **ARL<sub>1</sub> (ARL Fuera de Control)**

Se define como el número de puntos en una carta de control hasta que el gráfico muestre una condición fuera de control.

Se emplea bajo la restricción de que el proceso se encuentra fuera control.

$$ARL = \frac{1}{1 - \beta}$$

Donde  $\beta$  = probabilidad de obtener falso negativo (probabilidad de ocurrencia del error tipo II)

El esquema más eficiente para las cartas de control es donde el ARL0 toma el valor más alto posible mientras que el ARL1 toma el menor valor posible, así se asegura que revele más ágilmente los cambios en el proceso.

# CAPÍTULO 3

## 3. RESULTADOS Y ANÁLISIS:

En este capítulo los resultados obtenidos en base a la metodología utilizada fueron:

### 3.1 Análisis Especializado

Si una variable es estacionaria y se encuentra en un rango conocido, esta brinda información necesaria para ser considerada disparadora o de la cual dependen las restantes y que si sufre algún cambio se verá reflejado automáticamente en las demás variables.

A continuación, es desarrollada la metodología explicada en el capítulo 2, para de esta manera estudiar el comportamiento de la variable seleccionada para el estudio (variable piloto).

#### 3.1.1 Eliminación de valores sospechosos

Para la obtención de los gráficos de control para imputación por máximos y mínimos se usa la variable Temperatura del Aire (TA) como piloto, debido a que ésta en una zona presenta un comportamiento estacionario en los datos, con media y varianza constante a medida que transcurre el tiempo.

Basado en la evidencia estadística se resuelve adjuntar los siguientes análisis:

- Análisis de Máximos para Temperaturas del Aire
- Análisis de Mínimos para Temperaturas del Aire
- Análisis multivariado del Promedio de Todas las variables objeto de estudio.

Estos análisis se presentarán en las dos corridas de la Fase 1 en las que se llevará a efecto el proceso para eliminación de valores sospechosos.

Como cada mes tiene un comportamiento diferente también tiene una banda de confianza diferente.

Para determinar cuál de los dos modelos es el más robusto se comparó los parámetros de la función de **GUMBEL** y la función de **GEV** mediante la Prueba De Razón De Verosimilitud la cual permite comparar modelos anidados y mediante el cálculo del p-value indica que modelo es más robusto por lo tanto se ajusta mejor. (Coles, 2001)

Para esta prueba toma como modelo más robusto al modelo GEV, solo en caso de rechazarse se tomaría como mejor modelo el GUMBEL.

Se determinó el modelo que es mejor en cada mes.

A continuación, se muestran los resultados del análisis estadístico, para los datos divididos por meses, las respectivas bandas de confianza son generadas a partir el percentil 99% para el máximo, y 1% para el mínimo.

### 3.1.2 Fase 1 (Corrida 1)

Gráficos de Control para las Temperaturas Máximas del Aire

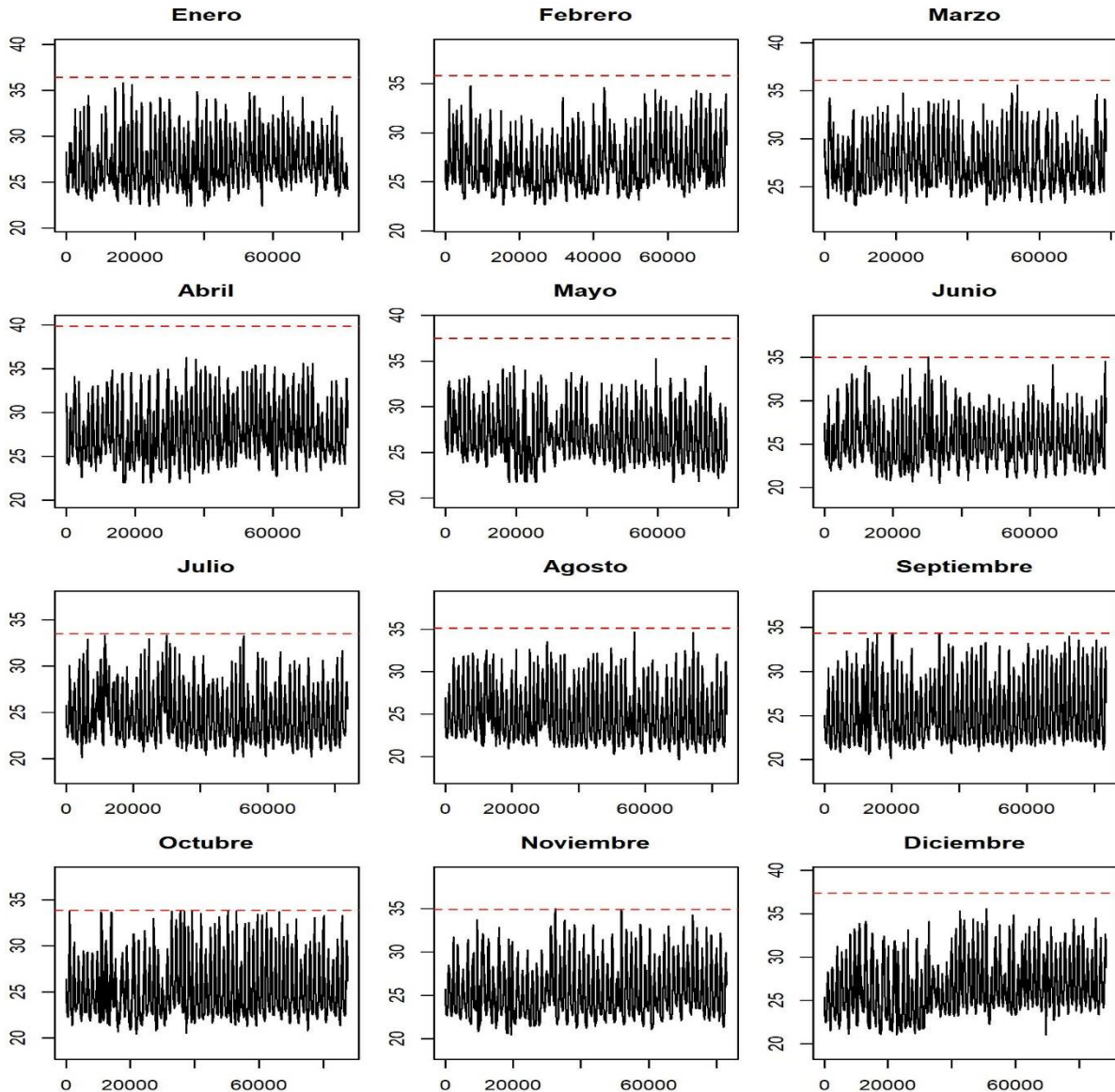


Figura 3.1 Gráficos de Control TA Máximas Fase 1 (Corrida 1)

Fuente: Elaboración propia

En el análisis que muestra la Figura 3.2 se pudo determinar que en los meses de Enero, Junio, Julio, Septiembre, Octubre y Noviembre existieron valores que excedan el límite confianza superior establecido. Siendo el mes de Octubre en el que se presentaron la mayor cantidad de valores por encima del umbral, lo que implica que en ese punto donde los valores exceden la banda de confianza, existe un minuto en el que se presentó una máxima absoluta.

Mediante un proceso de selección se agruparon todos los datos de cada mes por separado y para cada mes se realizó la estimación de la distribución de **GUMBEL** que devuelve el parámetro de localización, escala y forma, luego se aplicó el mismo procedimiento para la distribución de **GEV**.

Los parámetros para cada mes se presentan a continuación.

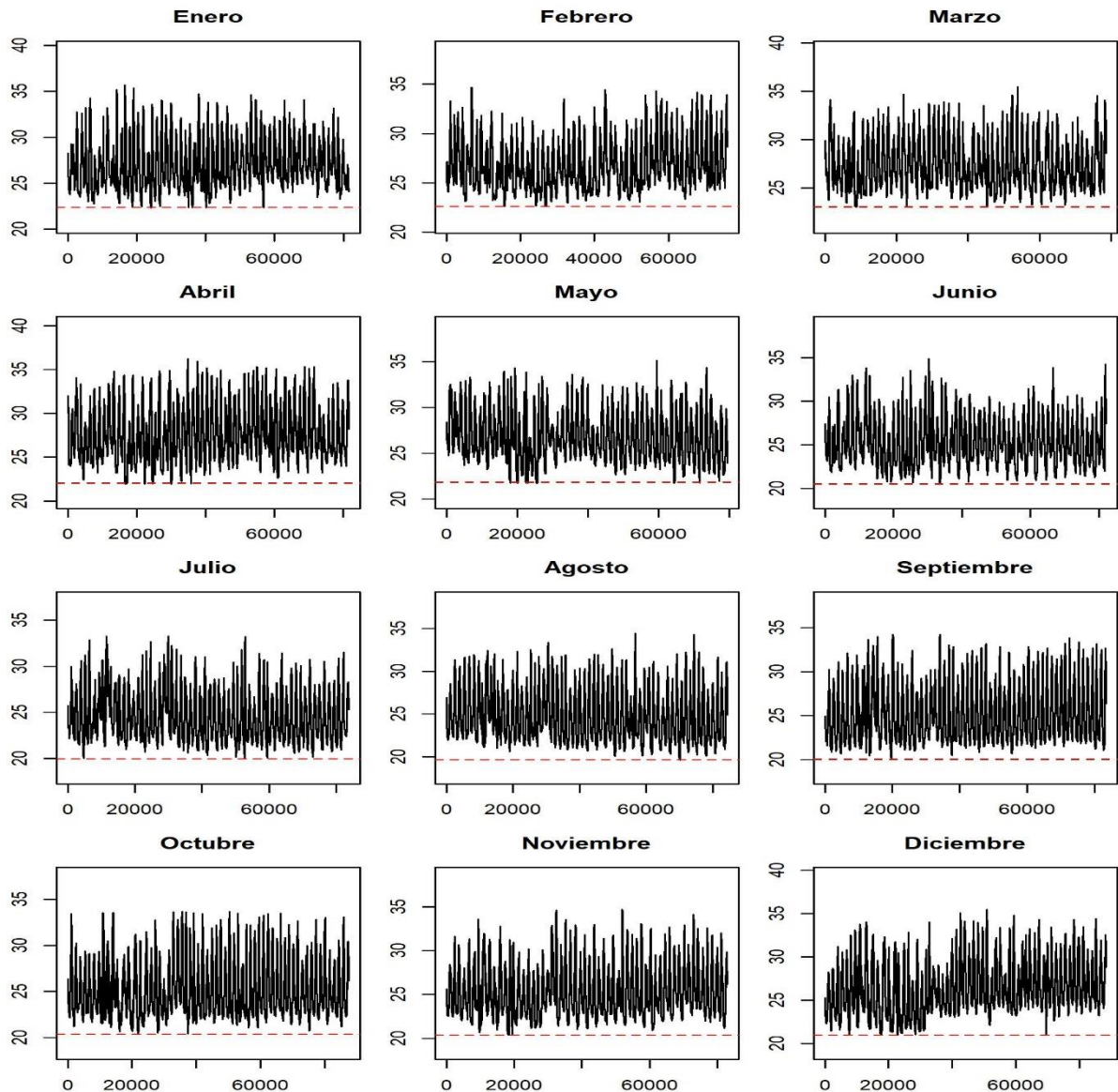
<b>Mes</b>	<b>Banda de Confianza</b>	<b>Parámetros de Localización</b>	<b>Parámetro de Escala</b>	<b>Parámetro de Forma</b>	<b>Modelo</b>
Enero	37.06504	26.21094	2.378114	NA	Gumbel
Febrero	35.97705	26.08768	2.169059	NA	Gumbel
Marzo	36.64899	26.65330	2.169934	NA	Gumbel
Abril	40.92354	26.90660	3.024019	-0.13982443	GEV
Mayo	38.49651	26.07054	2.737885	-0.16492729	GEV
Junio	35.69086	24.57215	2.455694	NA	Gumbel
Julio	34.37809	23.77367	2.335476	0.09653136	GEV
Agosto	36.52508	23.98708	2.715478	NA	Gumbel
Septiembre	35.38329	24.01770	2.475786	0.25168055	GEV
Octubre	34.85597	24.15596	2.317029	0.25259183	GEV
Noviembre	35.99872	24.44368	2.528478	0.06508476	GEV
Diciembre	38.31106	25.62032	2.782798	-0.08083862	GEV

**Tabla 3.1**

Fuente: Elaboración propia



### Gráficos de Control para las Temperaturas Mínimas del Aire



**Figura 3.2 Gráficos de Control TA Mínimas Fase 1 (Corrida 1)**

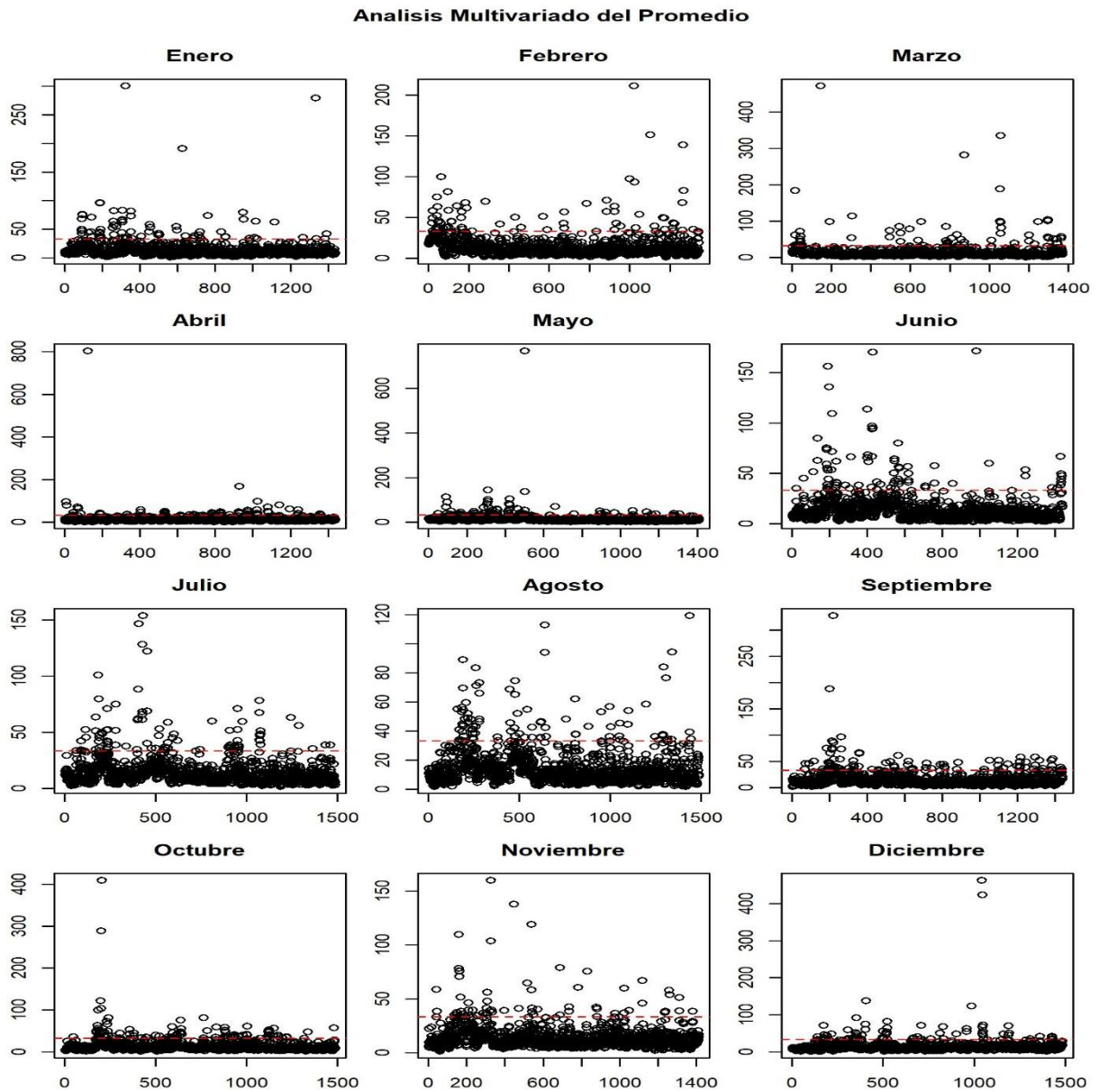
Fuente: Elaboración propia

En el análisis que muestra la Figura 3.3 se pudo determinar que solo en el mes de Julio no existen puntos que sobrepasan las bandas de confianza, en todos los demás meses del año si existen puntos fuera de los límites, siendo los meses de Abril y Mayo los que presentan el mayor excedente. Lo que quiere decir que en ese punto que excede la banda de confianza existió un minuto en el que se presentó una mínima absoluta la cual se muestra en los datos.

<b>Mes</b>	<b>Banda de Confianza</b>	<b>Parámetros de Localización</b>	<b>Parámetros de Escala</b>	<b>Parámetros de Forma</b>	<b>Modelo</b>
Enero	22.25564	25.40688	2.077824	-0.05407181	GEV
Febrero	22.53242	25.25443	1.775752	NA	Gumbel
Marzo	22.97396	25.74398	1.811213	NA	Gumbel
Abril	21.82692	25.85291	2.634931	-0.15943648	GEV
Mayo	21.51100	25.20659	2.433200	-0.20272631	GEV
Junio	20.45270	23.70571	2.126238	NA	Gumbel
Julio	19.91134	22.87941	1.945720	0.10066707	GEV
Agosto	19.58073	22.93241	2.184293	0.07643331	GEV
Septiembre	19.98630	22.98832	1.960590	0.26560203	GEV
Octubre	20.31853	23.20947	1.886608	0.20946328	GEV
Noviembre	20.28441	23.53526	2.136145	NA	Gumbel
Diciembre	20.88880	24.63747	2.453812	-0.11898913	GEV

**Tabla 3.2 Parámetros del Gráfico de Control para las Temperaturas Mínimas del Aire Fase 1 (Corrida 1)**

Fuente: Elaborado por el autor.



**Figura 3.3 Gráficos de Control T2 Hotelling Fase 1 (Corrida 1)**

Fuente: Elaboración propia

En el análisis del proceso multivariado presentado por la Figura 3.3 se muestra el índice calculado por la metodología T2 Hotelling, el gráfico contiene los promedios de cada hora divididos por meses de las variables RR, TA, PA, SR\_GLOBAL, TG1, TG2, TG3, TG4, TG5, TG6, TG7, WIND\_DIR, WIND\_SPD.

Con el índice T2 Hotelling la metodología brinda una banda de confianza obtenida por cada mes del estudio.

En los meses de Julio, Agosto, Septiembre y Octubre se presentó mayor cantidad de datos que sobrepasaron la banda de confianza.

<b>Mes</b>	<b>Banda de Confianza</b>
Enero	33.18132
Febrero	33.18035
Marzo	33.18064
Abril	33.18132
Mayo	33.18102
Junio	33.18135
Julio	33.18179
Agosto	33.18179
Septiembre	33.18135
Octubre	33.18179
Noviembre	33.18108
Diciembre	33.18176

**Tabla 3.3 Bandas de Confianza para T2 de Hotelling Análisis Multivariado del Promedio Fase 1 (Corrida 1)**

Fuente: Elaboración propia

**Nota:** Se debe recordar que este estudio es un análisis multivariado, las observaciones mostradas en el gráfico representan la combinación de todas las variables antes mencionadas, por lo que la metodología permite eliminar los valores que se encuentran por encima del umbral, cada hora representa 60 minutos por lo cual se debe filtrar esa hora y al eliminarla se eliminan cada minuto filtrado por hora.

En todos los meses observamos que existen datos que exceden la banda de confianza y estos valores se sugiere deben ser eliminados para obtener datos depurados en la 1era Fase.

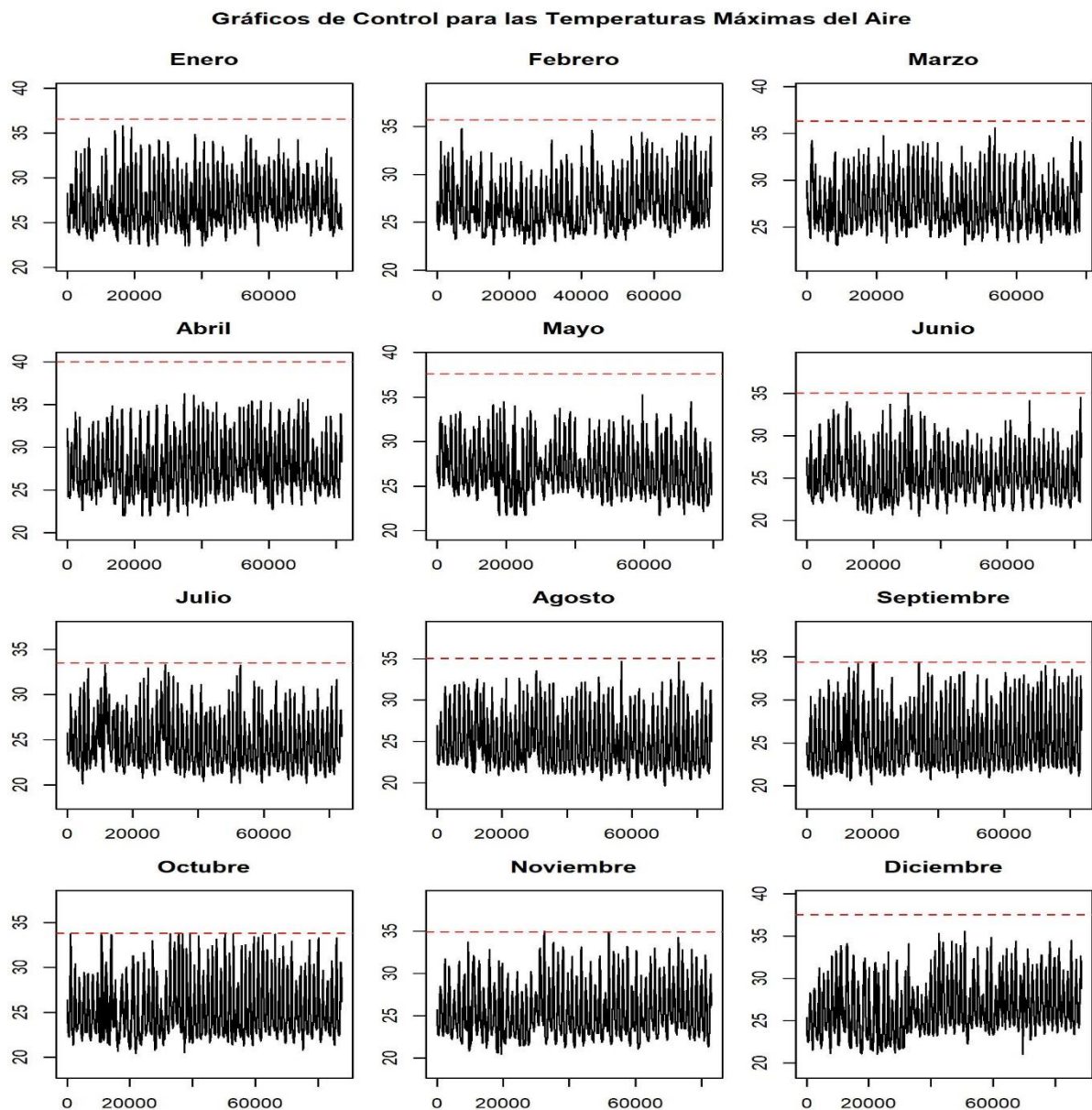
Con esta base de datos obtenida que ha sido filtrada de estos valores atípicos se ejecuta nuevamente la misma metodología para llegar a la Fase 2 y obtener observaciones estables de la base de datos.

### 3.1.3 Fase 1 (Corrida 2)

A continuación se vuelve a ejecutar el análisis previamente realizado en la Corrida 1 pero esta vez habiendo eliminado los valores que en dicha corrida sobrepasaron la banda de confianza establecida.

Al ejecutar nuevamente el análisis se volvieron a calcular las bandas de confianza y todos los parámetros para esta Corrida.

Los resultados se presentan a continuación:



**Figura 3.4 Gráficos de Control TA Máximas Fase 1 (Corrida 2)**

Fuente: Elaboración propia

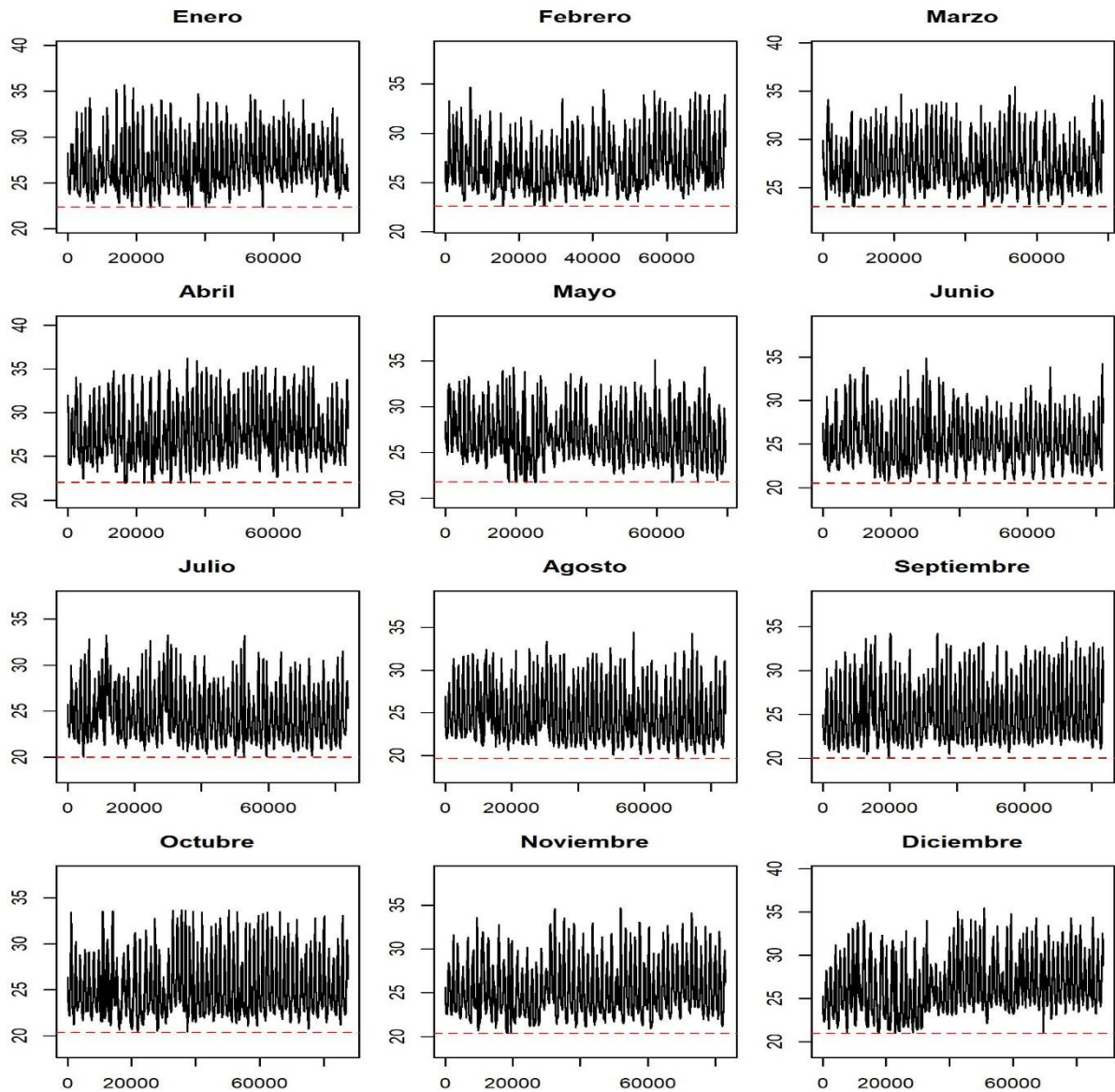
El análisis mostrado por la Figura 3.4 determina que en los meses de Julio, Septiembre, Octubre y Noviembre existen puntos que superan la banda de confianza, en estos puntos se presentaron máximos absolutos que se muestran en los datos. La mayor cantidad de datos que exceden la banda de confianza se presenta en los meses de Julio y Octubre.

<b>Mes</b>	<b>Banda de Confianza</b>	<b>Parámetros de Localización</b>	<b>Parámetros de Escala</b>	<b>Parámetros de Forma</b>	<b>Modelo</b>
Enero	36.54764	26.12047	2.266707	NA	Gumbel
Febrero	35.53381	26.02025	2.088515	NA	Gumbel
Marzo	36.14608	26.58102	2.098517	NA	Gumbel
Abril	40.11088	26.80232	2.878017	-0.12495831	GEV
Mayo	37.75703	25.98715	2.531606	-0.14331090	GEV
Junio	35.17102	24.40081	2.316604	NA	Gumbel
Julio	33.45488	23.55353	2.147591	0.12484229	GEV
Agosto	34.88484	23.66808	2.458276	0.09944924	GEV
Septiembre	34.23806	23.80523	2.293785	0.30304258	GEV
Octubre	33.78884	23.95965	2.149526	0.28205019	GEV
Noviembre	34.99377	24.24034	2.327409	0.12395009	GEV
Diciembre	37.53609	25.43080	2.619984	NA	Gumbel

**Tabla 3.4 Parámetros del Gráfico de Control para las Temperaturas Máximas del Aire Fase 1 (Corrida 2)**

Fuente: Elaboración propia

### Gráficos de Control para las Temperaturas Mínimas del Aire



**Figura 3.5 Gráficos de Control TA Mínimas Fase 1 (Corrida 2)**

Fuente: Elaboración propia

El análisis mostrado por la Figura 3.6 determina que durante los meses de Julio y Octubre no se presentaron puntos que excedan los límites, sin embargo en los demás meses del año si se presentan valores que exceden la banda de confianza, siendo los meses de Abril y Mayo donde existió la mayor cantidad de valores registrados.

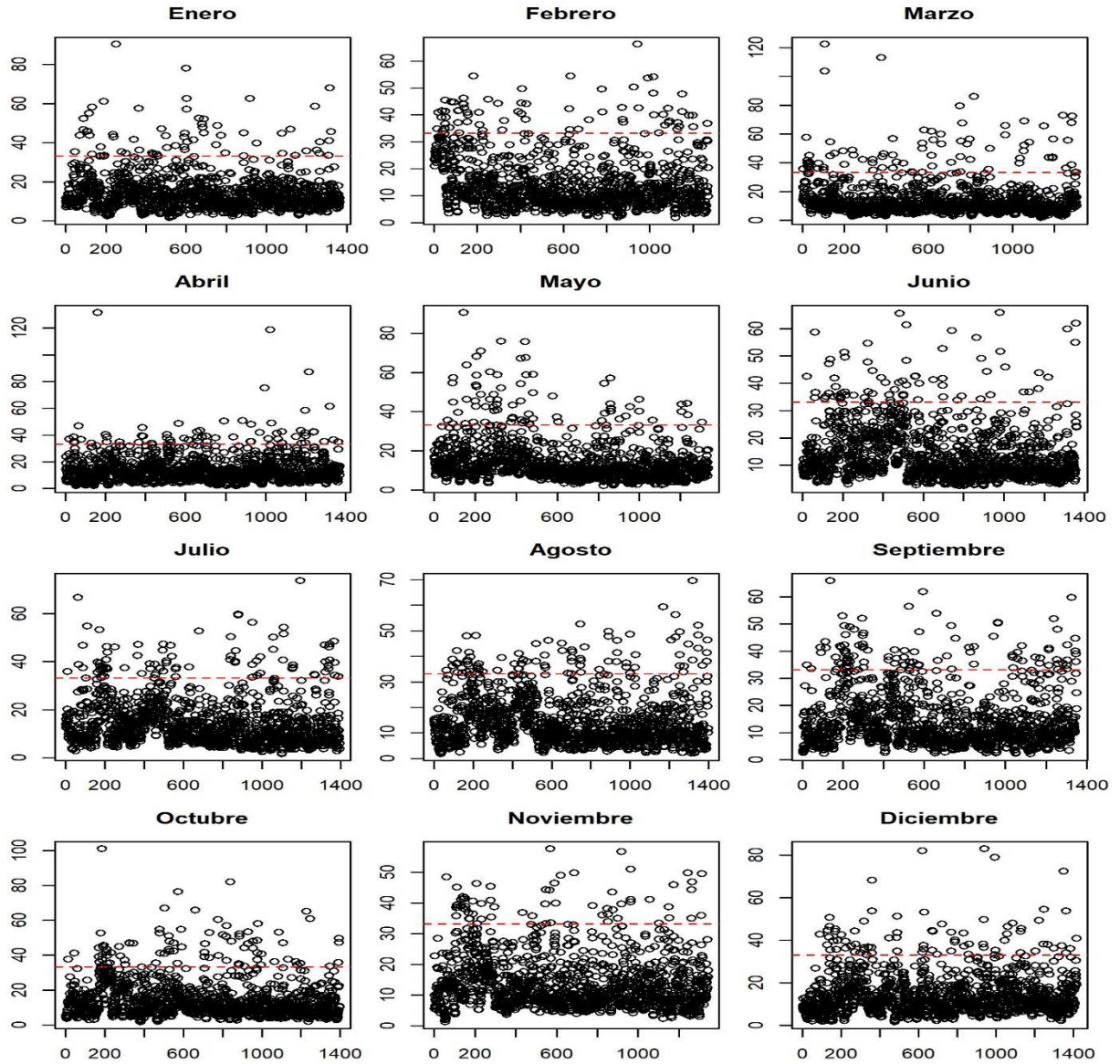
<b>Mes</b>	<b>Banda de Confianza</b>	<b>Parámetros de Localización</b>	<b>Parámetros de Escala</b>	<b>Parámetros de Forma</b>	<b>Modelo</b>
Enero	22.33201	25.29769	1.946230	NA	Gumbel
Febrero	22.60215	25.22855	1.718450	NA	Gumbel
Marzo	23.03141	25.70577	1.745541	NA	Gumbel
Abril	21.97637	25.77384	2.495127	-0.13677813	GEV
Mayo	21.73563	25.14908	2.230255	-0.16664675	GEV
Junio	20.49979	23.57650	2.023236	NA	Gumbel
Julio	19.97633	22.73982	1.815100	0.10042886	GEV
Agosto	19.67290	22.76724	2.044540	0.10437939	GEV
Septiembre	20.06547	22.89304	1.855325	0.27335750	GEV
Octubre	20.35529	23.10388	1.796166	0.21834809	GEV
Noviembre	20.36221	23.35245	1.957522	0.09685652	GEV
Diciembre	20.97445	24.57979	2.364868	-0.09220191	GEV

**Tabla 3.5 Parámetros del Gráfico de Control para las Temperaturas Mínimas del Aire Fase 1 (Corrida 2)**

Fuente: Elaboración propia



### Analisis Multivariado del Promedio



**Figura 3.6 Gráficos de Control T2 Hotelling Fase 1 (Corrida 2)**

Fuente: Elaboración propia

En análisis del proceso multivariado que presenta la Figura 3.7 se observa que en todos los meses del año siguen existiendo datos que exceden la banda de confianza y que se presenta igual cantidad de datos de este tipo para todos los meses.

Según la metodología estos valores se sugiere deben ser eliminados para obtener datos depurados en esta Fase, pero, debido a q el intervalo de confianza se está acercando a la masa de datos se decidió detener el proceso de eliminación debido a que se estaría eliminando valores correctos.

<b>Mes</b>	<b>Banda de Confianza</b>
Enero	33.18057
Febrero	33.17949
Marzo	33.17993
Abril	33.18062
Mayo	33.18025
Junio	33.18053
Julio	33.18088
Agosto	33.18093
Septiembre	33.18042
Octubre	33.18091
Noviembre	33.18034
Diciembre	33.1811

**Tabla 3.6 Bandas de Confianza para T2 de Hotelling Análisis Multivariado del Promedio Fase 1 (Corrida 2)**

Fuente: Elaboración propia

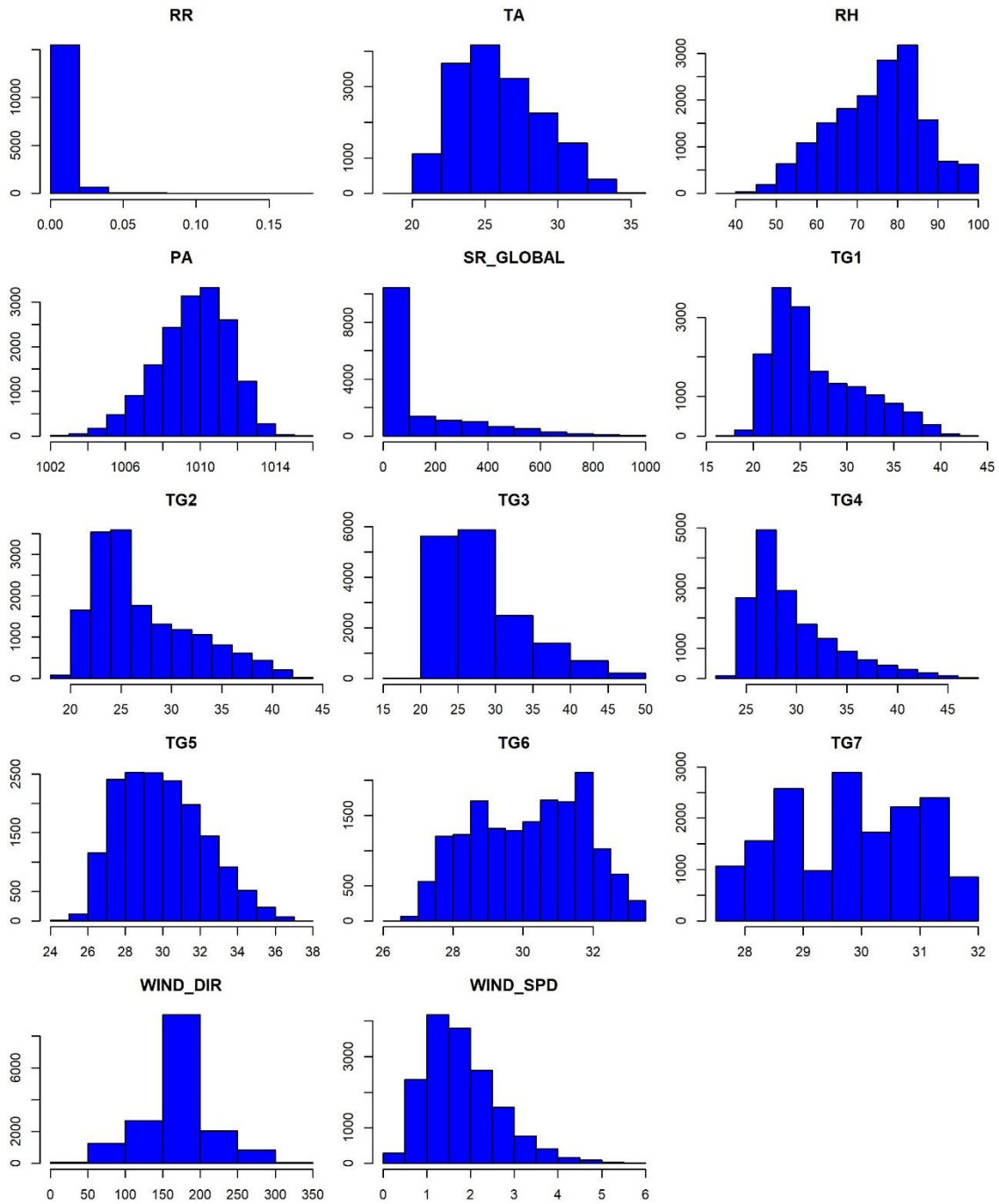
### **3.2 Análisis Descriptivo**

Las observaciones que se utilizaron para este análisis están comprendidas en un intervalo de tiempo entre mayo del 2017 y mayo del 2019.

Se tomaron en consideración variables promediadas como (RR, TA, RH, PA, SR\_GLOBAL, TG1, TG2, TG3, TG4, TG5, TG6, TG7, WIND\_DIR, WIND\_SPD).

A continuación se muestran los resultados mediante un Histograma de frecuencia del promedio horario de cada una de las variables con el fin de determinar si existe normalidad en alguna de las 14 variables que son objeto de estudio.

**Histogramas de Frecuencia del promedio horario**



**Figura 3.7 Histograma de Frecuencia del Promedio Horario**  
Fuente: Elaboración propia

La Figura 3.1 muestra histogramas de promedios de cada hora, se determinó que no existe normalidad en los datos de las variables examinadas, sin embargo, las variables TA, RH y PA muestran una cierta tendencia a normalidad.

Las demás variables muestran un comportamiento sesgado hacia la izquierda o hacia la derecha del histograma, mientras en otras gráficas, se presenta un comportamiento uniforme en los diferentes intervalos del eje x.

Para determinar si existe o no normalidad en los datos se realizó el Test de Shapiro, donde el contraste de hipótesis fue el siguiente:

**H<sub>0</sub>:** Los datos de la muestra son normales

**Vs.**

**H<sub>1</sub>:** Los datos de la muestra no son normales

Los p-value que arrojó la prueba para cada una de las variables fueron los siguientes:

Variable	Valor P
RR	9,77E-79
TA	6,08E-31
RH	2,66E-21
PA	6,14E-22
SR_GLOBAL	9,37E-71
TG1	1,62E-45
TG2	1,42E-47
TG3	1,77E-53
TG4	1,77E-51
TG5	2,00E-26
TG6	1,71E-32
TG7	1,97E-36
WIND_DIR	3,48E-34
WIND_SPD	1,24E-39

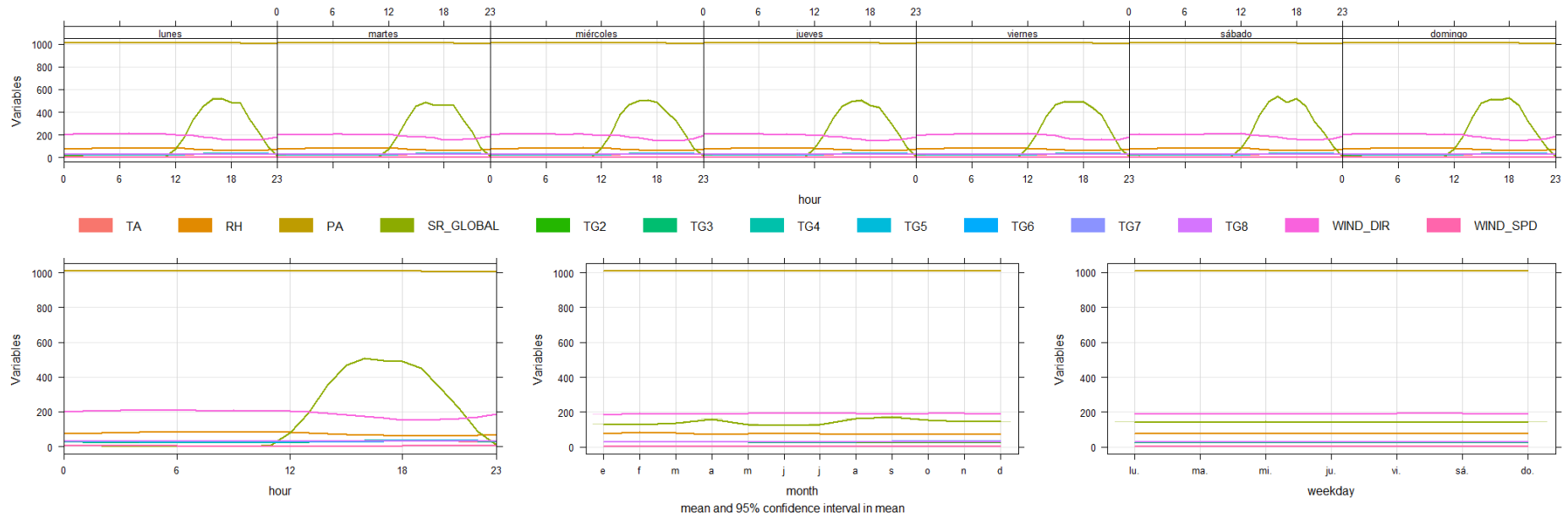
**Tabla 3.7 Valores p Test de Shapiro**

Fuente: Elaboración propia

Al tener un p-value menor que 0.05 en todos los casos se observa ausencia de normalidad en estos vectores.

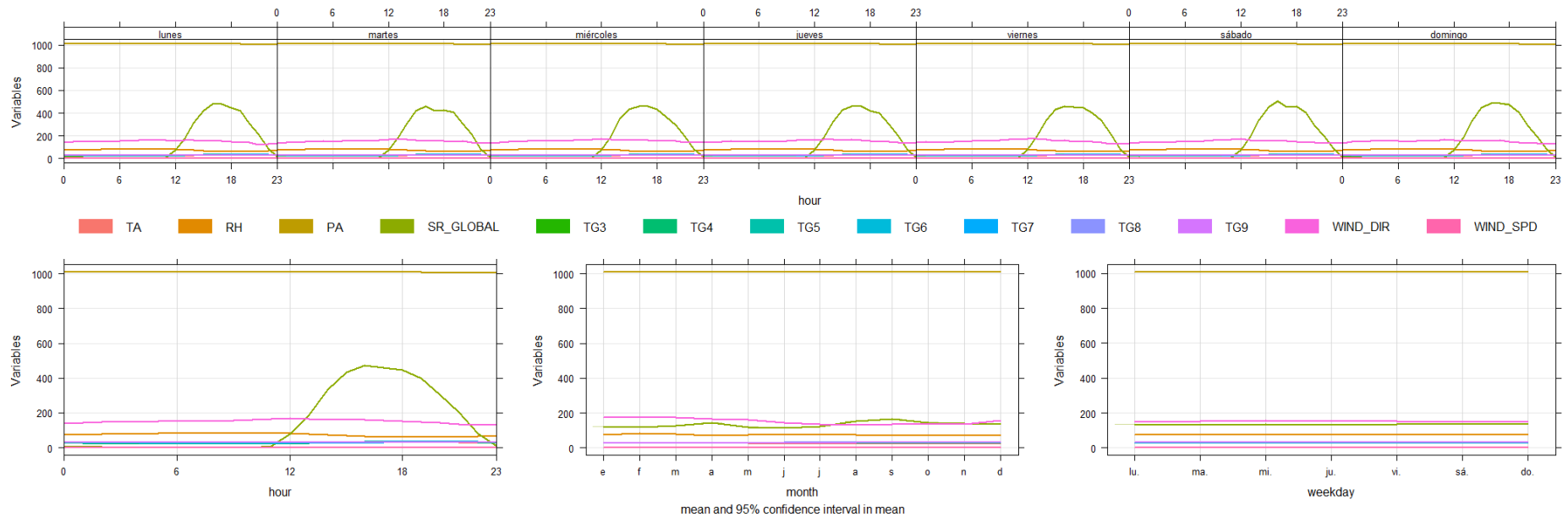
Adicionalmente en la realidad ninguna variable tiende a ser normal o es normal exactamente cómo podemos observar en la figura, es decir son distribuciones empíricas.

### 3.3 Comportamiento de las Variables en el Tiempo



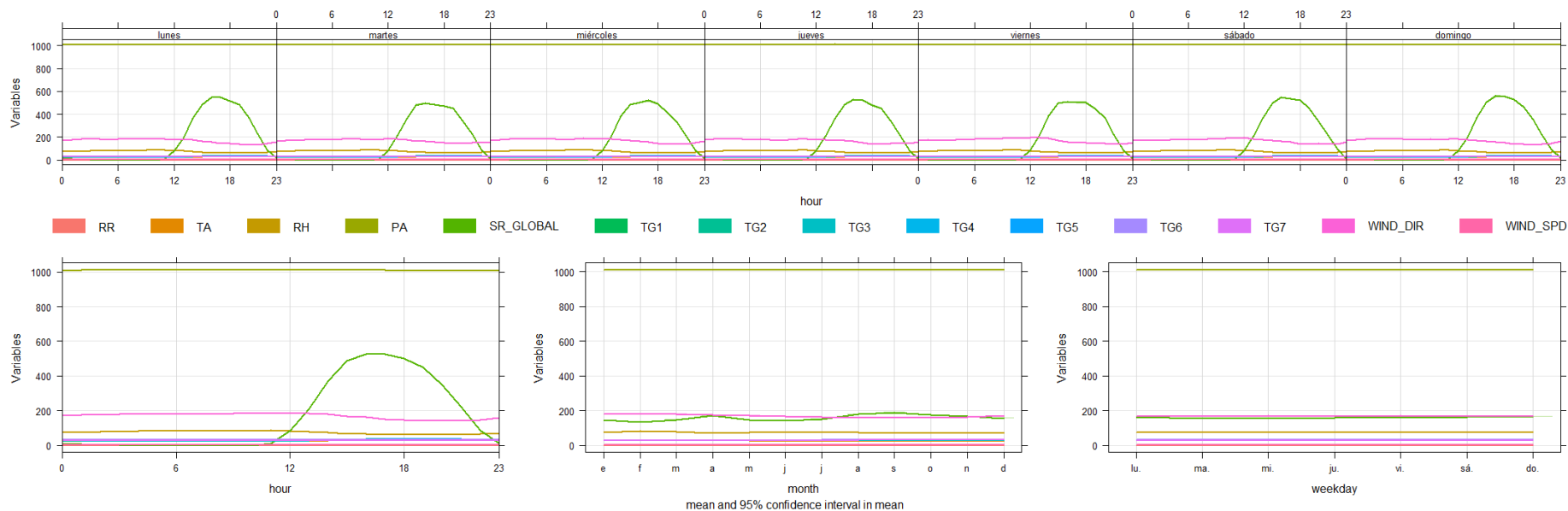
**Figura 3.8 Variación en el Tiempo para Máximos**  
Fuente: Elaboración propia

La figura 3.8 presenta la información del comportamiento de los datos de máximos de todas las variables en el tiempo. Las variables toman valores constantes todos los días de la semana con leves modificaciones en el transcurso del tiempo, SR\_GLOBAL todos los días entre las 11 y 17 horas aproximadamente se eleva, luego de esto disminuye hasta que alcanza nuevamente un comportamiento constante a partir de las 23 horas. WIND\_DIR toma valores aproximados a 200. De todas las variables, la que presenta el mayor valor constante es la presión atmosférica, la cual posee un coeficiente muy alto comparado con las demás variables.



**Figura 3.9 Variación en el Tiempo para Mínimos**  
**Fuente: Elaboración propia**

La figura 3.9 presenta la información del comportamiento de los datos de máximos de todas las variables en el tiempo. Las variables toman valores constantes todos los días de la semana con leves modificaciones en el transcurso del tiempo, SR\_GLOBAL todos los días entre las 11 y 17 horas aproximadamente se eleva, luego de esto disminuye hasta que alcanza nuevamente un comportamiento constante a partir de las 23 horas. WIND\_DIR toma valores menores a 200. De todas las variables, la que presenta el mayor valor constante es la presión atmosférica, la cual posee un coeficiente muy alto comparado con las demás variables.



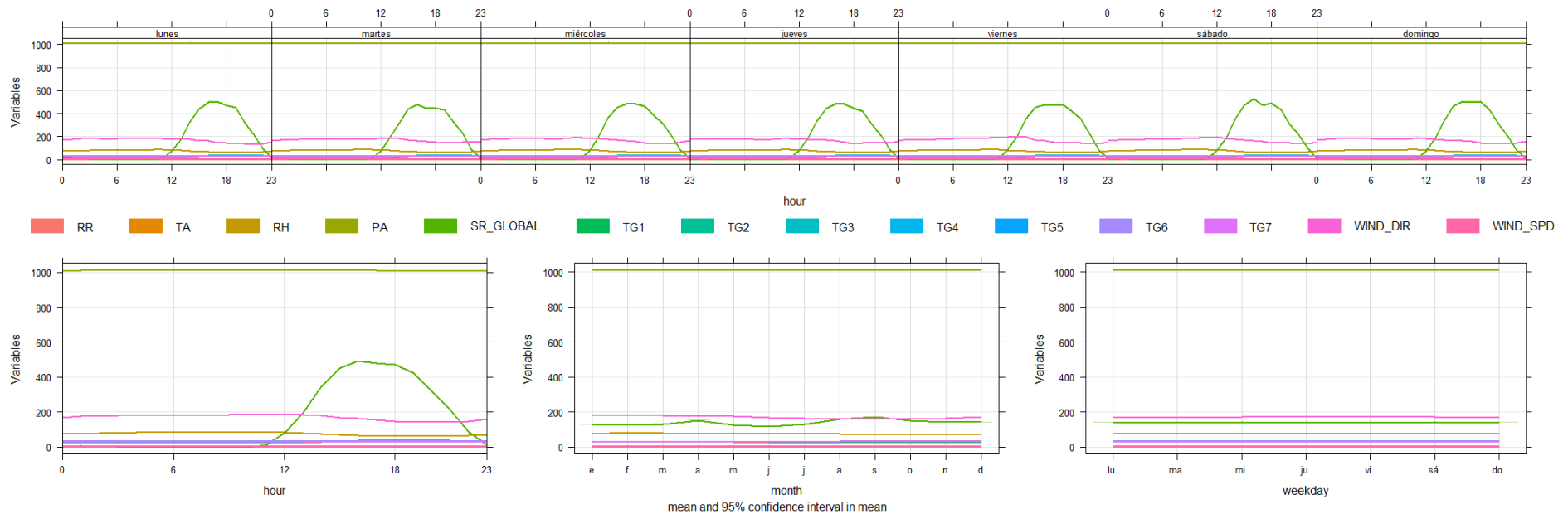
**Figura 3.10 Variación en el Tiempo Promedio Original**  
Fuente: Elaboración propia

La figura 3.10 presenta la información del comportamiento de los datos de Promedio de todas las variables en el tiempo antes de la depuración de los datos.

Las variables se comportan de un modo constante durante todos los días de la semana, SR\_GLOBAL todos los días entre las 11 y 16 horas aproximadamente se eleva, luego de esto disminuye hasta que alcanza nuevamente un comportamiento constante a partir de las 23 horas. De todas las variables, la que presenta el mayor valor constante es la presión atmosférica, la cual posee un coeficiente muy alto comparado con las demás variables.

Claramente se observa que WIND\_DIR en ciertos periodos de tiempo durante el día alcanza coeficientes de 200.





**Figura 3.11 Variación en el Tiempo Promedio Depurado**  
**Fuente: Elaboración propia**

La figura 3.11 presenta la información del comportamiento de los datos de Promedio de todas las variables en el tiempo después de la depuración de los datos, se muestra el comportamiento por horas, meses y por días.

Las variables se comportan de un modo constante durante todos los días de la semana, SR\_GLOBAL todos los días entre las 11 y 16 horas aproximadamente se eleva, luego de esto disminuye hasta que alcanza nuevamente un comportamiento constante a partir de las 23 horas. De todas las variables, la que presenta el mayor valor constante es la presión atmosférica, la cual posee un coeficiente muy alto comparado con las demás variables.

Una vez realizada la depuración se observa que WIND\_DIR no alcanza coeficientes de 200, es decir se eliminaron datos que sobrepasaban las bandas de confianza.

### 3.3.1 Análisis de las Correlaciones de las Variables

Una vez que el proceso de depuración de datos fue realizado mediante dos corridas de la Fase 1 en las cuales se procedió la eliminación de valores sospechosos, es necesario analizar las correlaciones existentes entre las variables de estudio y la intensidad de las mismas con el fin de determinar el grado de asociación que existe entre éstas.

A continuación se presentan los gráficos de correlación para cada uno de los casos estudiados máximo, mínimo y media.

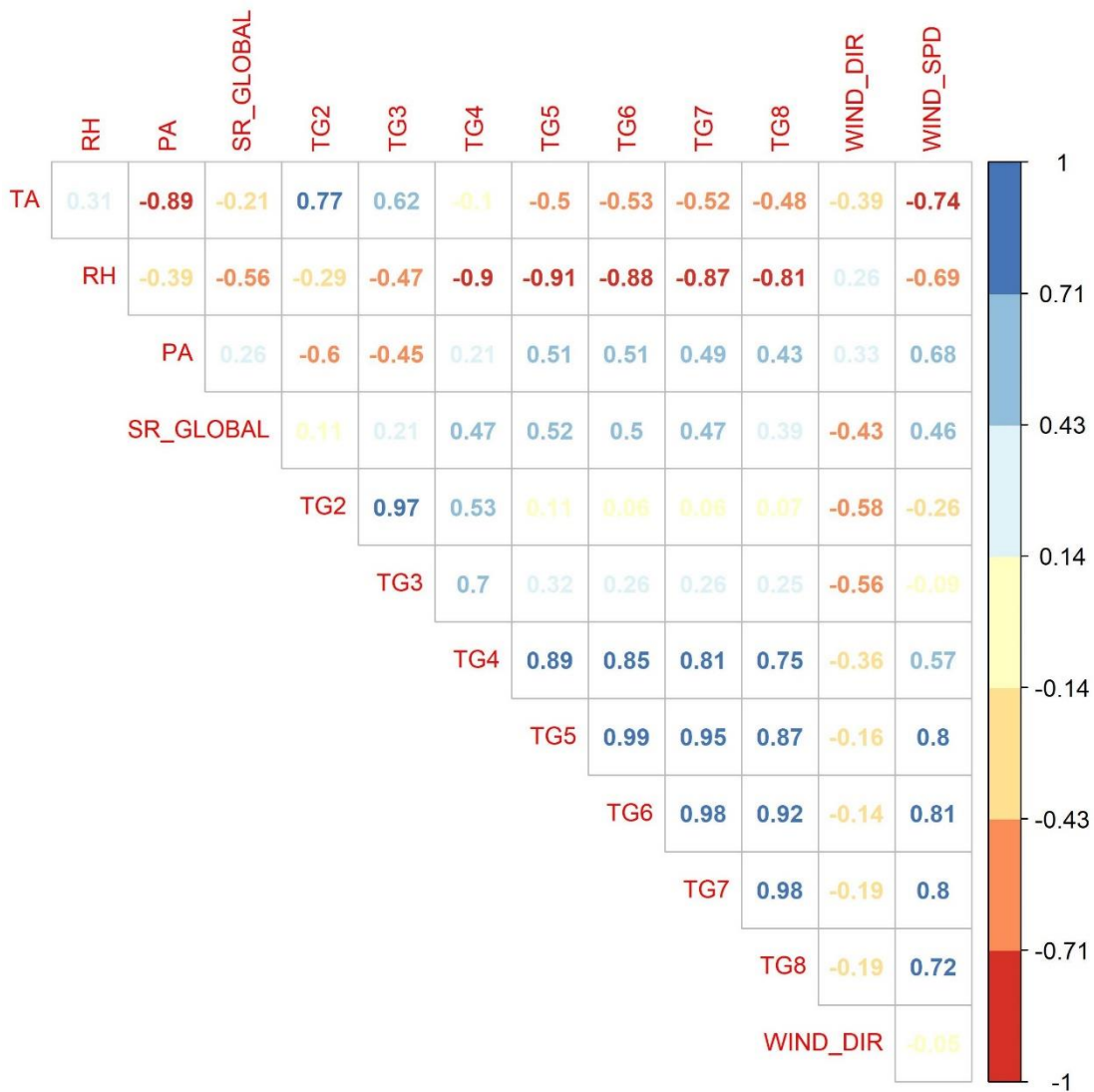


Figura 3.12 Correlaciones entre los Promedios de las Máximas Mensuales  
Fuente: Elaboración propia

La Figura 3.8 muestra la asociación entre los promedios de los valores máximos mensuales de las variables estudiadas y determina la intensidad con la cual éstas se encuentran correlacionadas.

**TA** presenta correlación positiva con **TG2** de 0.77, es decir a mayor temperatura del aire, la temperatura del suelo aumenta.

TA presenta correlación negativa con **PA** de -0.89, es decir a mayor temperatura del aire la presión atmosférica disminuye.

**RH** presenta correlación negativa con **TG4** de -0.9, con **TG5** de -0.91, con **TG6** de -0.88, con **TG7** de -0.87 y con **TG8** de -0.81.

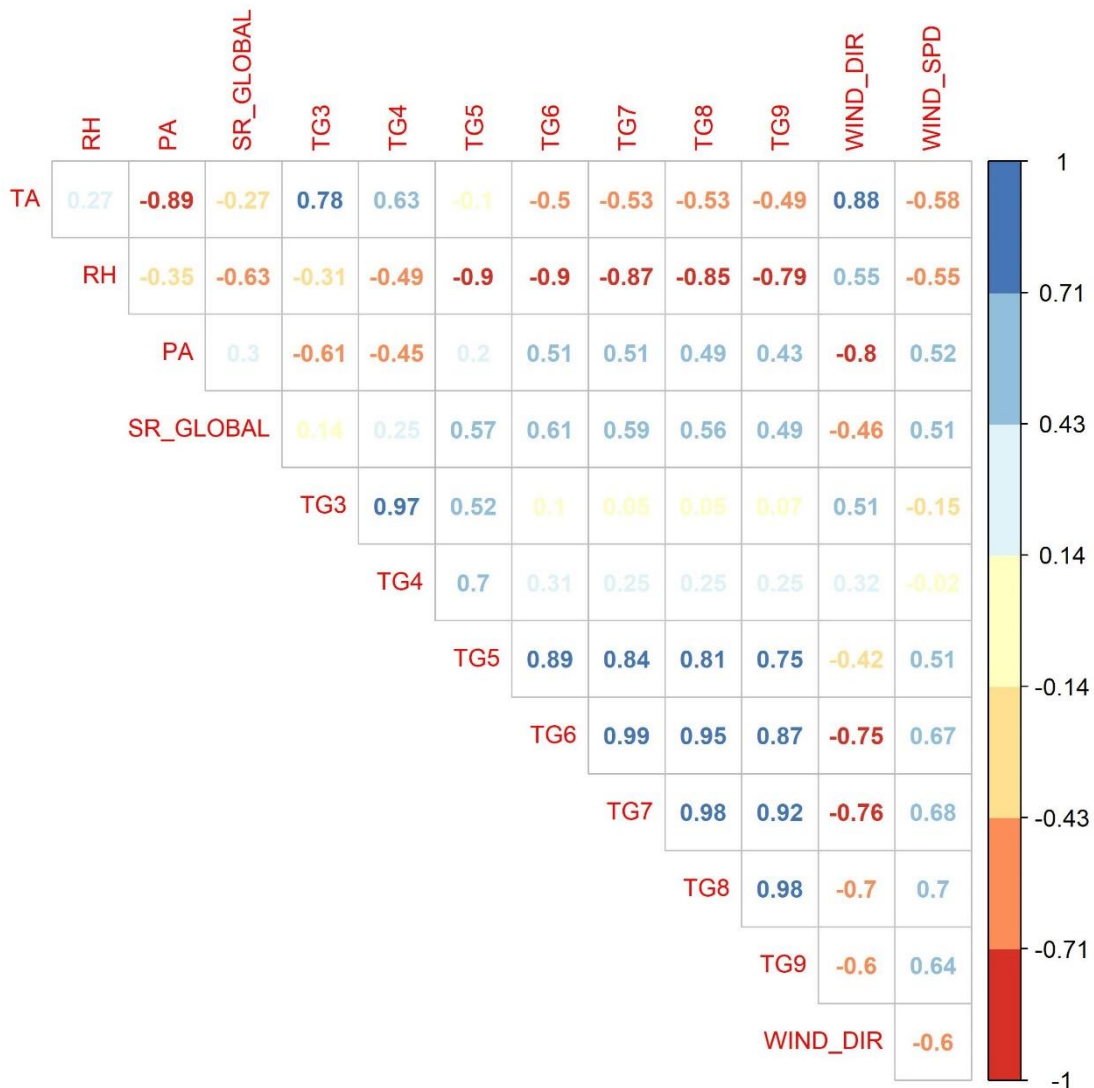
Por lo tanto si la humedad relativa aumenta, TG4, TG5, TG6, TG7, TG8 disminuyen.

**TG4** posee correlación positiva con TG5 de 0.89, con TG6 de 0.85, con TG7 de 0.81 y con TG8 de 0.75. Es decir si aumenta TG4 en consecuencia TG5, TG6, TG7 y TG8 también aumentarán.

**TG5** posee correlación positiva con TG6 de 0.99, con TG7 de 0.95 y con TG8 de 0.87, es decir si aumenta TG5 entonces TG6, TG7 y TG8 también aumentarán.

**TG6** posee correlación positiva con TG7 de 0.98 y con TG8 de 0.92, por lo tanto si TG6 aumenta, entonces TG7 y TG8 también aumentarán.

**TG7** posee correlación positiva con TG8 de 0.98, es decir que si TG7 aumenta en consecuencia TG8 aumenta.



**Figura 3.13 Correlaciones entre los Promedios de las Mínimas Mensuales**  
 Fuente: Elaboración propia

La Figura 3.9 muestra la asociación entre los promedios de los valores mínimos mensuales de las variables estudiadas y determina la intensidad con la cual éstas se encuentran correlacionadas.

**TA** presenta correlación positiva con **WIND\_DIR** de 0.88, es decir a mayor temperatura del aire, la dirección del viento aumenta. Una correlación positiva con **TG3** de 0.78, por lo tanto si TA aumenta, también aumentarán **WIND\_DIR** y **TG3**.

**TA** presenta correlación negativa con **PA** de -0.89, es decir a mayor temperatura del aire la presión atmosférica disminuye.

**RH** presenta correlación negativa con **TG5** de -0.9, con **TG6** de -0.9, con **TG7** de -0.87, con **TG8** de -0.85 y con **TG9** de -0.79.

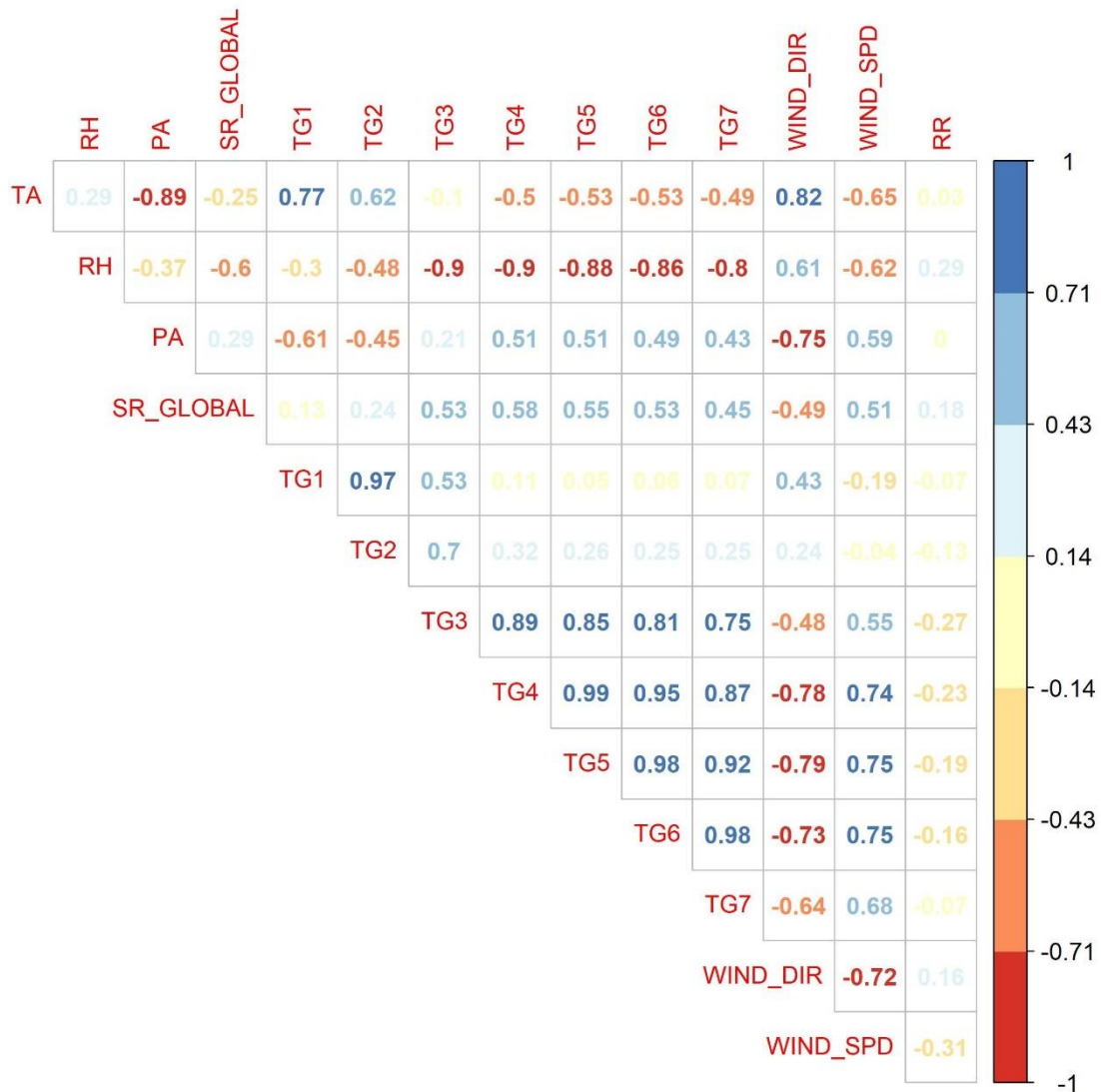
Por lo tanto si la humedad relativa aumenta, TG5, TG6, TG7, TG8 y TG9 disminuyen.

**TG5** posee correlación positiva con **TG6** de 0.89, con **TG7** de 0.84, con **TG8** de 0.81 y con **TG9** de 0.75. Es decir si aumenta TG4 también aumentarán TG5, TG6, TG7 y TG8.

**TG6** posee correlación positiva con **TG6** de 0.99, con **TG7** de 0.95 y con **TG8** de 0.87, es decir si aumenta TG5 entonces TG6, TG7 y TG8 también aumentarán.

**TG7** posee correlación positiva con **TG8** de 0.98 y con **TG9** de 0.92, por lo tanto si TG7 aumenta, entonces TG8 y TG9 también aumentarán.

**TG8** posee correlación positiva con **TG9** de 0.98, es decir que si TG7 aumenta también TG8 aumenta.



**Figura 3.14 Correlaciones entre los Promedios de las Medias Mensuales, RR acumulado**  
 Fuente: Elaboración propia

La Figura 3.10 muestra la asociación entre los promedios de las Medias Mensuales de las variables estudiadas y determina la intensidad con la cual éstas se encuentran correlacionadas.

**TA** presenta correlación positiva con **WIND\_DIR** de 0.82, es decir a mayor temperatura del aire, la dirección del viento aumenta. Una correlación positiva con **TG1** de 0.77, es decir que si TA aumenta, también aumentarán **WIND\_DIR** y **TG3**.

**TA** presenta correlación negativa con **PA** de -0.89, es decir a mayor temperatura del aire la presión atmosférica disminuye.

**RH** presenta correlación negativa con **TG3** de -0.9, con **TG4** de -0.9, con **TG5** de -0.88, con **TG6** de -0.86 y con **TG7** de -0.8.

Por lo tanto si la humedad relativa aumenta, TG5, TG6, TG7, TG8 y TG9 disminuyen.

**TG3** posee correlación positiva con **TG4** de 0.89, con **TG5** de 0.85, con **TG6** de 0.81, con **TG7** de 0.75 y con **WIND\_SPD** de 0.55. Es decir si aumenta TG4 también aumentarán TG4, TG5, TG6, TG7 y **WIND\_SPD**.

**TG4** posee correlación positiva con **TG5** de 0.99, con **TG6** de 0.95, **TG7** de 0.87 y con **WIND\_SPD** de 0.74, es decir si aumenta TG4 entonces TG5, TG6, TG7 y **WIND\_SPD** también aumentarán.

TG4 presenta correlación negativa con **WIND\_DEI** de -0.78, es decir a mayor TG4 la velocidad del viento disminuye.

**TG5** posee correlación positiva con **TG6** de 0.98 y con **TG7** de 0.92 y con **WIND\_SPD** de 0.75, por lo tanto si TG7 aumenta, entonces TG8 y TG9 también aumentarán.

TG5 presenta correlación negativa con **WIND\_DIR** de -0.79, es decir a mayor TG5 la velocidad del viento disminuye.

**TG6** posee correlación positiva con **TG7** de 0.98 y con **WIND\_SPD** de 0.75, es decir que si TG6 aumenta entonces **WIND\_SPD** aumenta también.

**TG6** presenta correlación negativa con **WIND\_DIR** de -0.79, es decir a mayor TG6 la velocidad del viento disminuye.

### **3.4 Análisis de Rendimiento para Gráficos de Control**

Como siguiente paso se medirá el rendimiento del proceso que se está desarrollando, para esto se necesita tener la base de datos libre de valores sospechosos.

Se decidió trabajar con un ARL de 370 con el objetivo de que no se generen falsas alarmas durante el proceso ya que con dicho valor se logra disminuir la tasa de falsas alarmas y solo generará una falsa alarma después de que se registren 370 observaciones en promedio.

Se simularon cambios en los parámetros de Location, Scale utilizando la distribución teórica para cada mes, se ejecutó nuevamente el procedimiento con el fin de obtener los parámetros ajustados en cada mes y se calcularon los ARL.

Mediante estas simulaciones se obtuvieron valores que brindarán información sobre el comportamiento del gráfico cuando los parámetros sufren cambios.

Los resultados se presentan a continuación.



**Análisis de Rendimientos para los Gráficos de Control de Máximos**

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Enero	BC	26,052	2,189	0	370,370	27,315	38,99
	FC	26,50	2,19	0	301,850	27,76	
	FC	27,00	2,19	0	240,286	28,26	
	FC	27,00	2,50	0	121,511	28,44	
	FC	27,00	3,00	0	54,912	28,73	
	FC	25,00	3,00	0	106,478	26,73	
	FC	27,00	2,00	0	401,849	28,15	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Febrero	BC	25,956	2,011	0	370,370	27,117	37,85
	FC	24,00	3,45	0	55,865	25,99	
	FC	26,33	1,95	0	317,541	27,46	
	FC	22,995	3,50	0	70	25,02	
	FC	25,96	2,99	0	53,880	27,68	
	FC	27,02	2,01	0	218,007	28,18	
	FC	26,67	5,00	0	9,861	29,56	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Marzo	BC	26,494	2,030	0	370,370	27,67	38,50
	FC	27,00	2,00	0	313,332	28,16	
	FC	24,42	2,46	0	308,053	25,84	
	FC	22,00	1,50	0	99,796	22,87	
	FC	28,68	2,65	0	135,955	30,21	
	FC	31,44	2,20	0	25,254	32,71	
	FC	27,00	2,00	0	313,942	28,15	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Abril	BC	26,688	2,791	-0,110	370,370	28,02	38,82
	FC	27,22	3,20	-0,11	102,850	28,75	
	FC	35,56	2,65	-0,11	4,274	36,83	
	FC	29,00	3,00	-0,11	58,400	30,43	
	FC	26,50	2,99	-0,11	242,742	27,93	
	FC	29,63	3,00	-0,11	42,34	27,07	
	FC	30,00	2,87	-0,11	43,067	31,37	

**Tabla 3.8 Análisis de Rendimientos para los Gráficos de Control de Máximos (Enero-Abril)**

Fuente: Elaboración propia

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Mayo	BC	25,838	2,423	-0,137	370,370	26,94	35,65
	FC	24,17	2,89	-0,14	313,47	25,48	
	FC	26,32	2,56	-0,14	155,47	27,49	
	FC	26,11	2,72	-0,14	120,07	27,35	
	FC	22,88	2,85	-0,14	55,02	24,18	
	FC	30,31	1,36	-0,14	287,08	30,93	
	FC	27,58	2,96	-0,14	30,91	28,93	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Julio	BC	23,381	1,993	0,136	370,370	24,839	41,50
	FC	23,45	2,00	0,14	359,84	24,91	
	FC	26,56	5,23	0,14	11,64	30,39	
	FC	22,01	2,23	0,14	316,38	23,64	
	FC	18,36	2,64	0,14	320,20	20,29	
	FC	29,24	2,24	0,14	60,63	30,87	
	FC	20,40	3,48	0,14	83,23	22,95	

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Junio	BC	24,255	2,200	0,000	370,370	25,52	37,26
	FC	23,67	2,57	0	257,82	25,16	
	FC	22,35	2,56	0	336,22	23,83	
	FC	25,00	2,16	0	289,05	26,25	
	FC	26,31	2,79	0	51,22	27,92	
	FC	29,01	2,43	0	30,41	30,41	
	FC	29,17	1,60	0	158,34	30,09	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Agosto	BC	23,467	2,297	0,129	370,370	25,127	43,86
	FC	27,33	2,70	0,13	91,50	29,28	
	FC	27,96	2,27	0,13	147,13	29,60	
	FC	15,65	3,50	0,13	250,68	18,18	
	FC	20,95	2,65	0,13	329,55	22,87	
	FC	25,35	2,15	0,13	324,17	26,90	
	FC	26,00	2,90	0,13	93,27	28,10	

**Tabla 3.9 Análisis de Rendimientos para los Gráficos de Control de Máximos (Mayo-Agosto)**

Fuente: Elaboración propia

**Análisis de Rendimientos para los Gráficos de Control de Máximos**

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Septiembre	BC	23,567	2,084	0,356	370,37	25,89	65,73
	FC	26,11	3,10	0,36	123,87	29,56	
	FC	26,50	2,63	0,36	176,77	29,43	
	FC	22,65	3,54	0,36	194,11	26,59	
	FC	14,25	2,70	0,36	320,10	17,26	
	FC	20,00	5,25	0,36	53,31	25,84	
	FC	18,56	3,42	0,36	147,99	22,37	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Noviembre	BC	24,142	2,237	0,137	370,37	25,78	44,51
	FC	22,46	2,50	0,14	326,00	24,29	
	FC	27,24	2,01	0,14	293,98	28,71	
	FC	21,96	2,98	0,14	181,19	24,14	
	FC	16,46	3,95	0,14	143,03	19,35	
	FC	30,46	2,00	0,14	137,98	31,92	
	FC	35,46	3,00	0,14	13,01	37,65	

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Octubre	BC	23,787	1,988	0,289	370,37	25,72	54,96
	FC	15,47	2,60	0,29	350,07	18,00	
	FC	25,12	3,12	0,29	97,93	28,16	
	FC	21,33	5,45	0,29	34,91	26,63	
	FC	24,76	2,10	0,29	290,50	26,81	
	FC	23,00	2,96	0,29	133,81	25,88	
	FC	26,45	3,13	0,29	86,66	29,50	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Diciembre	BC	25,343	2,544	0	370,37	26,81	40,39
	FC	23,25	2,94	0	340,04	24,95	
	FC	32,26	2,00	0	58,79	33,41	
	FC	22,54	3,76	0	115,69	24,71	
	FC	27,36	2,30	0	289,23	28,68	
	FC	24,60	3,65	0	75,73	26,71	
	FC	30,46	5,12	0	7,45	33,42	

**Tabla 3.10 Análisis de Rendimientos para los Gráficos de Control de Máximos (Septiembre-Diciembre)**

Fuente: Elaboración propia

**Análisis de Rendimientos para los Gráficos de Control de Mínimos**

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Enero	BC	25,26	1,90	0	370,37	26,36	36,49
	FC	23,24	2,30	0	317,81	24,57	
	FC	27,63	1,99	0	86,37	28,78	
	FC	35,42	2,50	0	2,08	36,87	
	FC	22,13	2,99	0	122,45	23,85	
	FC	26,60	2,20	0	90,07	27,87	
	FC	29,72	1,54	0	81,36	30,61	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Febrero	BC	25,18	1,66	0	370,37	26,14	35,01
	FC	24,19	1,87	0	325,45	25,27	
	FC	21,71	2,75	0	126,25	23,30	
	FC	15,91	3,94	0	127,79	18,19	
	FC	32,95	2,00	0	3,32	34,11	
	FC	27,22	2,54	0	21,94	28,69	
	FC	26,72	3,45	0	11,54	28,72	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Marzo	BC	25,65	1,68	0	370,37	26,62	35,60
	FC	25,54	2,80	0	102,69	27,16	
	FC	36,27	1,86	0	3,83	37,34	
	FC	28,49	1,99	0	152,43	29,64	
	FC	28,44	2,76	0	38,72	30,03	
	FC	32,54	2,63	0	10,16	34,06	
	FC	17,25	3,88	0	240,49	19,49	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Abril	BC	25,70	2,44	-0,13	370,37	26,83	35,79
	FC	26,11	2,80	-0,13	32,32	27,40	
	FC	25,24	4,45	-0,13	17,44	27,29	
	FC	27,17	3,65	-0,13	17,34	28,85	
	FC	21,96	5,89	-0,13	16,92	24,68	
	FC	33,27	0,65	-0,13	230,38	33,56	
	FC	16,24	42,54	-0,13	2,16	35,91	

**Tabla 3.11 Análisis de Rendimientos para los Gráficos de Control de Mínimos (Enero-Abril)**

Fuente: Elaboración propia

**Análisis de Rendimientos para los Gráficos de Control de Mínimos**

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Mayo	BC	25,03	2,16	-0,15	370,37	25,99	33,46
	FC	32,98	2,30	-0,15	1,80	34,00	
	FC	27,24	3,52	-0,15	8,34	28,80	
	FC	21,57	4,76	-0,15	23,63	23,69	
	FC	26,97	3,13	-0,15	12,59	28,36	
	FC	27,64	2,89	-0,15	11,55	28,93	
	FC	34,27	3,56	-0,15	1,40	35,85	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Julio	BC	22,60	1,69	0,10	370,37	23,76	36,27
	FC	23,20	1,92	0,10	180,30	24,52	
	FC	26,27	2,80	0,10	21,73	28,19	
	FC	18,89	2,37	0,10	244,92	20,51	
	FC	21,56	2,53	0,10	98,38	23,29	
	FC	27,07	2,13	0,10	36,81	28,53	
	FC	33,27	1,52	0,10	6,61	34,31	

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Junio	BC	23,46	1,94	0	370,37	24,58	34,91
	FC	27,33	2,47	0	22,06	28,75	
	FC	28,24	2,16	0	22,35	29,49	
	FC	24,25	2,65	0	56,08	25,78	
	FC	30,55	1,56	0	17,03	31,45	
	FC	26,65	4,77	0	6,17	29,40	
	FC	34,65	2,85	0	1,67	36,29	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Agosto	BC	22,62	1,93	0,12	370,37	23,99	39,32
	FC	23,16	2,21	0,12	185,58	24,74	
	FC	25,65	2,85	0,12	44,14	27,69	
	FC	27,62	2,20	0,12	60,68	29,19	
	FC	31,59	0,55	0,12	3573,98	31,98	
	FC	27,88	1,44	0,12	258,69	28,90	
	FC	35,30	2,20	0,12	5,75	36,86	

**Tabla 3.12 Análisis de Rendimientos para los Gráficos de Control de Mínimos (Mayo-Agosto)**

Fuente: Elaboración propia

**Análisis de Rendimientos para los Gráficos de Control de Mínimos**

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Septiembre	BC	22,71	1,72	0,28	370,37	24,37	49,14
	FC	27,02	2,98	0,28	54,98	29,89	
	FC	15,65	3,12	0,28	138,32	18,65	
	FC	23,67	2,35	0,28	143,12	25,93	
	FC	26,05	2,36	0,28	109,29	28,32	
	FC	24,25	2,56	0,28	106,63	26,72	
	FC	13,70	5,23	0,28	44,33	18,73	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Noviembre	BC	23,26	1,88	0,11	370,37	24,57	38,74
	FC	35,92	2,16	0,11	3,92	37,42	
	FC	27,39	4,25	0,11	11,00	30,34	
	FC	22,65	2,98	0,11	72,15	24,72	
	FC	29,70	2,16	0,11	32,30	31,20	
	FC	27,66	2,95	0,11	24,25	29,70	
	FC	24,93	2,56	0,11	71,85	26,71	

Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Octubre	BC	22,96	1,68	0,21	105263,20	24,35	42,42
	FC	31,25	1,72	0,21	62,53	32,67	
	FC	27,20	1,56	0,21	208,39	28,50	
	FC	22,62	2,35	0,21	132,75	24,57	
	FC	28,68	1,72	0,21	361,83	30,11	
	FC	24,91	2,98	0,21	47,55	27,38	
	FC	24,19	2,34	0,21	105,16	26,13	
Mes	Estado	Location	Scale	Shape	ARL	Media BC	Umbral
Diciembre	BC	24,48	2,32	-0,08	370,37	25,65	35,46
	FC	23,78	2,77	-0,08	167,08	25,18	
	FC	25,66	2,97	-0,08	45,96	27,16	
	FC	29,58	3,34	-0,08	7,18	31,26	
	FC	24,48	3,24	-0,08	52,19	26,11	
	FC	32,55	1,24	-0,08	14,04	33,17	
	FC	24,65	4,00	-0,08	21,30	24,65	

**Tabla 3.13 Análisis de Rendimientos para los Gráficos de Control de Mínimos (Septiembre-Diciembre)**

Fuente: Elaboración propia

### 3.6 Discusión y Análisis Final

Valores excedentes para Gráfico de Control de Temperaturas Máximas del Aire Fase 1 (Corrida 1)					
Enero	Junio	Julio	Septiembre	Octubre	Noviembre
7	15	164	41	354	6

Valores excedentes para Gráfico de Control de Temperaturas Mínimas del Aire Fase 1 (Corrida 1)										
Enero	Febrero	Marzo	Abril	Mayo	Junio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
671	361	373	1451	1903	65	103	2	4	241	483

Valores excedentes para Gráfico de Control de $T^2$ Hotelling Análisis Multivariado del Promedio Fase 1 (Corrida 1)											
Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
4200	4387	3891	3717	4024	4904	5710	5455	5675	5960	4817	4345

**Tabla 3.14 Valores que exceden los Límites de Control en la Corrida 1**

Fuente: Elaboración propia

En la tabla 3.9 se registraron las cantidades de valores que excedieron las bandas de confianza para la Corrida 1 (Fase 1).

Se puede apreciar que para los gráficos de control por máximo el mes de Octubre fue el mes que registró mayor cantidad de observaciones por encima de la banda de confianza.

Se puede apreciar que para los gráficos de control por mínimo el mes de Mayo fue el mes que registró mayor cantidad de observaciones por debajo de la banda de confianza.

Para  $T^2$  de Hotelling el mes que registró más observaciones que excedieron la banda de confianza fue el mes de Octubre.

Valores excedentes para Gráfico de Control de Temperaturas Máximas del Aire Fase 1 (Corrida 2)			
Julio	Septiembre	Octubre	Noviembre
156	40	218	4

Valores excedentes para Gráfico de Control de Temperaturas Mínimas del Aire Fase 1 (Corrida 2)									
Enero	Febrero	Marzo	Abril	Mayo	Junio	Agosto	Septiembre	Noviembre	Diciembre
205	185	128	387	303	14	15	1	38	33

Valores excedentes para Gráfico de Control de $T^2$ Hotelling Análisis Multivariado del Promedio Fase 1 (Corrida 2)											
Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
60	60	60	60	60	60	60	60	60	60	60	60

**Tabla 3.15 Valores que exceden los Límites de Control en la Corrida 2**

Fuente: Elaboración propia

En la tabla 3.10 se registraron las cantidades de valores que excedieron las bandas de confianza para la Corrida 2 (Fase 1).

Se puede apreciar que para los gráficos de control por máximo el mes de Octubre fue el mes que registró mayor cantidad de observaciones por encima de la banda de confianza.

Se puede apreciar que para los gráficos de control por mínimo el mes de Abril fue el mes que registró mayor cantidad de observaciones por debajo de la banda de confianza.

Para  $T^2$  de Hotelling en esta corrida todos los meses se registró igual cantidad de observaciones que excedieron la banda de confianza.



En Fase 1 (Corrida 1) se observa con claridad que tanto para el Gráfico de Control de Temperaturas Máximas del Aire como para el Gráfico de Control de T2 Hotelling Análisis Multivariado del Promedio, el mes en que se registraron mayor número de observaciones por encima de las bandas de confianza fue el mes de Octubre.

En esta etapa el Gráfico de Control de T2 Hotelling detecta una cantidad de observaciones fuera de control desde 3717 observaciones en adelante por encima de las bandas de confianza, lo que significa que en esta etapa el proceso se encuentra sumamente descontrolado.

En Fase 1 (Corrida 2) se observa de igual manera que para el Gráfico de Control de Temperaturas Máximas del Aire, el mes en que se registraron mayor número de observaciones por encima de las bandas de confianza fue el mes de Octubre. En esta etapa el Gráfico de Control de T2 Hotelling detecta una cantidad de observaciones fuera de control de 60 observaciones por encima de las bandas de confianza para todos los meses, lo que significa que en esta etapa el proceso ha mejorado considerablemente en comparación a la primera corrida.

El proceso de depuración de datos consigue buenos resultados en tan solo dos corridas de la Fase 1, y dado que los valores que exceden las bandas de confianza han disminuido se puede considerar que el proceso se acerca a un punto dentro de lo aceptable.

# CAPÍTULO 4

## 4. CONCLUSIONES Y RECOMENDACIONES

### 4.1 Conclusiones

Del análisis realizado para este estudio se concluyó que:

- Fue posible establecer umbrales estadísticos que permitieron la eliminación de valores sospechosos mediante la imputación de máximos y mínimos en dos corridas de la Fase 1.
- Se logró discriminar o eliminar valores sospechosos mediante la imputación por media en la Fase 1.
- Mediante las dos corridas de la Fase 1 realizadas, se descartaron aproximadamente el 5% de datos los datos en bruto.
- Se puede concluir que el proceso está bajo control dado que en ninguno de los casos los valores de los ARL de temperatura máxima superan el valor de

$$\frac{1}{1 - \alpha} = \frac{1}{1 - 0,9973} = 370.3704$$

## 4.2 Recomendaciones

- Para la correcta depuración de datos se sugiere trabajar con un intervalo mayor a 5 años si se estuviera trabajando con datos minutales de estaciones meteorológicas automáticas.
- Se recomienda realizar una comparación de la base de datos depurada con otras estaciones cercanas para explorar si los datos son equivalentes.
- Se sugiere explorar otros métodos que permitan agregarle mayor riqueza al análisis dado que el estudio de valores extremos no es la única metodología que permite realizar control de calidad a los datos.
- Se sugiere implementar un gráfico de control combinado de máximos y mínimos para establecer la relación conjunta entre estos para luego poder comparar ese gráfico con el  $T^2$  de Hotelling.

## 5. BIBLIOGRAFÍA

- Aprendiendo Meteorología*. (2001). Obtenido de La Meteorología:  
<http://web.archive.org/web/20080820122403/http://nimbus.com.uy/weather/pdf/cap2.pdf>
- Araya López, J. L. (14 de Junio de 2010). Resultados de un control de calidad de datos de temperatura superficial. *Tecnología en Marcha*, 33-49.
- Archive Internet. (s.f.). *Wayback Machine*. Obtenido de <https://web.archive.org/web/20090522121225/http://www.automaticweatherstation.com/index.html>
- Burke, E., Perry, R., & Brown, S. (25 de June de 2010). An extreme value analysis of UK drought and projections of change in the future. (Elsevier, Ed.) *Journal of Hydrology*. Recuperado el 06 de Agosto de 2019, de <https://www.sciencedirect.com/science/article/pii/S0022169410002349?via%3Dihub>
- Coles, S. (2001). Una introducción al modelado estadístico de valores extremos. *Springer-Verlag*, 208.
- Cruz-Saco Vilela, R., Natividad Alvarado, J., & Espinoza Paredes, R. (s.f.). *Software para el control de calidad y validación de datos meteorológicos y generación de rosa de vientos*. Lima: Universidad Dacional de Ingeniería.
- Cuatrecasas Arbós, L., & González Babón, J. (2017). *Gestión integral de la calidad: Implantación, control y certificación*. Barcelona: Profil Editorial I. .
- Fisher, R., & Tippett, L. (1928). *Limiting forms of the frequency distribution of the largest and smallest member of a sample*. *Proc . Cambridge Phil. Soc.* 24: 180-190, doi:10.1017/s0305004100015681.
- Fréchet, M. (1927). *Sur la loi de probabilité de l'écart maximum*. *Ann. .Soc. Polon. Math, Cracow.* 6, 93.
- Gnedenko, B. V. (1943). *Sur la distribution limite du terme maximum d'une série aléatoire*. *Ann. Math.* 4a, 423-453.

- Guayaquil es mi Destino ®. (2019). Recuperado el 25 de 07 de 2019, de <https://www.guayaquilesmidestino.com/es/descubre-guayaquil/generalidades/clima>
- Gumbel, E. (1958). *Statistics of extremes*. New York: Columbia University Press.
- Gumbel, E. J. (1934). *Les moments des distributions finales de la première et de la dernière valeur. Comptes Rendus de l'Académie des Sciences, Paris, 198, 141-143.*
- Gumbel, E. J. (1935). *Les valeurs extremes des distributions statistiques. Annales de l'institut Henri Poincaré, 5, 115-158.*
- Hamel, G. (1 de Febrero de 2018). *Geniolandia*. Obtenido de <https://www.geniolandia.com/13181757/que-es-la-velocidad-del-viento>
- IBM®. (s.f.). Obtenido de [https://www.ibm.com/support/knowledgecenter/es/SSEP7J\\_10.1.1/com.ibm.swg.ba.cognos.ug\\_cr\\_rptstd.10.1.1.doc/c\\_id\\_obj\\_desc\\_tables.html](https://www.ibm.com/support/knowledgecenter/es/SSEP7J_10.1.1/com.ibm.swg.ba.cognos.ug_cr_rptstd.10.1.1.doc/c_id_obj_desc_tables.html)
- Jenkinson, A. F. (1955). *The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Quart. J. Roy. Meteo. Soc. 1(81): 158-171.*
- Llinás Solano, H., & Rojas Álvarez, C. (2006). *Estadística descriptiva y distribuciones de probabilidad*. Barranquilla: Ediciones Uninorte.
- Luceño Vázquez, A., & González Ortiz, F. (2015). *Métodos estadísticos para medir, describir y controlar la variabilidad*. Santander, Cantabria: Universidad Cantabria.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *National Institute of Science of India, 49-55.*
- Martínez Vaca, W. (s.f.). *Estadística Descriptiva con énfasis en Salud Pública*. La Hoguera.
- Méndez Muñiz, Javier María; Cuervo García, Rafael; Bureau Veritas Formación. (2010).
- Montgomery, D. C. (2006). *Introduction to Statistical Quality Control* (6 ed.). USA: Wiley.
- Montgomery, D. C. (2006). *Introduction to Statistical Quality Control*. 198-199. Recuperado el 08 de 2019

- National Weather Services. (s.f.). *Origin of Wind*. Obtenido de <https://www.weather.gov/jetstream/wind>
- Oakland, J. S. (2003). *Statistical Process Control* (Fifth Edition ed.). Burlington, Massachusetts, EE.UU.: Butterworth-Heinemann. Recuperado el 04 de 08 de 2019, de <http://diliev.com/Home/materiali/KHEA/referati/6812268-Statistical-Process-Control-eBook-VG.pdf>
- Oosterbaan, R. (1994). Chapter 6 Frequency and Regression Analysis of Hydrologic Data. *Drainage Principles and Applications*, 16. Recuperado el 06 de 08 de 2019, de [Commons.wikimedia.org/wiki/File:Extrema\\_example\\_es.svg](https://commons.wikimedia.org/wiki/File:Extrema_example_es.svg)
- Pérez Porto, J., & Gardey, A. (2016). *Definicion.de*. Obtenido de <https://definicion.de/temperatura-atmosferica/>
- Pérez Porto, J., & Merino, M. (2012). *Definicion.de*. Obtenido de <https://definicion.de/presion-atmosferica/>
- Ruiz-Falcó Rojas, A. (2006). *Control Estadístico de Procesos (Apuntes)*. Madrid: Universidad Pontificia COMILLAS. Recuperado el 04 de 08 de 2019, de <https://web.cortland.edu/matresearch/ControlProcesos.pdf>
- Seber, G. A. (1984). *Multivariate Observations*. Wiley Series in probability and mathematical statistics.
- Steven M. Babin, M. P. (25 de Enero de 2016). *Water Vapor Myths: A Brief Tutorial*. Obtenido de <https://web.archive.org/web/20160125031338/http://www.atmos.umd.edu/~stevenb/vapor/>
- Tippett, L., & Fisher, R. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180-190. doi:10.1017/S0305004100015681
- Varela Ledesma, N., Díaz Pereira, Y., & Pomares, I. (2015). Control de calidad de las mediciones de temperatura seca y humedad relativa en la estación meteorológica Sancti Spíritus. *Ciencias de la Tierra y el Espacio*, 29-39.

- Veiga, H., Herrera , N., Skansi, M., & Podestá, G. (2014). *Descripción de controles de calidad de datos climáticos diarios implementados por el Centro Regional del Clima para el Sur de América del Sur .*
- Weckerly , D., Mendenhall III, W., & Scheaffer, R. (2010). *Estadística matemática con aplicaciones*. México, D.F.: Cengage Learning Editores, S.A.
- Weibull, W. (1951). *A statistical distribution function of wide applicability. J. Appl. Mech.-Trans. ASME 18 (3): 293-297.*
- Wikipedia. (16 de Julio de 2019). Obtenido de [https://es.wikipedia.org/wiki/Precipitaci%C3%B3n\\_\(meteorolog%C3%ADa\)](https://es.wikipedia.org/wiki/Precipitaci%C3%B3n_(meteorolog%C3%ADa))
- Wikipedia®. (9 de Julio de 2019). *Wikipedia*. Obtenido de [https://es.wikipedia.org/wiki/Estaci%C3%B3n\\_meteorol%C3%B3gica](https://es.wikipedia.org/wiki/Estaci%C3%B3n_meteorol%C3%B3gica)