

	GESTIÓN DE RECURSOS Y SERVICIOS BIBLIOTECARIOS	Código	FO-SB-12/v0
	ESQUEMA HOJA DE RESUMEN	Página	1/106

RESUMEN TRABAJO DE GRADO

AUTOR(ES):

NOMBRE(S): CARLOS FEDERICO APELLIDOS: CALDERÓN DELGADO

NOMBRE(S): LEIDY YERALDIN APELLIDOS: RAMIREZ FLOREZ

FACULTAD: INGENIERÍA

PLAN DE ESTUDIOS: INGENIERIA CIVIL

DIRECTOR:

NOMBRE(S): GUSTAVO ADOLFO APELLIDOS: CARRILLO SOTO

NOMBRE(S): _____ APELLIDOS: _____

TÍTULO DEL TRABAJO (TESIS): RECONSTRUCCIÓN DE SERIES DE TIEMPO DE TEMPERATURA DIARIA MEDIANTE REGRESIÓN LINEAL MULTIVARIADA EN NORTE DE SANTANDER

RESUMEN

El trabajo trata sobre la Reconstrucción de series de tiempo de temperatura diaria mediante regresión lineal multivariada en Norte de Santander. Se propone, Reconstruir las series de tiempo de temperatura diaria mediante regresión lineal multivariada en el departamento de Norte de Santander. Para ello, se analiza las series temporales de temperatura diaria para estaciones en Norte de Santander. Seguido de, aplicar procedimientos estándar para determinar datos faltantes en las series seleccionadas; y finalmente, presentar la información recolectada usando una herramienta tecnológica que permita reconstruir las series de temperatura, apoyando la iniciativa del Observatorio Hidro-Climatológico (HCO-UFPS) liderada por el grupo de investigación HYDROS-UFPS. El tipo de investigación de este proyecto según el nivel corresponde a descriptivo. La población, son las 8858 estaciones meteorológicas de las cuales 294 se encuentran en el Departamento Norte de Santander. Y para la muestra, se selecciona las estaciones meteorológicas del IDEAM con información útil que permita la transformación de la data. Para el procesamiento de la información se utiliza la página principal del IDEAM para realizar la consulta y descarga de la data, obtenidos por medio de las estaciones Hidrometeorológicas del Departamento Norte de Santander, que permitan suministrar datos para la investigación descriptiva - documental del presente proyecto.

PALABRAS CLAVE: Series de tiempo, metereologica, Hydros-climatica, UFPS, regresion lineal.

CARACTERÍSTICAS:

PÁGINAS: 102 **PLANOS:** _____ **ILUSTRACIONES:** _____ **CD ROOM:** 1

Elaboró		Revisó		Aprobó	
Equipo Operativo del Proceso		Comité de Calidad		Comité de Calidad	
Fecha	24/10/2014	Fecha	05/12/2014	Fecha	05/12/2014

COPIA NO CONTROLADA

RECONSTRUCCIÓN DE SERIES DE TIEMPO DE TEMPERATURA DIARIA MEDIANTE
REGRESIÓN LINEAL MULTIVARIADA EN NORTE DE SANTANDER

CARLOS FEDERICO CALDERÓN DELGADO

LEIDY YERALDIN RAMIREZ FLOREZ

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERIA

PLAN DE ESTUDIOS DE INGENIERIA CIVIL

SAN JOSE DE CUCUTA

2020

RECONSTRUCCIÓN DE SERIES DE TIEMPO DE TEMPERATURA DIARIA MEDIANTE
REGRESIÓN LINEAL MULTIVARIADA EN NORTE DE SANTANDER

CARLOS FEDERICO CALDERÓN DELGADO

LEIDY YERALDIN RAMIREZ FLOREZ

Trabajo de grado presentado como requisito para optar el título de:

Ingeniero Civil

Director:

PhD. GUSTAVO ADOLFO CARRILLO SOTO

UNIVERSIDAD FRANCISCO DE PAULA SANTANDER

FACULTAD DE INGENIERIA

PLAN DE ESTUDIOS DE INGENIERIA CIVIL

SAN JOSE DE CUCUTA

2020

ACTA DE SUSTENTACION DE TRABAJO DE GRADO

FECHA: 14 DE MAYO DE 2020 HORA: 2:00 p. m.

LUGAR: VIDEO CONFERENCIA GOOGLE MEET

PLAN DE ESTUDIOS: INGENIERIA CIVIL

TITULO DE LA TESIS: "RECONSTRUCCIÓN DE SERIES DE TIEMPO DE TEMPERATURA DIARIA MEDIANTE REGRESIÓN LINEAL MULTIVARIADA EN NORTE DE SANTANDER".


JURADOS: ING. NELSON JAVIER CELY CALIXTO
ING. JUAN CARLOS SAYAGO ORTEGA

DIRECTOR: Ph.D. GUSTAVO ADOLFO CARRILLO SOTO.

NOMBRE DE LOS ESTUDIANTES:	CODIGO	CALIFICACION	
		NUMERO	LETRA
CARLOS FEDERICO CALDERÓN DELGADO	2110191	4,5	CUATRO, CINCO

MERITORIA


ING. NELSON JAVIER CELY CALIXTO


ING. JUAN CARLOS SAYAGO ORTEGA


Vo. Bo. JAVIER ALFONSO CARDENAS GUTIERREZ
Coordinador Comité Curricular

Betty M.

ACTA DE SUSTENTACION DE TRABAJO DE GRADO

FECHA: 20 DE DICIEMBRE DE 2019 HORA: 9:00 a. m.

LUGAR: FU 309 - UFPS

PLAN DE ESTUDIOS: INGENIERIA CIVIL

TITULO DE LA TESIS: "RECONSTRUCCION DE SERIES DE TIEMPO DE TEMPERATURA DIARIA MEDIANTE REGRESION LINEAL MULTIVARIADA EN NORTE DE SANTANDER".

JURADOS: ING. NELSON JAVIER CELY CALIXTO
ING. JUAN CARLOS SAYAGO ORTEGA

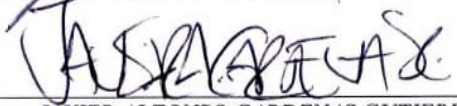
DIRECTOR: Ph.D. GUSTAVO ADOLFO CARRILLO SOTO.

NOMBRE DE LOS ESTUDIANTES:	CODIGO	CALIFICACION	
		NUMERO	LETRA
LEIDY YERALDIN RAMIREZ FLOREZ	2110094	4,5	CUATRO, CINCO

MERITORIA


ING. NELSON JAVIER CELY CALIXTO


ING. JUAN CARLOS SAYAGO ORTEGA

Vo. Bo. 
JAVIER ALFONSO CARDENAS GUTIERREZ
Coordinador Comité Curricular

Betty M.



**CARTA DE AUTORIZACIÓN DE LOS AUTORES PARA
LA CONSULTA, LA REPRODUCCIÓN PARCIAL O TOTAL Y LA PUBLICACIÓN
ELECTRÓNICA DEL TEXTO COMPLETO**

Cúcuta,

Señores
BIBLIOTECA EDUARDO COTE LAMUS
Ciudad

Cordial saludo:

LEIDY YERALDIN RAMIREZ FLOREZ y CARLOS FEDERICO CALDERÓN DELGADO, identificado(s) con la C.C. N.º 1.094.532.868 y C.C. N.º 1.090.175.751, autor(es) de la tesis y/o trabajo de grado titulado RECONSTRUCCIÓN DE SERIES DE TIEMPO DE TEMPERATURA DIARIA MEDIANTE REGRESIÓN LINEAL MULTIVARIADA EN NORTE DE SANTANDER presentado y aprobado en el año 2020 como requisito para optar al título de INGENIERO CIVIL; autorizo(amos) a la biblioteca de la Universidad Francisco de Paula Santander, Eduardo Cote Lamus, para que con fines académicos, muestre a la comunidad en general a la producción intelectual de esta institución educativa, a través de la visibilidad de su contenido de la siguiente manera:

- Los usuarios pueden consultar el contenido de este trabajo de grado en la página web de la Biblioteca Eduardo Cote Lamus y en las redes de información del país y el exterior, con las cuales tenga convenio la Universidad Francisco de Paula Santander.
- Permita la consulta, la reproducción, a los usuarios interesados en el contenido de este trabajo, para todos los usos que tengan finalidad académica, ya sea en formato CD-ROM o digital desde Internet, Intranet etc.; y en general para cualquier formato conocido o por conocer.

Lo anterior, de conformidad con lo establecido en el artículo 30 de la ley 1982 y el artículo 11 de la decisión andina 351 de 1993, que establece que “**los derechos morales del trabajo son propiedad de los autores**”, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.

FEDERICO CALDERÓN

Carlos Federico Calderón Delgado.
C.C. 1.090.175.751.
Expedida en Chinácota,
Norte de Santander.

Yeraldin Ramirez

Leidy Yeraldin Ramirez Florez.
C.C. 1.094.532.868.
Expedida en Lourdes,
Norte de Santander.

Contenido

	pág.
Introducción	15
1. Problema	16
1.1 Título	16
1.2 Planteamiento del Problema	16
1.3 Formulación del Problema	17
1.4 Objetivos	17
1.4.1 Objetivo general	17
1.4.2 Objetivos específicos	17
1.5 Justificación	17
1.6 Alcances y Limitaciones	18
1.6.1 Alcances	18
1.6.2 Limitaciones	18
1.7 Delimitaciones	18
1.7.1 Delimitación espacial	18
1.7.2 Delimitación temporal	19
1.7.3 Delimitación conceptual	19
2. Marco Referencial	21
2.1 Antecedentes	21
2.1.1 Antecedentes empíricos	21
2.1.2 Antecedentes bibliográficos	21

2.2 Marco Teórico	23
2.3 Marco Conceptual	23
2.4 Marco Contextual	37
2.5 Marco Legal	39
3. Diseño Metodológico	41
3.1 Tipo de Investigación	41
3.2 Población y Muestra	41
3.2.1 Población	41
3.2.2 Muestra	41
3.3 Instrumentos para la Recolección de Información	41
3.4 Técnica de Análisis y Procesamiento de Datos	42
3.5 Presentación de Resultados	42
4. Contenido del proyecto	43
4.1 Selección de Estaciones Según Catálogo Nacional de Estaciones del IDEAM y Otras Entidades	43
4.2 Recolección de la Data	45
4.3 Análisis de Series Temporales de Temperatura	49
4.4 Selección de Estaciones para Llenado de Datos	78
4.5 Aplicación de Procedimientos Estándar para Determinar Datos Faltantes en las Series Seleccionadas	79
4.6 Presentación de la Información Recolectada Usando una Herramienta Tecnológica que Permita Reconstruir las Series de Temperatura	91

5. Conclusiones	94
6. Recomendaciones	97
Referencias Bibliográficas	99

Lista de Figuras

	pág.
Figura 1. Asimetría	33
Figura 2. San José de Cúcuta	39
Figura 3. Sitio web Consulta - Descarga Datos Hidrometeorológicos	45
Figura 4 Organización de información reportada en formatos de Excel	58
Figura 5. Gráfica de dispersión - Valores Máximos AEROPUERTO CAMILO DAZA año 1976	58
Figura 5. Box plot - Valores Máximos	72
Figura 6. Matriz de coeficiente de correlación lineal - Valores Máximos	79
Figura 7. Matriz de coeficiente de correlación lineal - Valores Medios	80
Figura 8. Matriz de coeficiente de correlación lineal - Valores mínimos	80
Figura 9. Generación de EMC de estación 3 simulada	83
Figura 10. Dispersión EST. 3 vs. EST. 3 SIM	84
Figura 11. Diferencias de Error	84
Figura 12. Rangos de Error	85
Figura 13. Localización de Estaciones mediante Google Earth	92

Lista de Tablas

	pág.
Tabla 1. Estaciones meteorológicas seleccionadas	44
Tabla 2. Variables con parámetro temperatura	46
Tabla 3. Data solicitada en radicado 1	47
Tabla 4. Data solicitada en radicado 2	47
Tabla 5. Data solicitada en radicado 3	48
Tabla 6. Data solicitada en radicado 4	48
Tabla 7. Data solicitada en radicado 5	49
Tabla 8. Recibido 1	50
Tabla 9. Recibido 2	51
Tabla 10. Recibido 2 solicitado	53
Tabla 11. Recibido 2 no solicitado	53
Tabla 12. Recibido 2 estaciones duplicadas - no solicitadas	54
Tabla 13. Estaciones con Datos Dudosos	59
Tabla 14. Análisis de series de temperatura diaria - Valores Máximos Aeropuerto Camilo Daza	62
Tabla 15. Comparación estaciones repetidas en recibidos 1 y 2	65
Tabla 16. Denominación de estaciones	66
Tabla 17. Porcentaje de Datos Registrados (PDR) – Valores Máximos	68
Tabla 18. Tmax16015010 - Temperatura Máxima 1976-2018 Apto Camilo Daza	69
Tabla 19. Análisis general de series de temperatura diaria – Valores Máximos	70
Tabla 20. Estaciones seleccionadas para llenado de datos	78

Tabla 21. Gráfica Error Medio Cuadrático (EMC)	86
Tabla 22. Regresión 1 Variable EST. 2 - Valores Máximos	87
Tabla 23. Regresión 2 Variables EST. 2 - Valores Máximos	88
Tabla 24. Regresión 3 Variables EST. 2 - Valores Máximos	88
Tabla 25. Regresión 4 Variables EST. 2 - Valores Máximos	89
Tabla 26. EST. 2 Real vs EST. 2 Simulada Valores Máximos	90
Tabla 27. Comparación EST.2 Real vs EST.2 Simulada Valores Máximos	91

Resumen

El trabajo trata sobre la Reconstrucción de series de tiempo de temperatura diaria mediante regresión lineal multivariada en Norte de Santander. Debido a, que La Universidad Francisco de Paula Santander es una Institución Pública de Educación Superior, orientada al mejoramiento continuo y la calidad en los procesos de docencia, investigación y extensión, en el marco de estrategias metodológicas presenciales, a distancia y virtuales, cuyo propósito fundamental es la formación integral de profesionales comprometidos con la solución de problemas del entorno, en busca del desarrollo sostenible de la región. Se propone, Reconstruir las series de tiempo de temperatura diaria mediante regresión lineal multivariada en el departamento de Norte de Santander. Para ello, se analiza las series temporales de temperatura diaria para estaciones en Norte de Santander. Seguido de, aplicar procedimientos estándar para determinar datos faltantes en las series seleccionadas; y finalmente, presentar la información recolectada usando una herramienta tecnológica que permita reconstruir las series de temperatura, apoyando la iniciativa del Observatorio Hidro-Climatológico (HCO-UFPS) liderada por el grupo de investigación HYDROS-UFPS. El tipo de investigación de este proyecto según el nivel corresponde a descriptivo. La población, son las 8858 estaciones meteorológicas de las cuales 294 se encuentran en el Departamento Norte de Santander. Y para la muestra, se selecciona las estaciones meteorológicas del IDEAM con información útil que permita la transformación de la data. Para el procesamiento de la información se utiliza la página principal del IDEAM para realizar la consulta y descarga de la data, obtenidos por medio de las estaciones Hidrometeorológicas del Departamento Norte de Santander, que permitan suministrar datos para la investigación descriptiva - documental del presente proyecto.

Abstract

The work deals with the Reconstruction of daily temperature time series using multivariate linear regression in Norte de Santander. Due to, the Francisco de Paula Santander University is a Public Institution of Higher Education, aimed at continuous improvement and quality in the teaching, research and extension processes, within the framework of face-to-face and virtual methodological strategies, whose fundamental purpose It is the comprehensive training of professionals committed to solving environmental problems, in search of sustainable development in the region. It is proposed, to reconstruct the daily temperature time series by means of multivariate linear regression in the department of Norte de Santander. To do this, it analyzes the daily temperature time series for stations in Norte de Santander. Following, apply standard procedures to determine missing data in the selected series; and finally, to present the information collected using a technological tool that allows the reconstruction of the temperature series, supporting the initiative of the Hydro-Climatological Observatory (HCO-UFPS) led by the HYDROS-UFPS research group. The type of investigation of this project according to the level corresponds to descriptive. The population is the 8858 meteorological stations of which 294 are in the North Department of Santander. And for the sample, select the IDEAM weather stations with useful information that allows data transformation. For the processing of the information, the main page of the IDEAM is used to consult and download the data, to detect through the Hydrometeorological stations of the North Department of Santander, that data can be sent for the descriptive - documentary investigation of the present draft.

Introducción

La Universidad Francisco de Paula Santander es una Institución Pública de Educación Superior, orientada al mejoramiento continuo y la calidad en los procesos de docencia, investigación y extensión, en el marco de estrategias metodológicas presenciales, a distancia y virtuales, cuyo propósito fundamental es la formación integral de profesionales comprometidos con la solución de problemas del entorno, en busca del desarrollo sostenible de la región (Universidad Francisco de Paula Santander, 2019).

El Observatorio Hidro-Climatológico (HCO-UFPS), liderado por el grupo de investigación HYDROS-UFPS, se plantea como un centro regional de modelación hidroclimatológica que busca conocer el comportamiento de sus variables principales.

El Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM tiene como entidad la misión de ... “apoyo técnico y científico al Sistema Nacional Ambiental, que genera conocimiento, produce información confiable, consistente y oportuna, sobre el estado y las dinámicas de los recursos naturales y del medio ambiente” ... dado que para la recopilación de la información a través de esta entidad se obtiene la misma en formatos que difieren en la utilización para procesos académicos, como es la reconstrucción de series de tiempo de temperatura diaria en el departamento de Norte de Santander (Instituto de Hidrología, Meteorología y Estudios Ambientales- IDEAM, 2019).

1. Problema

1.1 Título

Reconstrucción de series de tiempo de temperatura diaria mediante regresión lineal multivariada en Norte de Santander

1.2 Planteamiento del Problema

La Universidad Francisco de Paula Santander es una Institución Pública de Educación Superior, orientada al mejoramiento continuo y la calidad en los procesos de docencia, investigación y extensión, en el marco de estrategias metodológicas presenciales, a distancia y virtuales, cuyo propósito fundamental es la formación integral de profesionales comprometidos con la solución de problemas del entorno, en busca del desarrollo sostenible de la región (Universidad Francisco de Paula Santander, 2019).

El Observatorio Hidro-Climatológico (HCO-UFPS), liderado por el grupo de investigación HYDROS-UFPS, se plantea como un centro regional de modelación hidroclimatológica que busca conocer el comportamiento de sus variables principales.

El Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM tiene como entidad la misión de ... “apoyo técnico y científico al Sistema Nacional Ambiental, que genera conocimiento, produce información confiable, consistente y oportuna, sobre el estado y las dinámicas de los recursos naturales y del medio ambiente” ... dado que para la recopilación de la información a través de esta entidad se obtiene la misma en formatos que difieren en la utilización para procesos académicos, como es la reconstrucción de series de tiempo de

temperatura diaria en el departamento de Norte de Santander (Instituto de Hidrología, Meteorología y Estudios Ambientales- IDEAM, 2019).

1.3 Formulación del Problema

¿Existen series de tiempo de temperatura diaria, con registros validados, homogenizados y sin datos faltantes, en formatos de fácil acceso y manipulación, en el departamento de Norte de Santander?

1.4 Objetivos

1.4.1 Objetivo general. Reconstruir las series de tiempo de temperatura diaria mediante regresión lineal multivariada en el departamento de Norte de Santander.

1.4.2 Objetivos específicos. Analizar las series temporales de temperatura diaria para estaciones en Norte de Santander

Aplicar procedimientos estándar para determinar datos faltantes en las series seleccionadas.

Presentar la información recolectada usando una herramienta tecnológica que permita reconstruir las series de temperatura, apoyando la iniciativa del Observatorio Hidro-Climatológico (HCO-UFPS) liderada por el grupo de investigación HYDROS-UFPS.

1.5 Justificación

A raíz de la iniciativa de un Observatorio Hidro-Climatológico regional (HCO-UFPS), se pretende recolectar la data de las tres variables principales para la modelación hidroclimatologica, las cuales son: Caudal, precipitación y temperatura.

Partiendo de la variable temperatura, la cual influye en el cambio climático; se lleva a cabo este proyecto de la reconstrucción de series de tiempo de temperatura diaria mediante regresión lineal multivariada en el departamento de Norte de Santander.

1.6 Alcances y Limitaciones

1.6.1 Alcances. El propósito de este proyecto es la reconstrucción de series de tiempo de temperatura diaria mediante regresión lineal multivariada en el departamento de Norte de Santander utilizando la información obtenida en las diferentes estaciones meteorológicas del IDEAM que proporcionen datos útiles.

1.6.2 Limitaciones. En el proceso de reconstrucción de series de temperatura diaria se pueden presentar una serie de limitaciones referentes a:

La demora en la obtención de la data.

La selección de las estaciones objeto de estudio.

La transformación de la data a información.

El proceso de llenado de datos faltantes.

1.7 Delimitaciones

1.7.1 Delimitación espacial. El proyecto se lleva a cabo en el departamento de Norte de Santander. Norte de Santander es uno de los 32 departamentos de Colombia. Está ubicado en la zona nororiental del país, sobre la frontera con Venezuela. Se localiza geográficamente entre los $06^{\circ}56'42''$ y $09^{\circ}18'01''$ de latitud norte y los $72^{\circ}01'13''$ y $73^{\circ}38'25''$ de longitud oeste.

Tiene una extensión de 22.130 km², que equivalen al 1.91% del millón ciento cincuenta y nueve mil ochocientos setenta y un kilómetros cuadrados (1.159.871,41 km²) del territorio nacional. Limita al norte y al este con Venezuela, al sur con los departamentos de Boyacá y Santander, y al oeste con Santander y Cesar. Forma parte de la Región Andina junto con los departamentos de Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Santander, Quindío, Risaralda, y Tolima, la más densamente poblada del país, donde reside más del 70% de la población colombiana.

Importante a destacar es la trascendencia histórica que tiene para el continente americano; en este departamento fue creada la Gran Colombia.

Ha dado al país varias personas importantes como Francisco de Paula Santander (primer presidente de la República y fundador del Ejército Nacional), Camilo Daza (fundador de la Fuerza Aérea Colombiana), Rafael García Herreros (fundador del Minuto de Dios), Fabiola Zuluaga (la mejor tenista del país), Virgilio Barco (ex-presidente de la República) y José Eusebio Caro, importante literato y fundador del Partido Conservador (Gobernación de Norte de Santander, 2019).

1.7.2 Delimitación temporal. El proyecto se desarrollará en el transcurso del primer semestre académico del año 2019 programado por la Universidad Francisco de Paula Santander.

1.7.3 Delimitación conceptual. Se manejarán conceptos claves como son:

Modelación hidroclimatológica

Temperatura

Precipitación

Caudal

Hidrología

Climatología

Parámetros estadísticos

Llenado de datos faltantes

Regresión lineal multivariada

2. Marco Referencial

2.1 Antecedentes

2.1.1 Antecedentes empíricos. Medina (2008) estimación estadística de valores faltantes en series históricas de lluvia. Trabajo de investigación para optar al grado de magister en investigación operativa y estadística, Universidad Tecnológica de Pereira, Facultad de Ingeniería Industrial, Escuela de Postgrados, Pereira, abril, 2008. Las observaciones de lluvia que se registran en las distintas estaciones de la red meteorológica de la Federación Nacional de Cafeteros de Colombia, es uno de los insumos para la realización de estudios, investigaciones, actividades de seguimiento y manejo de la información que sirven para fundamentar la toma de decisiones en materia de recursos, costos de producción y política ambiental. En ocasiones, no se cuenta con información completa, debido a situaciones como equipos no calibrados, deterioro en unidades de medición, mediciones no oportunas o deficiencias en el procedimiento mismo de la observación y la medición entre otros. Con el fin de tener series históricas completas de lluvia, se propuso un método de estimación de datos faltantes de fácil aplicación, basado en series de datos de precipitación acumulada por ciclos en periodos de tiempo mensuales, teniendo en cuenta la agrupación por años El Niño, La Niña y Neutro, fundamentado y validado estadísticamente.

2.1.2 Antecedentes bibliográficos. Castro (2009) Metodología de validación de productos MODIS para la estimación de temperatura de la superficie en zonas heterogéneas y homogéneas de Colombia. Trabajo de grado para optar por el título de Magíster en Geomática, Universidad Nacional de Colombia, Facultad de Agronomía, Escuela de Posgrados, Bogotá D.C., 2009. La acelerada producción científica en geomática, ha permitido conocer fenómenos sobre la faz de la Tierra. Sus herramientas como el modelamiento espacial, los geodatos de sensores remotos, la

administración de la información geoespacial y los estudios sobre la dinámica del planeta, han suministrado claves de acceso al mejoramiento del entendimiento de los fenómenos naturales y antropogénicos.

El avance tecnológico provee el uso de mayor información para estudiar variables ambientales como la temperatura de la superficie (del aire a 2 metros). Sin embargo, obtener estos datos de manera tradicional involucra un costo económico y un cubrimiento espacial insuficiente, ya que esta información es requerida para adelantar estudios agronómicos expansivos, modelos matemáticos de interacción Tierra-Atmósfera y cambio climático.

Frente a esta necesidad, el uso de información proveniente de datos satelitales crece a medida que los objetivos y metas en la investigación también aumentan. En Colombia, la baja cobertura espacial de las estaciones meteorológicas deja vacíos de información que pueden ser estimados a través de las herramientas geomáticas.

En esta investigación, la explotación de los datos satelitales del sensor Modis en conjunto con los termodatos de las estaciones en terreno, permitieron establecer los parámetros que explican espacialmente el fenómeno de temperatura en el país y cómo ésta se comporta de acuerdo a las diferencias geográficas propias del territorio.

A través del modelamiento geoestadístico, el conocimiento empírico y los ajustes teóricos, se estableció un Modelo de Regresión Lineal Múltiple, que estima las temperaturas de la superficie con alta fiabilidad.

2.2 Marco Teórico

IDEAM – Instituto de Hidrología, Meteorología y Estudios Ambientales

Misión: El IDEAM es una institución pública de apoyo técnico y científico al Sistema Nacional Ambiental, que genera conocimiento, produce información confiable, consistente y oportuna, sobre el estado y las dinámicas de los recursos naturales y del medio ambiente, que facilite la definición y ajustes de las políticas ambientales y la toma de decisiones por parte de los sectores público, privado y la ciudadanía en general.”

Visión: En el año 2026 el IDEAM será el Instituto modelo por excelencia, reconocido nacional e internacionalmente como la Entidad que genera y suministra información hidrológica, meteorológica y ambiental para la definición de políticas públicas y toma de decisiones relacionadas con el desarrollo sostenible y la prevención de los efectos de cambio climático (IDEAM, 2019).

2.3 Marco Conceptual

Modelo hidrológico: Un modelo hidrológico es pues una representación simplificada de un sistema real complejo llamado prototipo, bajo forma física o matemática. De manera matemática, el sistema real está representado por una expresión analítica.

En un modelo hidrológico, el sistema físico real que generalmente representamos es la 'cuenca hidrográfica' y cada uno de los componentes del ciclo hidrológico. De esta manera un modelo matemático nos ayudará a tomar decisiones en materia de hidrología, por lo que es necesario tener conocimiento de entradas (inputs) al sistema y salidas (outputs) a partir del sistema, para verificar si el modelo es representativo del prototipo.

La salida de los modelos hidrológicos varía - dependiendo de las metas y objetivos del modelo. Algunos modelos se utilizan para predecir los totales mensuales de escorrentía, mientras que otros están diseñados para ver a las tormentas individuales. El resultado más común es el hidrograma o hidrograma de escurrimiento (IDEAM, 2019).

Temperatura: Es una magnitud física que refleja la cantidad de calor, ya sea de un cuerpo, de un objeto o del ambiente. Dicha magnitud está vinculada a la noción de frío (menor temperatura) y caliente (mayor temperatura).

La temperatura está relacionada con la energía interior de los sistemas termodinámicos, de acuerdo al movimiento de sus partículas, y cuantifica la actividad de las moléculas de la materia: A mayor energía sensible, más temperatura.

El estado, la solubilidad de la materia y el volumen, entre otras cuestiones, dependen de la temperatura. En el caso del agua a presión atmosférica normal, si se encuentra a una temperatura inferior a los 0°C, se mostrará en estado sólido (congelada); si aparece a una temperatura de entre 1°C y 99°C, se encontrará en estado líquido; si la temperatura es de 100°C o superior, por último, el agua presentará un estado gaseoso (vapor) (Definición, 2019).

Precipitación: La definición más sencilla de precipitación indica que es cualquier forma de agua que cae del cielo. Por supuesto, la precipitación en cualquiera de sus tipos tiene su formación en la atmósfera de la Tierra y cae después sobre la superficie de ésta, mayormente como lluvia.

La precipitación forma parte del ciclo del agua y gracias a este componente, los seres vivos obtienen el agua dulce que necesitan para vivir. Después de caer, se evapora y se eleva a la

atmósfera en forma de gas, se condensa y desciende de nuevo. El proceso se repite una y otra vez.

Cada año caen unos 505,000 km³ de agua; 398,000 km³ sobre los océanos y 107,000 km³ sobre la superficie terrestre. Esto significa que en promedio, la precipitación anual es de 99 centímetros.

Pero la precipitación es diferente en cada región geográfica del mundo. Por ejemplo, en Georgia, Estados Unidos, las lluvias caen de manera uniforme durante todo el año, es decir, con un nivel de 102-127 centímetros anuales mientras que en Arica, Chile, no llovió durante 14 años consecutivos. Entonces, las precipitaciones pueden variar en una región, área e incluso ciudad (Geoenciclopedia, 2019).

Caudal: Como definición general, se conoce como caudal, a la cantidad de fluido que circula a través de una sección de un ducto, ya sea tubería, cañería, oleoducto, río, canal, por unidad de tiempo. Generalmente, el caudal se identifica con el flujo volumétrico o volumen que pasa por un área determinada en una unidad de tiempo específica.

El caudal de un río puede aumentar o disminuir dependiendo de la estación del año, por ejemplo, los ríos que se alimentan principalmente del deshielo de las capas de nieve, aumentan su caudal en primavera, mientras que ríos cuya agua procede de las precipitaciones de aguas meteóricas, presentan niveles máximos de caudal en épocas de lluvias y niveles mínimos en las estaciones o meses más secos. Es por esto que, si la fuente hídrica se localiza en zonas con altos índices de precipitación a lo largo del año, el caudal será constante y regular, y si se localiza en zonas donde la precipitación sea irregular, sufrirá fuertes crecidas en las épocas de lluvia y bajará

su nivel de agua el resto del año; este fenómeno se conoce como estiaje (Blog Fibras y Normas de Colombia S.A.S, 2019).

Hidrología: Es la ciencia natural que estudia el agua, ocurrencia, circulación y distribución en la superficie terrestre, sus propiedades químicas y físicas y su relación con el medio ambiente, incluyendo a los seres vivos (Aparicio, 1992).

Climatología: Hace parte de las ciencias de la tierra. La climatología estudia los fenómenos del clima, estableciendo su comportamiento por zonas, de acuerdo a las variaciones que se han presentado en el tiempo, es decir, basado en lo ocurrido en el pasado, se conoce lo que ocurre en el presente y se prevé lo que sucederá en el futuro, respecto al clima.

La climatología tiende a ser confundida con la meteorología (ciencia que estudia el tiempo atmosférico), porque ambas utilizan los mismos parámetros de evaluación, pero la diferencia está en que la meteorología realiza previsiones a corto plazo o de manera inmediata, mientras que la finalidad de la climatología está en el estudio y previsiones del comportamiento del clima a futuro o largo plazo.

La climatología está relacionada prácticamente con todas las ciencias que hacen parte de la cotidianidad del ser humano de hoy, aquellas como la Ingeniería o la arquitectura, que implican la interacción del hombre con la naturaleza y las alteraciones que produce (Conceptodefinición, 2019).

Las observaciones de temperatura, presión atmosférica, vientos, humedad y precipitaciones, así como el tipo o tipos de tiempo que se recogen en las estaciones meteorológicas. Con estos

datos se elaboran tablas de valores medios que se trasladan a climogramas, representaciones gráficas de la variación anual de temperatura y precipitaciones, como variables principales.

Estudio: Para el estudio de la climatología tenemos que considerar tres métodos que son fundamentales:

La climatología analítica; que está basado en el análisis estadístico de las características que se consideran más significativas. En ella se establecen los valores medios de los elementos atmosféricos y establecer la probabilidad de que se alcancen determinados valores extremos.

La climatología dinámica; que proporciona una visión dinámica y de conjunto de las manifestaciones cambiantes que se registran en la atmósfera como una unidad física. Es una explicación matemática de la atmósfera mediante las leyes de la mecánica de fluidos y de la termodinámica.

La climatología sinóptica; consiste en el análisis de la configuración de los elementos atmosféricos en un espacio tridimensional y a unas horas concretas y de su evolución. Se pretenden descubrir leyes empíricas e incrementar el conocimiento acerca de la atmósfera.

Factores climáticos: La atmósfera está en continuo movimiento, lo que se denomina circulación atmosférica. Concretamente los factores climáticos obedecen a las características geográficas influyentes en el clima y hay que distinguir:

Energía solar: Nivel de radiación de los rayos solares.

Latitud: Distancia al Norte o Sur del Ecuador terrestre.

Altitud y relieve: Altura a partir del nivel del mar.

Orientación: Forma y posición de la Tierra en el sistema solar.

Continentalidad: Ubicación geográfica de las tierras continentales.

Otros factores relacionados con la distribución geográfica de tierras, mares, llanuras, bosques, montañas y desiertos (Tutiempo.net, 2019).

Clima: El clima es el conjunto fluctuante de las condiciones atmosféricas, caracterizado por los estados y evoluciones del estado del tiempo, durante un periodo de tiempo y un lugar o región dados, y controlado por los denominados factores forzantes, factores determinantes y por la interacción entre los diferentes componentes del denominado sistema climático (atmósfera, hidrosfera, litosfera, criósfera, biosfera y antropósfera).

Debido a que el clima se relaciona generalmente con las condiciones predominantes en la atmósfera, este se describe a partir de variables atmosféricas como la temperatura y la precipitación, denominados elementos climáticos; sin embargo, se podría identificar también con las variables de otros de los componentes del sistema climático (IDEAM, 2019).

Tiempo: El tiempo es la manifestación de la dinámica de la atmósfera en un lugar y momento determinados. La dinámica de la atmósfera al distribuir la masa (vapor de agua y otros gases) y la energía (calor y movimiento) genera variaciones espaciotemporales de elementos como la temperatura, la presión y la humedad, lo cual produce en un lugar y tiempo determinados condiciones cálidas o frías, húmedas o secas, de cielo nublado o de cielo despejado, situaciones de lluvia, etc. Estos fenómenos se conocen como estado del tiempo (IDEAM, 2019).

Parámetro: Es una cantidad numérica calculada sobre una población y resume los valores que esta toma en algún atributo (Guardiola, 2019).

Estadístico: Es una cantidad numérica calculada sobre una muestra que resume su información sobre algún aspecto (Guardiola, 2019).

Parámetros estadísticos: El objetivo de la estadística es extraer la información esencial de un conjunto de datos, reduciendo un conjunto grande de números a un conjunto pequeño de números. Las estadísticas son números calculados de una muestra los cuales resumen sus características más importantes. Los parámetros estadísticos son características de una población, tales como media aritmética y desviación estándar como lo explica (Chow, & Mays, 2010.).

Hay tres tipos de parámetros estadísticos, de tendencia central o punto medio, de variabilidad o dispersión y de simetría. Las medidas de tendencia central son medidas estadísticas que pretenden resumir en un solo valor a un conjunto de valores. Representan un centro en torno al cual se encuentra ubicado el conjunto de los datos. Las medidas de tendencia central más utilizadas son: Media, mediana y moda. Las medidas de dispersión en cambio miden el grado de dispersión de los valores de la variable. Dicho en otros términos las medidas de dispersión pretenden evaluar en qué medida los datos difieren entre sí. De esta forma, ambos tipos de medidas usadas en conjunto permiten describir un conjunto de datos entregando información acerca de su posición y su dispersión (Quevedo, 2011).

Las medidas de dispersión más utilizadas son el rango, la desviación estándar y la varianza (Ecured, 2011).

La asimetría y curtosis es la medida que indica la simetría de la distribución de una variable respecto a la media aritmética, sin necesidad de hacer la representación gráfica. Los coeficientes de asimetría indican si hay el mismo número de elementos a izquierda y derecha de la media.

Existen tres tipos de curva de distribución según su asimetría:

Asimetría negativa: La cola de la distribución se alarga para valores inferiores a la media.

Simétrica: Hay el mismo número de elementos a izquierda y derecha de la media. En este caso, coinciden la media, la mediana y la moda. La distribución se adapta a la forma de la campana de Gauss, o distribución normal.

Asimetría positiva: La cola de la distribución se alarga (a la derecha) para valores superiores a la media. (Universo Formulas, s.f.).

Media aritmética: La media aritmética es el valor promedio de las muestras y es independiente de las amplitudes de los intervalos. Se simboliza como \bar{x} y se encuentra sólo para variables cuantitativas. Se encuentra sumando todos los valores y dividiendo por el número total de datos.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Mediana: La mediana de un conjunto de datos es el valor que cumple que la mitad de valores están por encima y la otra mitad por debajo. Así pues, para encontrarla basta con ordenar los elementos de menor a mayor y escoger el valor central.

Media geométrica: La media geométrica de un conjunto de n datos es el resultado de multiplicarlos entre si y aplicar la n-enésima raíz. Si en la media aritmética sumáramos los valores para luego dividirlos, ahora debemos multiplicarlos para luego aplicar la n -ésima raíz pertinente. Es decir,

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n}$$

Varianza: La varianza de unos datos es la media aritmética del cuadrado de las desviaciones respecto a la media de la misma. Se simboliza como σ^2 y se calcula aplicando:

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

Desviación típica o estándar: La desviación típica es la raíz cuadrada de la varianza y se representa por la letra σ . Para calcularla se calcula la varianza y se saca la raíz. Las interpretaciones que se deducen de la desviación típica son, por lo tanto, parecidas a las que se deducían de la varianza.

Comparando con el mismo tipo de datos, una desviación típica elevada significa que los datos están dispersos, mientras que un valor bajo indica que los valores son próximos los unos de los otros, y por lo tanto de la media.

Propiedades de la desviación típica

La desviación típica es un valor positivo, la igualdad sólo se da en el caso de que todas las muestras sean iguales.

Si a todos los datos se les suma una constante, la desviación típica sigue siendo la misma.

Si todos los datos se multiplican por una constante, la desviación típica queda multiplicada por dicha constante.

Si se dispone de varias distribuciones con la misma media y se calculan las distintas desviaciones típicas, se puede hallar la desviación típica total aplicando:

$$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}{n}}$$

En el caso de que las distribuciones tengan distinto tamaño, se pondera y queda como:

$$\sigma = \sqrt{\frac{\sigma_1^2 k_1 + \sigma_2^2 k_1 + \dots + \sigma_n^2 k_n}{k_1 + k_2 + \dots + k_n}}$$

Los conceptos de media aritmética, mediana, media geométrica, varianza y desviación típica o estándar se recopilan a través de (Sangaku S.L, 2019).

Coefficiente de variación: El coeficiente de variación, también denominado como coeficiente de variación de Spearman, es una medida estadística que nos informa acerca de la dispersión relativa de un conjunto de datos. Su cálculo se obtiene de dividir la desviación típica o estándar entre el valor absoluto de la media del conjunto y por lo general se expresa en porcentaje para su mejor comprensión. El coeficiente de variación se puede ver expresado con las letras CV o r, dependiendo del manual o la fuente utilizada (Economipedia, 2019).

$$CV = \frac{\sigma}{|\bar{x}|}$$

Coefficiente de asimetría: Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se dice que la asimetría es positiva cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es Simétrica cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se

conoce como asimetría negativa cuando la mayor cantidad de datos se aglomeran en los valores menores que la media. (SPSS FREE, version 24).

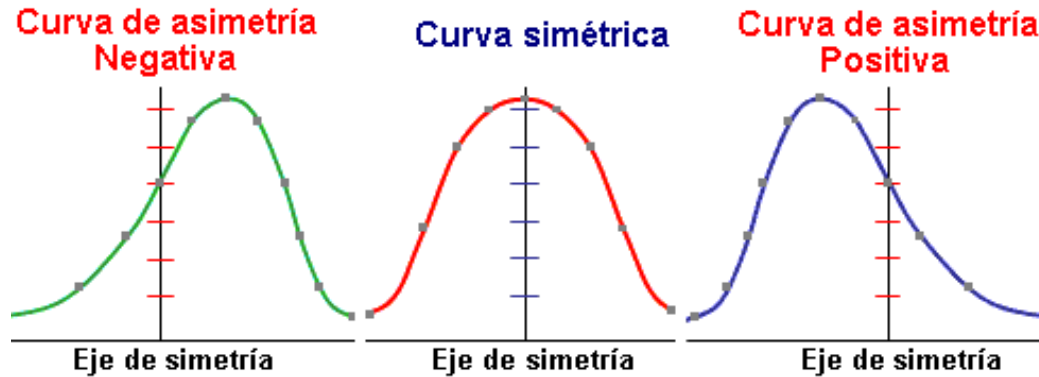


Figura 1. Asimetría

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

$$\mu_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

σ es la desviación estándar.

Los resultados pueden ser los siguientes:

$\gamma_1 = 0$ La distribución es simétrica: Existe la misma concentración de valores a la derecha y a la izquierda de la media.

$\gamma_1 > 0$ La distribución es asimétrica positiva: Existe mayor concentración de valores a la derecha de la media que a su izquierda. La cola derecha es más larga.

$\gamma_1 < 0$ La distribución es asimétrica negativa: Existe mayor concentración de valores a la izquierda de la media que a su derecha. La cola izquierda es más larga (Grupo Camaleon, 2017).

Coeficiente de correlación lineal: La correlación, también conocida como coeficiente de correlación lineal (de Pearson), es una medida de regresión que pretende cuantificar el grado de variación conjunta entre dos variables.

Por tanto, es una medida estadística que cuantifica la dependencia lineal entre dos variables, es decir, si se representan en un diagrama de dispersión los valores que toman dos variables, el coeficiente de correlación lineal señalará lo bien o lo mal que el conjunto de puntos representados se aproxima a una recta.

Valores que puede tomar la correlación:

$\rho = -1$ Correlación perfecta negativa

$\rho = 0$ No existe correlación

$\rho = +1$ Correlación perfecta positiva

Hablamos de correlación positiva si siempre que el valor «x» sube, el valor «y» sube, y además con la misma intensidad (+1).

En el caso opuesto, si siempre que el valor «x» sube, y el valor «y» baja, y además con la misma intensidad, entonces estamos hablando de correlación negativa (-1).

Es importante saber que esto no quiere decir que lo hagan en la misma proporción (salvo que tengan la misma desviación típica) (Economipedia, 2019).

Regresión lineal múltiple: La regresión lineal es una técnica estadística destinada a analizar las causas de por qué pasan las cosas. A partir de los análisis de regresión lineal múltiple podemos:

Identificar que variables independientes (causas) explican una variable dependiente (resultado)

Comparar y comprobar modelos causales

Predecir valores de una variable, es decir, a partir de unas características predecir de forma aproximada un comportamiento o estado

La regresión lineal múltiple es la gran técnica estadística para comprobar hipótesis y relaciones causales. Ante de empezar, una serie de condiciones que se deben cumplir para poder aplicar la regresión lineal múltiple:

La variable dependiente (resultado) debe ser ordinal o escalar, es decir, que las categorías de la variable tengan orden interno o jerarquía.

Las variables independientes (causas) deben ser ordinales o escalares.

Hay otras condiciones como: Las variables independientes no pueden estar altamente correlacionadas entre sí, las relaciones entre las causas y el resultado deben ser lineales, todas variables deben seguir la distribución normal y deben tener varianzas iguales. Estas condiciones no son tan estrictas y hay maneras de tratar los datos si se incumple.

Cómo analizar la regresión lineal múltiple en 4 pasos:

Los dos primeros pasos hacen referencia a la bondad del modelo, es decir, si el conjunto de variables independientes (causas) explican la variable dependiente (resultado)

Significación de F-test: Si es menor de 0,05 es que el modelo es estadísticamente significativo y por tanto las variables independientes explican “algo” la variable dependiente, cuánto “algo” es la R-cuadrado

R cuadrado: Es cuánto las variables independientes explican la variable dependiente, indica el porcentaje de la varianza de la variable dependiente explicado por el conjunto de variables independientes. Cuanto mayor sea la R-cuadrado más explicativo y mejor es el modelo causal.

Los dos siguientes pasos hacen referencia a la influencia de cada una de las variables independientes:

Significación de t-test: Si es menor de 0,05 es que esa variable independiente se relaciona de forma significativa con la variable dependiente, por tanto, influye sobre ella, es explicativa

Coefficiente beta (β): Indica la intensidad y la dirección de la relación entre esa variable independiente (VI) y la variable dependiente (VD):

Cuanto más se aleja de 0 más fuerte es la relación

El signo indica la dirección (signo + indica que al aumentar los valores de la VI aumentan los valores de la VD; signo – indica que al aumentar los valores de la VI, los valores de la VD descienden). (Networkianos, s.f.).

Error medio cuadrático (EMC): Definido matemáticamente como la raíz cuadrada del cociente entre la suma de los cuadrados de los errores aleatorios y el número de errores menos uno, se minimiza con una solución por el método de los cuadrados mínimos. Él da una medida

estadística de la dispersión de las posiciones calculadas en torno a la "posición mejor ajustada". A menor EMC mayor precisión (Grupo Camaleon, 2018).

2.4 Marco Contextual

San José de Cúcuta es la capital del Departamento Norte de Santander y la frontera más dinámica de Colombia, por este lugar pasa aproximadamente el 80 por ciento de las exportaciones dirigidas a Venezuela. Cúcuta crece al ritmo de los tiempos. Su ubicación fronteriza ha hecho que se extienda cada vez más para dar cabida a miles de habitantes provenientes de todos los rincones del país y del mundo.

Historia: Fue fundada en el año 1733 en terrenos donados por Doña Juana Rangel de Cuellar.

Posteriormente en el año 1875 sufre un devastador terremoto, que dio pie para ser reconstruida en el mismo lugar, siguiendo los planes del ingeniero venezolano Francisco de Paula Andrade Troconis.

En San José de Cúcuta se instaló el primer ferrocarril privado del país, la primera instalación de servicio telefónico en el año 1890 y la primera instalación de alumbrado público eléctrico en el año 1897. Su actividad económica principal es el comercio.

La Historia de Cúcuta se remonta a la llegada de los conquistadores españoles al continente americano, cuando era un asentamiento precolombino que Pedro de Ursúa le dio en encomienda a Sebastián Lorenzo en 1550. La conducta hostil de los indios Motilones para con los habitantes blancos del valle, fue factor determinante para que se construyera una parroquia que llevaría el nombre de “Patriarca San José”. Juana Rangel de Cuéllar donó 782 hectáreas en el sitio

Guaimaral y se inició la formación del poblado. La iglesia fue el centro alrededor de la cual se fundó el pueblo de Cúcuta, la cual tuvo un considerable progreso debido al comercio.

La ciudad fue cuna del núcleo humano más significativo de la historia de Colombia, donde nacieron hombres como Francisco de Paula Santander, Fernando Galvis Cáceres, Pedro Fortul, Mercedes Abrego, Francisco Montes de Oca y Juan Nepomuceno Piedor.

En 1875, Cúcuta ocupaba un área de 2,5 km² y abrigaba en sus 52 manzanas de casas a unos 12 mil habitantes. Poseía tres iglesias: La Catedral de San José, la del Carmen (ubicada al costado de la Biblioteca Pública Julio Pérez Ferrero) y la de San Antonio. Tenía también la Casa Municipal, dos teatros, un gran mercado cubierto, locales para escuelas, un magnífico hospital (San Juan de Dios) y se levantaba un puente de siete arcos sobre el río Pamplonita.

En la actualidad, el área de la ciudad es de 1.176 km² y se destaca por ser una de las ciudades principales de Colombia. Tiene una aglomeración urbana oficialmente constituida (denominada Área Metropolitana de Cúcuta), cuya población asciende a más de 1'200.000 personas. Tanto en su población neta como en la de su conurbación ocupa el puesto número cinco a nivel nacional. (AMC - Área Metropolitana de Cúcuta, 2019).

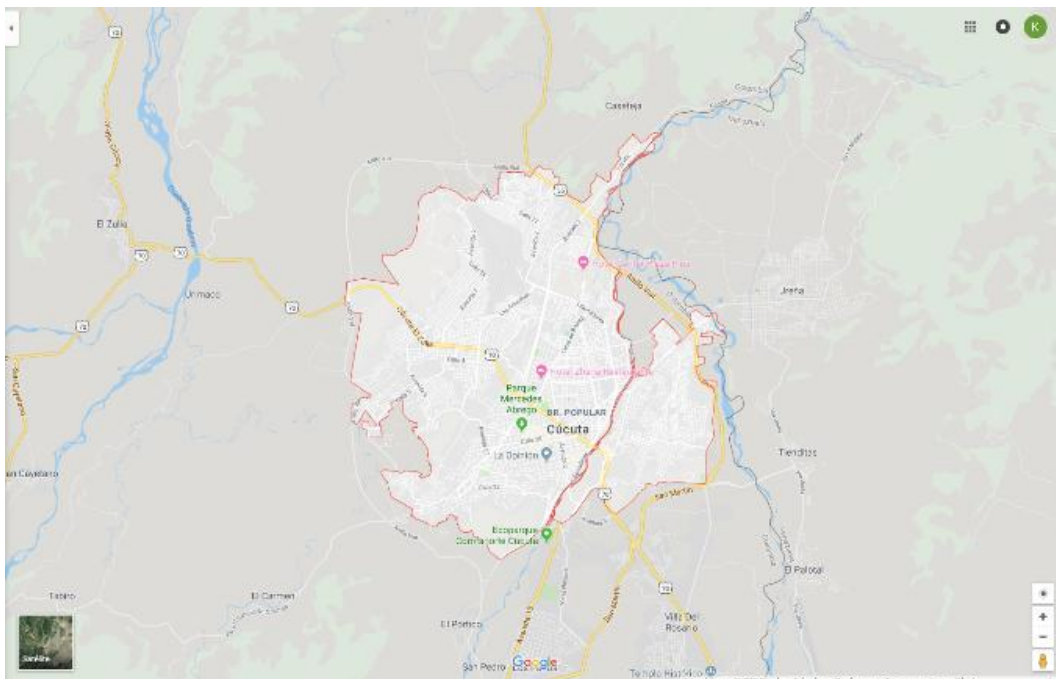


Figura 2. San José de Cúcuta

2.5 Marco Legal

El Consejo Superior Universitario de la Universidad Francisco de Paula Santander, estableció el Estatuto estudiantil el día 26 de agosto de 1996, mediante el acuerdo N° 065, donde el artículo 140, define las diferentes opciones que tiene el estudiante para realizar su trabajo de grado, que contempla posibles proyectos, como los trabajos de investigación y sistematización del conocimiento o, proyectos de extensión como las pasantías, trabajo de grado y reglamentado por el acuerdo 069 del 5 de septiembre de 1997, Inciso b de este acuerdo (Universidad Francisco de Paula Santander, 1996).

b. Investigación: Es la actividad intelectual encaminada a la construcción de conocimientos en las diversas esferas de la actividad humana, utilizando instrumentos racionales y materiales concebidos a través del tiempo, dentro del rigor y los cánones aceptados como científicos y cuyo

fin último es el progreso del conocimiento y su aplicación en beneficio de la sociedad (UFPS, 1996)

.

3. Diseño Metodológico

3.1 Tipo de Investigación

El tipo de investigación de este proyecto según el nivel corresponde a descriptivo, que; “consiste en la caracterización de un hecho, fenómeno, individuo o grupo, con el fin de establecer su estructura o comportamiento” (Arias, 2006, p. 24). Y respecto del diseño a la investigación documental, donde se realiza la interpretación de datos secundarios.

3.2 Población y Muestra

3.2.1 Población. Colombia cuenta en la actualidad con 8858 estaciones meteorológicas de las cuales 294 se encuentran en el Departamento Norte de Santander.

3.2.2 Muestra. Se selecciona las estaciones meteorológicas del IDEAM con información útil que permita la transformación de la data.

3.3 Instrumentos para la Recolección de Información

Los instrumentos que se utilizan para la recolección de la información en el transcurso de la ejecución del proyecto, son los siguientes:

Computador

Impresora

Recopilación y análisis de datos

Papelería

Formatos

Trabajos de grado

3.4 Técnica de Análisis y Procesamiento de Datos

Para el procesamiento de la información se utiliza la página principal del IDEAM para realizar la consulta y descarga de la data, obtenidos por medio de las estaciones Hidrometeorológicas del Departamento Norte de Santander, que permitan suministrar datos para la investigación descriptiva - documental del presente proyecto.

En conjunto el programa de Excel procede a la transformación y análisis de la información.

3.5 Presentación de Resultados

Los resultados se presentan por medio de un documento final, con formatos diligenciados en Excel que corresponden a las estaciones meteorológicas objeto de análisis.

4. Contenido del proyecto

4.1 Selección de Estaciones Según Catálogo Nacional de Estaciones del IDEAM y Otras

Entidades

La actividad consiste en una preselección de estaciones que se torna de gran importancia debido a que con esta depuración se puntualiza que series temporales de temperatura solicitar al IDEAM.

El primer paso es descargar el Catálogo Nacional de Estaciones del IDEAM y otras entidades desde su página principal www.ideam.gov.co, como se muestra en Anexo 1.

Según el catálogo nacional de estaciones del IDEAM, el cual relaciona la información de las estaciones hidrometeorológicas de propiedad del IDEAM y de otras entidades, actualizado el 22 de febrero de 2019, Colombia cuenta con 8858 estaciones hidrometeorológicas, cuyas categorías varían entre: Agrometeorológica, climática ordinaria, climática principal, limnigráfica, limnimétrica, mareográfica, meteorológica especial, pluviográfica, pluviométrica, radio sonda, sinóptica principal y sinóptica secundaria; y su estado está entre: Activo, en mantenimiento y suspendido. Se encuentran activas 5447, esto es solo el 61,49% de la totalidad descrita anteriormente.

Realizando un primer filtro por departamentos, 294 (3,32%) estaciones se encuentran en territorio Nortesantandereano y secuencialmente por estado se reducen a 146 (1,65%) las estaciones activas en el departamento.

De acuerdo a la categoría de este proyecto, las que proveen información referente son: Agrometeorológicas, climática ordinaria, climática principal, meteorológica especial y sinóptica principal; donde respectivamente se encuentran 42 (0,47%) estaciones.

Dado las fechas de instalación, estado activo y en relación con la variable temperatura, se seleccionan 21 estaciones meteorológicas, que posiblemente puedan aportar datos útiles y se reportan en actividad desde el año 1940 según el Catálogo Nacional de Estaciones del IDEAM y otras entidades.

Tal selección se da de acuerdo a que las estaciones meteorológicas definidas aparecen entre las más antiguas y activas actualmente por tanto pueden generar una información veraz, para analizar el comportamiento de la temperatura en la región. Ver Anexo 2. En la Tabla 1 se plasman las 21 estaciones seleccionadas.

Tabla 1. Estaciones meteorológicas seleccionadas

N°	Código	Nombre	Instalación
1	16035010	TIBU	1940
2	16015010	AEROPUERTO CAMILO DAZA	1941
3	16015060	BLONAY	1951
4	16025060	FRANCISCO ROMERO	1952
5	16025040	CINERA-VILLA OLGA	1965
6	16025010	CARMEN DE TONCHALA	1968
7	16055040	ABREGO CENTRO ADMO	1969
8	23195180	ESCUELA AGRICOLA CACHIRA	1972
9	16015020	ISER PAMPLONA	1972
10	37035010	TUNEBIA	1972
11	16015030	LA ESPERANZA	1972
12	16025030	SALAZAR	1973
13	16035030	SARDINATA	1973
14	16055010	AEROPUERTO AGUAS CLARAS	1973
15	16055020	TEORAMA	1973

N°	Código	Nombre	Instalación
16	16055060	LA PLAYA	1984
17	16015090	TAMA PARQUE NAL	1989
18	16055090	INSTITUCION AGRICOLA CONVENCION	1990
19	16055110	GABRIEL M BARRIGA	1991
20	16055100	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	1991
21	16015100	RAGONVALIA	2000

4.2 Recolección de la Data

La recolección de la data, parte de las 21 estaciones meteorológicas seleccionadas.

Para esta actividad se plantea como opción inicial el uso de la página principal del IDEAM que cuenta con la opción “consulta - descarga datos hidrometeorológicos” que enlaza con el sitio web DHIME <http://dhime.ideam.gov.co/atencionciudadano/>, donde para realizar una consulta se debe diligenciar las casillas que se muestran en la Figura 3 Sitio web Consulta - Descarga Datos Hidrometeorológicos y para este caso como parámetro temperatura, se encuentran 163 variables de entre las cuales se plasma algunas en la Tabla 2 y figuran en totalidad en Anexo 3.

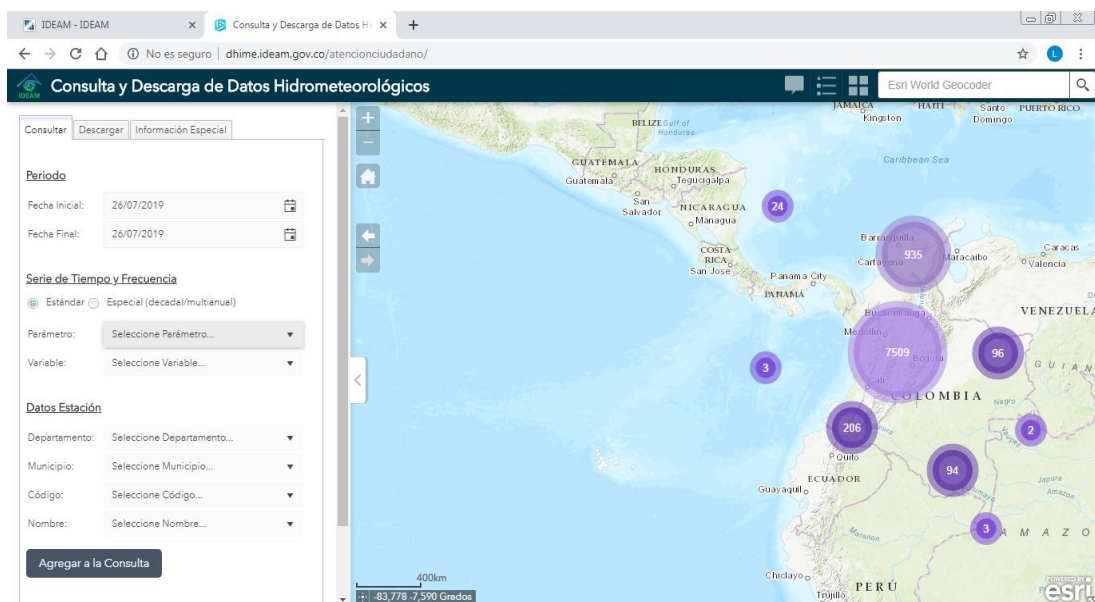


Figura 3. Sitio web Consulta - Descarga Datos Hidrometeorológicos

Tabla 2. Variables con parámetro temperatura

Variables con parámetro temperatura
Amplitud entre la temperatura máxima y mínima [T_AMPLITUD]
Amplitud entre la temperatura máxima y mínima a 10 centímetros [TA_10_AMPLITUD]
Amplitud entre la temperatura máxima y mínima a 2 metros [TA_2_AMPLITUD]
Punto de las 10 AM del máximo horario [TAMX10_MX_10_D]
Variables con parámetro temperatura
Temperatura del Aire a 10 cm máxima anual [TA10_MX_A]
Temperatura del Aire a 10 cm media diaria [TA10_MEDIA_D]
Temperatura del Aire a 2 metros media anual [TA2_MEDIA_A]
Temperatura del Aire a 2 metros media mensual [TA2_2_MEDIA_M]
Temperatura húmeda media anual [THSM_MEDIA_A]
Temperatura máxima absoluta anual [TMX_MX_A]
Temperatura máxima del Aire a 10 cm mínima diaria [TAMX10_MN_D]
Temperatura máxima del Aire a 2 metros máxima absoluta diaria [TAMX2_MX_D]
Temperatura máxima diaria [TMX_CON]
Temperatura mínima media anual [TMN_MEDIA_A]
Temperatura seca máxima anual [TSSM_MX_A]
Temperatura seca mínima diaria [TSTG_MN_D]
TEMPERATURA SECA PROMEDIO MENSUAL [TSSM_MEDIA_M_Mala]
TPR_CAL_MEDIA_D [TPR_CAL_MEDIA_D]

La consulta a través de esta plataforma se dificulta, pues sin saber concretamente que variable de temperatura usar para este proyecto y realizando en repetidas ocasiones consulta con diferentes variables, se imposibilita la descarga de datos y no se obtiene la información a través de esta plataforma.

Por lo anterior, se procede a realizar el registro en la página del Ideam, realizando una inscripción inicial con un correo electrónico en el cual se recibiría un link de verificación, dándose espera de una semana en la cual no se recibe dicho correo de secuencia a la inscripción.

Dado que no se ha podido obtener la información requerida se procede a realizar una llamada a la oficina del Ideam y siguiendo las instrucciones dadas se radica la primera solicitud de información realizando los pasos tal como se muestra en Anexo 4.

Radicado 1: 20199050021262

En radicado 1 se solicita registro de datos históricos de temperatura media de cinco (5) estaciones meteorológicas, que se muestran en Tabla 3.

Tabla 3. Data solicitada en radicado 1

Código	Nombre
16035010	TIBU [16035010]
16015010	AEROPUERTO CAMILO DAZA [16015010]
16015060	BLONAY [16015060]
16025060	FRANCISCO ROMERO [16025060]
16025040	CINERA-VILLA OLGA [16025040]

Radicado 2: 20199050021712

En la segunda solicitud y siguiendo los mismos pasos se solicita cinco estaciones más, relacionadas en la Tabla 4.

Tabla 4. Data solicitada en radicado 2

Código	Nombre
16025010	CARMEN DE TONCHALA [16025010]
16055040	ABREGO CENTRO ADMO [16055040]
23195180	ESCUELA AGRICOLA CACHIRA [23195180]
16015020	ISER PAMPLONA [16015020]
37035010	TUNEBIA [37035010]

Para las respuestas a los radicados 1 y 2 se recibe la información en archivos sin formato. Ésta es transformada y trabajada en formatos de Excel, la información llega en conjunto con valores Máximos, Medios y Mínimos de las estaciones solicitadas. En Anexo 5 se muestra el documento recibido y como se transforma a Excel para poder trabajar esta información.

Continuando con la recolección se radican dos solicitudes más, realizando este proceso por separado y por grupos de cinco, ver Tabla 5, y seis estaciones, ver Tabla 6, a la vez para que sea más ágil el envío. Radicado 3: 20199050029012

Tabla 5. Data solicitada en radicado 3

Código	Nombre
16015030	ESPERANZA LA [16015030]
16025030	SALAZAR [16025030]
16035030	SARDINATA [16035030]
16055010	AEROPUERTO AGUAS CLARAS [16055010]
16055020	TEORAMA [16055020]

Radicado 4: 20199050029022

Tabla 6. Data solicitada en radicado 4

Código	Nombre
16055060	PLAYA LA [16055060]
16015090	TAMA PARQUE NAL [16015090]
16055090	INSTITUCION AGRICOLA CONVENCION [16055090]
16055110	GABRIEL M BARRIGA [16055110]
16055100	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER [16055100]
16015100	RAGONVALIA [16015100]

En recibidos de radicados 3 y 4 solo llegan las 10 estaciones solicitadas y otras 24 no solicitadas. Esta información llega como documento de texto y su transformación se muestra en Anexo 6.

Finalizando la recolección se radica una última solicitud correspondiente a radicado 5. Ver Tabla 7.

Radicado 5: 20199050029022

Tabla 7. Data solicitada en radicado 5

Código	Nombre
16055110	GABRIEL M BARRIGA [16055110]
16055050	SANTA LUCIA [16055050]
16055070	SOLEDAD LA [16055070]
16015010	AEROPUERTO CAMILO DAZA [16015010]

De la selección dada según el Catálogo Nacional de Estaciones del IDEAM y otras entidades, se solicita la estación GABRIEL M BARRIGA la cual no llega en radicados anteriores, además SANTA LUCIA y SOLEDAD LA muestran una posible data entre 45 y 27 años de operación respectivamente desde su instalación y debido a que los valores mínimos de AEROPUERTO CAMILO DAZA no se reciben en la primera solicitud se realiza este radicado 5.

Como respuesta al radicado 5 sólo se obtiene información de AEROPUERTO CAMILO DAZA a partir del año 1990, por lo tanto, se decide no dar uso a ésta.

4.3 Análisis de Series Temporales de Temperatura

Para efectos de claridad de los datos, se expresa que la temperatura en este proyecto se recibe, transforma y entrega en grados centígrados (°C).

Se separa la información así:

Recibido 1: Información obtenida mediante radicados 1 y 2

Recibido 2: Información obtenida mediante radicados 3 y 4

En recibido 1 se obtiene data de 10 estaciones, que se relacionan en Tabla 8.

Tabla 8. Recibido 1

N°	Código	Nombre	Instalación	Data desde
1	16015010	AEROPUERTO CAMILO DAZA	1941	1976
2	16015020	ISER PAMPLONA	1972	1977
3	16015060	BLONAY	1951	1984
4	16025010	CARMEN DE TONCHALA	1968	1978
5	16025040	CINERA-VILLA OLGA	1965	1977
6	16025060	FRANCISCO ROMERO	1952	1984
7	16035010	TIBU	1940	1978
8	16055040	ABREGO CENTRO ADMO	1969	1977
9	23195180	ESCUELA AGRICOLA CACHIRA	1972	1978
10	37035010	TUNEBIA	1972	1977

Esta información llega en un archivo sin formato el cual se transforma a hojas de cálculo de Excel como se muestra en el Anexo 5 y es arreglada y trabajada como se muestra en el Anexo 7 para así realizar su respectivo análisis.

En recibido 2, obtenido mediante radicados 3 y 4, data información de 34 estaciones. (Tabla 9).

Tabla 9. Recibido 2

Nº	Código	Nombre	Instalación	Data desde
1	16010330	SAN FAUSTINO	1983	1979
2	16015010	AEROPUERTO CAMILO DAZA	1941	1976
3	16015020	ISER PAMPLONA	1972	1977
4	16015030	ESPERANZA LA	1972	1979
5	16015040	SANTA ISABEL	1968	1976
6	16015090	TAMA PARQUE NAL	1989	1989
7	16015100	RAGONVALIA	2000	2000
8	16015110	UNIV FRANCISO DE	2005	2005
9	16015120	UNIV DE PAMPLONA	2004	2011
10	16015130	ALCALDIA DE HERRAN	2005	2006
11	16015140	FCA.LA PALMITA	2010	2010
12	16015501	AEROPUERTO CAMILO DAZA - AUT	2015	2015
13	16025010	CARMEN DE TONCHALA	1968	1978
14	16025020	RISARALDA	1973	1979
15	16025030	SALAZAR	1973	1979
16	16025040	CINERA-VILLA OLGA	1965	1977
17	16025070	CEEZ ICA	1983	1984
18	16025080	H.J.C GRANJA	1994	1995
19	16035010	TIBU	1940	1978
20	16035020	PETROLEA	1973	1978
21	16035030	SARDINATA	1973	1978
22	16035040	ESTANCIA LA	1983	1984
23	16045020	EL TARRA	2014	1983
24	16055010	AEROPUERTO AGUAS CLARAS	1973	1979
25	16055020	TEORAMA	1973	1978
26	16055040	ABREGO CENTRO ADMO	1969	1977
27	16055060	PLAYA LA	1984	1985
28	16055080	MARAVILLA LA	1986	1987
29	16055090	INSTITUCION AGRICOLA CONVENCION	1990	1990
30	16055100	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	1991	1992
31	16055120	AGUAS DE LA VIRGEN	2006	2006
32	23195180	ESCUELA AGRICOLA CACHIRA	1972	1978
33	37015010	SILOS	1972	1983
34	37035010	TUNEBIA	1972	1977

Esta información llega en un archivo en formato de texto el cual se transforma a hojas de cálculo de Excel como se muestra en el Anexo 6 y es arreglada y trabajada como se muestra en el Anexo 8 para así realizar su respectivo análisis.

El recibido 1 se tiene que las fechas de instalación del Catálogo Nacional de Estaciones del IDEAM y otras entidades son diferentes a lo que se recibe, dónde por ejemplo en estación AEROPUERTO CAMILO DAZA se instala en 1941 y en lo reportado data desde 1976, también se muestra estado activo, pero, no data a la actualidad.

Se encuentra que en la misma estación del ejemplo anterior no incluye valores mínimos y en su lugar se repitieron los valores máximos, en las otras 9 estaciones si se recibe valores máximos, medios y mínimos.

A pesar de los inconvenientes de la estación AEROPUERTO CAMILO DAZA presenta un buen registro de datos desde 1976.

En los valores mínimos de ISER PAMPLONA se encontraron valores negativos, lo que indica un tratamiento de cuidado con esta estación a la hora de realizar su respectivo análisis.

Las estaciones BLONAY y FRANCISCO ROMERO cuentan con información de 8 años entre 1984 y 1994; a pesar de ser muy poco el registro, los 8 años reportados cuentan con muy buena información.

En recibido 2 llega información de 34 estaciones totales.

De las 11 solicitadas llegan 10 estaciones tal como se muestra en Tabla 10. Se adicionan 16 estaciones no solicitadas que se muestran en Tabla 11 y sucede atípicamente que llega

información de 8 estaciones que se solicitaron y se obtuvieron en recibido 1, relacionadas en la Tabla 12.

Tabla 10. Recibido 2 solicitado

N°	Código	Nombre	Instalación	Data desde
1	16015030	ESPERANZA LA	1972	1979
2	16015090	TAMA PARQUE NAL	1989	1989
3	16015100	RAGONVALIA	2000	2000
4	16025030	SALAZAR	1973	1979
5	16035030	SARDINATA	1973	1978
6	16055010	AEROPUERTO AGUAS CLARAS	1973	1979
7	16055020	TEORAMA	1973	1978
8	16055060	PLAYA LA	1984	1985
9	16055090	INSTITUCION AGRICOLA CONVENCION	1990	1990
10	16055100	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	1991	1992

Tabla 11. Recibido 2 no solicitado

N°	Código	Nombre	Instalación	Data desde
1	16010330	SAN FAUSTINO	1983	1979
2	16015040	SANTA ISABEL	1968	1976
3	16015110	UNIV FRANCISO DE	2005	2005
4	16015120	UNIV DE PAMPLONA	2004	2011
5	16015130	ALCALDIA DE HERRAN	2005	2006
6	16015140	FCA.LA PALMITA	2010	2010
7	16015501	AEROPUERTO CAMILO DAZA - AUT	2015	2015
8	16025020	RISARALDA	1973	1979
9	16025070	CEEZ ICA	1983	1984
10	16025080	H.J.C GRANJA	1994	1995
11	16035020	PETROLEA	1973	1978
12	16035040	ESTANCIA LA	1983	1984
13	16045020	EL TARRA	2014	1983
14	16055080	MARAVILLA LA	1986	1987
15	16055120	AGUAS DE LA VIRGEN	2006	2006
16	37015010	SILOS	1972	1983

Tabla 12. Recibido 2 estaciones duplicadas - no solicitadas

N°	Código	Nombre	Instalación	Data desde
1	16015010	AEROPUERTO CAMILO DAZA	1941	1976
2	16015020	ISER PAMPLONA	1972	1977
3	16025010	CARMEN DE TONCHALA	1968	1978
4	16025040	CINERA-VILLA OLGA	1965	1977
5	16035010	TIBU	1940	1978
6	16055040	ABREGO CENTRO ADMO	1969	1977
7	23195180	ESCUELA AGRICOLA CACHIRA	1972	1978
8	37035010	TUNEBIA	1972	1977

Dando claridad a la Tabla 12 y haciendo una breve comparación entre la data obtenida en el recibido 1 y el recibido 2 se muestra dispersión entre estas, encontrando que no coincide inicialmente los años desde donde datan, donde en recibido 1 como ejemplo estación AEROPUERTO CAMILO DAZA reporta desde 1976 en todos los valores y que en recibido 2 reporta desde 1976 sólo valores medios, este recibido sí reporta valores mínimos y conjunto con valores máximos datan desde 1983.

Observando se deduce que, generalmente, se recibe valores máximos y mínimos con mayor información y que para el caso de los valores medios es muy poca la data suministrada.

Entre el análisis del recibido 2 se encuentra la estación SAN FAUSTINO que en valores máximos y mínimos cuenta con 64 datos en el año 1983 y en valores medios tiene 4 años incompletos entre 1979 – 1982.

Se identifica que la estación ISER PAMPLONA no cuenta con valores negativos como si ocurre en recibido 1, donde reporta instalación desde el año 1972 y data desde 1977 en valores

máximos, medios y mínimos, y para el recibido 2, reporta instalación desde 1972 y data en valores medios desde 1977, pero, en valores máximos y mínimos data desde 1983.

SANTA ISABEL, muestra en valores máximos y mínimos información de 17 años entre 1983 y 2002, considerando que este es un buen reporte, sin embargo, en valores medios hay 8 años de información a partir del año 1976 hasta 1995, apreciando que sólo 5 años tienen buena información pero que para el lapso de 19 años no es sustancial.

La estación TAMA PARQUE NAL tiene buena información en valores máximos y mínimos, pero, en valores medios solo cuenta con 3 años de información, el año 1989 con 1 dato, el año 2000 con 30 datos y el año 2018 con 222 datos.

Se consideran estaciones nuevas las instaladas a partir del 2000, coincidiendo con la fecha de instalación de catálogo y de inicio de reporte, nombradas a continuación: RAGONVALIA, UNIV FRANCISO DE, UNIV DE PAMPLONA, ALCALDIA DE HERRAN, FCA.LA PALMITA, AEROPUERTO CAMILO DAZA – AUT y AGUAS DE LA VIRGEN reportando información a partir de 2000, 2005, 2011, 2006, 2010, 2015, 2006, respectivamente, para valores máximos, medios y mínimos.

Tenemos las estaciones RAGONVALIA, UNIV FRANCISO DE y la estación automática AEROPUERTO CAMILO DAZA – AUT, que cuentan con muy buena data, a partir de su fecha de instalación, respecto a las mencionadas anteriormente.

Destacando como información importante está que las estaciones UNIV FRANCISO DE y UNIV DE PAMPLONA, presentan datos dudosos.

Para la estación UNIV FRANCISO DE la elevación es de 311 m.s.n.m., ubicada en el Municipio de Cúcuta y contiene buena información entre lo reportado desde el año 2005 hasta el año 2018, pero, cuenta con una serie de datos dudosos. Analizando se tiene que la estación UNIV FRANCISO DE cuenta en valores máximos con 4199 datos, donde 145 corresponden a datos dudosos entre 0 y 10.4 °C, y 47 °C, en valores medios con 4061 datos, donde 146 corresponden a datos dudosos entre 0 y 20 °C, y valores mínimos con 4077 datos, donde 161 corresponden a datos dudosos entre 0 y 12 °C. De los 145 datos dudosos de valores máximos, 126 corresponden a 0°C, en valores medios, de los 146 datos dudosos, 124 corresponden a 0°C y de los 161 datos dudosos, de valores mínimos, 158 corresponden a 0°C. Los años que contemplan estas fallas son 2005, 2006, 2013 y 2014.

La estación UNIV DE PAMPLONA presenta una elevación de 2362 m.s.n.m., ubicada en el municipio de Pamplona. Dados los datos dudosos con los que cuenta, se detalla que en valores máximos tiene 1877 datos, donde 87 corresponden a datos dudosos entre 0 y 10 °C, y entre 29.2 y 48.4 °C, para valores medios presenta 1832 datos, de los cuales 74 corresponden a datos dudosos entre 0 y 1.4 °C, y entre 33.4 y 63.4 °C, y para valores mínimos presenta 1887 datos, de los cuales 94 corresponden a datos dudosos de 0°C. De los 87 datos considerados como dudosos, entre los valores máximos, 68 corresponden a 0°C, de los 74 datos considerados dudosos, en valores medios, 57 corresponden a 0°C. Al darse que todos estos datos dudosos se encuentran en el año 2014, y que la mayoría de estos son valores de 0°C, se considera una posible falla de la estación y por lo tanto estos valores serán catalogados como vacíos (NaN).

CEEZ ICA, cuenta con 4 años de información entre 1985 y 1989 para valores máximos y mínimos y en valores medios cuenta con 2 años, 1984 y 1987.

Las estaciones que llegan con 0 % de valores medios son H.J.C GRANJA, ESTANCIA LA, EL TARRA, MARAVILLA LA, INSTITUCION AGRICOLA CONVENCION.

Se observa generalmente que en valores máximos y mínimos los años 1984 y 1986, fueron años críticos para la recolección de temperaturas diarias, 19 estaciones de las 34 de recibido 2 no cuentan con alguno de estos 2 años mencionados, de estas 19 estaciones, sólo 2 tienen información en 1984 y 4 estaciones tienen en 1986.

El análisis de las series temporales permite indagar sobre la calidad de la información, el punto medio, variabilidad y simetría de los datos, además, los valores anuales máximos, medios y mínimos y obtener valores, vacíos y porcentaje de estos. Tal análisis se realiza siguiendo el capítulo 11 Estadística Hidrológica de Hidrología Aplicada por Ven Te Chow, David R. Maidment, Larry w. Mays (Chow, & Mays, 2010).

Se organizan las estaciones en formatos de Excel de forma anual, en tabla que identifica los meses, días, valores y características tales como: Código, nombre, fecha de instalación, latitud, longitud, elevación y municipio, como lo muestra la Figura 4. Utilizando los datos anuales se genera una gráfica de dispersión como se puede observar en la figura anterior, que muestra el comportamiento de la temperatura anualmente, detallada en la Figura 5.

ESTACION → 16015010 AEROPUERTO CAMILO DAZA
 INSTALACION → 1941 FEB
 LATITUD → 7.93027778 N(+)
 LONGITUD → 72.50916667 W(-)
 ELEVACION → 250 m.s.n.m.
 MUNICIPIO → CUCUTA

VALORES 363 99.2%
 VACIOS (NaN) 3 0.8%
 TOTAL 366 100%
 VALOR MAXIMO (°C) 37.6
 VALOR MEDIO (°C) 30.5
 VALOR MINIMO (°C) 22.0

MEDIA ARITMETICA (°C) 30.5
 MEDIANA (°C) 30.6
 MEDIA GEOMETRICA (°C) 30.3

VARIANZA (°C) 8.1
 DESVIACION ESTANDAR (°C) 2.8
 COEFICIENTE DE VARIACION 9.3%

COEFICIENTE DE ASIMETRIA -0.271

MES	DIA	VALOR (°C)
ENERO	1	28.0
	2	27.7
	3	27.3
	4	28.4
	5	27.2
	6	23.3
	7	29.2
	8	28.6
	9	27.8
	10	26.7
	11	27.4
	12	28.7
	13	28.6
	14	27.1
	15	26.2
	16	30.0
	17	28.8
	18	29.8
	19	27.0
	20	28.4
	21	27.0
	22	28.8
	23	28.8
	24	29.3
	25	30.0
	26	29.0
	27	24.6
	28	26.3
	29	26.4
	30	26.8
	31	25.2

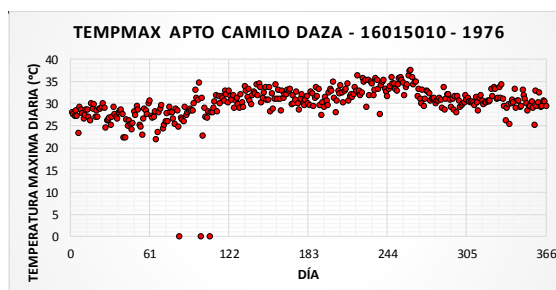


Figura 4 Organización de información reportada en formatos de Excel

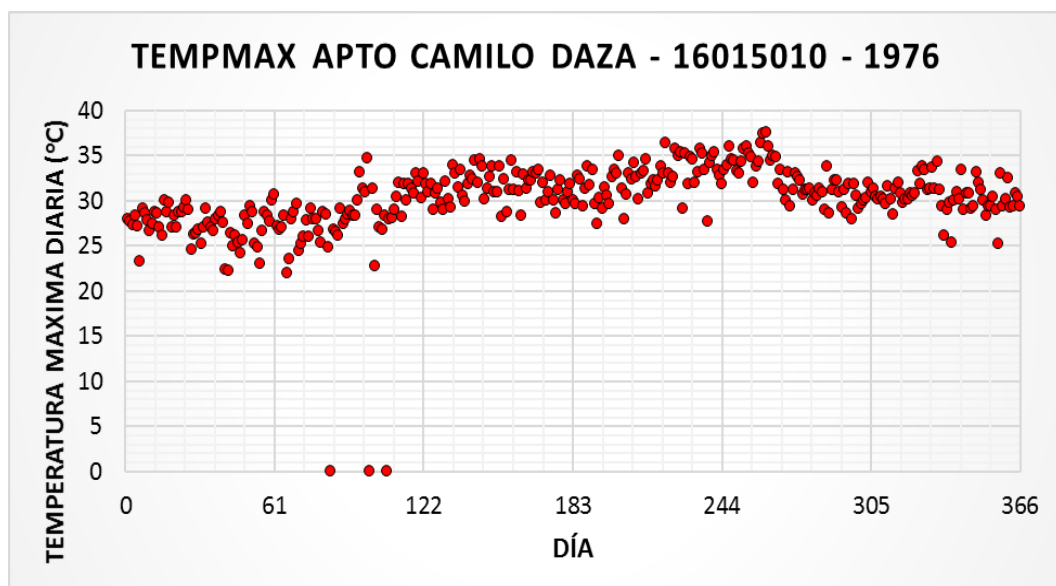


Figura 5. Gráfica de dispersión - Valores Máximos AEROPUERTO CAMILO DAZA año 1976

Analizando esta gráfica estación a estación, anualmente, se detecta una serie de datos dudosos, se relacionan estas estaciones en Tabla 13, las cuales, como se planteó para las estaciones UNIV FRANCISCO DE y UNIV DE PAMPLONA, se cambian como vacíos (NaN) y se demarcan en el Anexo 9.

Tabla 13. Estaciones con Datos Dudosos

PRIMER RECIBIDO	
ESTACIÓN	VARIABLE
APTO CAMILO DAZA	MEDIOS
ISER PAMPLONA	MEDIOS
BLONAY	MAXIMOS
TIBU	MAXIMOS
TIBU	MEDIOS
ABREGO CENTRO ADMO	MAXIMOS
ABREGO CENTRO ADMO	MINIMOS
ESCUELA AGRICOLA CACHIRA	MINIMOS
SEGUNDO RECIBIDO	
ESTACIÓN	VARIABLE
APTO CAMILO DAZA	MAXIMOS
APTO CAMILO DAZA	MEDIOS
APTO CAMILO DAZA	MINIMOS
ISER PAMPLONA	MEDIOS
SANTA ISABEL	MEDIOS
SANTA ISABEL	MINIMOS
TAMA PARQUE NAL	MAXIMOS
TAMA PARQUE NAL	MEDIOS
RAGONVALIA	MEDIOS
RAGONVALIA	MINIMOS
UNIV FRANCISCO DE	MAXIMOS
UNIV FRANCISCO DE	MEDIOS
UNIV FRANCISCO DE	MINIMOS
UNIV DE PAMPLONA	MAXIMOS
UNIV DE PAMPLONA	MEDIOS
UNIV DE PAMPLONA	MINIMOS
ALCALDIA DE HERRAN	MINIMOS

PRIMER RECIBIDO	
ESTACIÓN	VARIABLE
AEROPUERTO CAMILO DAZA - AUT	MINIMOS
CARMEN DE TONCHALA	MEDIOS
RISARALDA	MAXIMOS
SALAZAR	MEDIOS
CEEZ ICA	MAXIMOS
HJC GRANJA	MAXIMOS
TIBU	MAXIMOS
TIBU	MEDIOS
TIBU	MINIMOS
PETROLEA	MINIMOS
SARDINATA	MAXIMOS
LA ESTANCIA	MEDIOS
TEORAMA	MAXIMOS
ABREGO CENTRO ADMO	MEDIOS
LA MARAVILLA	MINIMOS
UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	MAXIMOS
UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	MEDIOS
UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	MINIMOS
AGUAS DE LA VIRGEN	MINIMOS
ESCUELA AGRICULA CACHIRA	MAXIMOS
ESCUELA AGRICULA CACHIRA	MEDIOS
ESCUELA AGRICULA CACHIRA	MINIMOS
SILOS	MINIMOS

De tal manera, sin datos dudosos, se organiza la información reportada, adjunta en el Anexo 10.

Generando los parámetros estadísticos descritos en Chow, & Mays (2010) media aritmética, mediana, media geométrica, varianza, desviación estándar, coeficiente de variación, coeficiente de asimetría, se analiza año a año entre 1976 hasta 2018 las series de temperatura diaria, ver Tabla 14, que ejemplifica este análisis hecho a todas las estaciones, (Ver Anexo 11). Estos pasos se realizan tanto en recibido 1 como en recibido 2.

En vista de que ciertas estaciones se repitieron en el segundo recibido y no cuentan con la misma información del recibido anterior, se efectúa una comparación.

Las estaciones repetidas son: aeropuerto Camilo Daza, ISER Pamplona, Carmen de Tonchala, Cinera-Villa Olga, Tibu, Abrego centro admo, Escuela Agrícola Cachira y Tunebia.

Según el rango de años utilizado entre 1976 y 2018, la cantidad de valores en una sola variable son 15706, y se deben obtener entre valores máximos, medios y mínimos 47118 como un 100 % de valores; con esto se observa:

La estación AEROPUERTO CAMILO DAZA cuenta en totalidad con 28206 valores (59,9%) en primer recibido, mientras que para el segundo con 25094 valores (53.3%). Se tiene que en el primer recibido solo cuenta con valores máximos y medios y aun así es superior en valores al segundo.

Tabla 14. Análisis De Series De Temperatura Diaria - Valores Máximos Aeropuerto Camilo Daza

AÑO	PUNTO MEDIO (°C)			VARIABILIDAD			SIMETR IA	VALORES (°C)			DATOS			
	Med.arit me.	Mediana	Med.geomet.	Varianza	Desv.estan.	Coef.varia.	Coef.asime.	Maximos	Medios	Minimos	Valores	%	Vacio s	%
1976	30.5	30.6	30.3	8.1	2.8	9.3%	-0.271	37.6	30.5	22.0	363	99.18%	3	0.82%
1977	32.4	32.7	32.3	4.7	2.2	6.7%	-0.682	36.4	32.4	24.2	361	98.90%	4	1.10%
1978	32.0	32.0	31.9	5.0	2.2	7.0%	-0.531	37.0	32.0	25.0	360	98.63%	5	1.37%
1979	31.2	31.2	31.1	5.1	2.3	7.3%	-0.089	36.5	31.2	25.0	242	66.30%	123	33.70 %
1980	32.2	32.2	32.1	5.5	2.3	7.3%	-0.676	37.0	32.2	22.8	331	90.44%	35	9.56%
1981	31.1	31.0	31.1	3.9	2.0	6.4%	-0.424	35.8	31.1	24.0	363	99.45%	2	0.55%
1982	31.3	31.4	31.2	5.2	2.3	7.3%	-0.386	36.8	31.3	22.0	365	100.00%	0	0.00%
1983	32.7	33.0	32.6	4.8	2.2	6.7%	-0.515	37.5	32.7	25.2	364	99.73%	1	0.27%
1984	31.3	31.4	31.2	6.1	2.5	7.9%	-0.536	36.4	31.3	23.6	272	74.32%	94	25.68 %
1985	30.9	31.0	30.8	5.5	2.3	7.6%	-0.153	37.2	30.9	24.2	362	99.18%	3	0.82%
1986	31.2	31.3	31.1	6.0	2.4	7.8%	-0.209	36.9	31.2	23.5	362	99.18%	3	0.82%
1987	32.9	33.2	32.9	5.2	2.3	6.9%	-0.663	37.8	32.9	24.3	321	87.95%	44	12.05 %
1988	31.8	31.8	31.7	5.7	2.4	7.5%	-0.505	37.0	31.8	23.4	366	100.00%	0	0.00%
1989	30.0	30.0	29.9	6.4	2.5	8.5%	-0.157	36.0	30.0	22.4	197	53.97%	168	46.03 %
1990	31.5	31.7	31.4	6.7	2.6	8.2%	-0.404	37.5	31.5	22.3	365	100.00%	0	0.00%
1991	32.3	32.6	32.2	5.3	2.3	7.2%	-0.327	37.3	32.3	25.0	362	99.18%	3	0.82%
1992	32.4	32.8	32.4	5.0	2.2	6.9%	-0.310	38.2	32.4	25.4	366	100.00%	0	0.00%
1993	32.4	32.6	32.3	4.9	2.2	6.9%	-0.335	37.6	32.4	25.5	362	99.18%	3	0.82%
1994	31.6	31.5	31.5	6.9	2.6	8.3%	-0.292	37.7	31.6	22.6	363	99.45%	2	0.55%
1995	32.5	32.5	32.4	4.3	2.1	6.4%	-0.341	36.8	32.5	24.6	365	100.00%	0	0.00%
1996	31.8	32.0	31.7	5.9	2.4	7.6%	-0.445	37.0	31.8	23.8	366	100.00%	0	0.00%
1997	32.7	33.2	32.6	7.6	2.8	8.5%	-0.406	38.7	32.7	24.8	363	99.45%	2	0.55%
1998	33.2	33.7	33.1	6.0	2.5	7.4%	-0.738	38.0	33.2	24.5	354	96.99%	11	3.01%
1999	31.0	31.0	30.9	7.8	2.8	9.0%	-0.267	36.8	31.0	23.5	363	99.45%	2	0.55%
2000	31.3	31.9	31.2	8.5	2.9	9.3%	-0.525	37.0	31.3	23.0	360	98.36%	6	1.64%
2001	32.8	33.0	32.7	5.6	2.4	7.2%	-0.386	37.9	32.8	25.8	363	99.45%	2	0.55%
2002	32.5	32.5	32.4	5.6	2.4	7.3%	-0.170	37.9	32.5	25.6	359	98.36%	6	1.64%
2003	33.0	33.1	33.0	5.4	2.3	7.0%	-0.568	37.6	33.0	22.4	365	100.00%	0	0.00%

AÑO	PUNTO MEDIO (°C)			VARIABILIDAD			SIMETR IA	VALORES (°C)			DATOS			
	Med.arit me.	Mediana	Med.geomet.	Varianza	Desv.estan.	Coef.varia.	Coef.asime.	Maximos	Medios	Minimos	Valores	%	Vacio s	%
2004	31.8	32.0	31.7	6.6	2.6	8.1%	-0.287	37.4	31.8	23.8	364	99.45%	2	0.55%
2005	31.9	32.0	31.8	7.5	2.7	8.6%	-0.348	37.9	31.9	24.1	351	96.16%	14	3.84%
2006	32.3	32.2	32.2	7.8	2.8	8.6%	-0.230	38.2	32.3	23.8	364	99.73%	1	0.27%
2007	32.4	32.4	32.4	5.2	2.3	7.0%	-0.397	37.3	32.4	24.4	365	100.00%	0	0.00%
2008	31.6	31.6	31.5	6.0	2.5	7.8%	-0.215	37.2	31.6	23.8	364	99.45%	2	0.55%
2009	32.0	32.0	31.9	7.8	2.8	8.7%	-0.346	38.2	32.0	22.8	365	100.00%	0	0.00%
2010	32.7	33.0	32.6	5.8	2.4	7.4%	-0.382	38.8	32.7	22.8	365	100.00%	0	0.00%
2011	31.3	31.3	31.2	7.5	2.7	8.7%	-0.338	37.2	31.3	21.6	361	98.90%	4	1.10%
2012	32.3	32.2	32.2	6.3	2.5	7.7%	-0.133	37.9	32.3	24.5	365	99.73%	1	0.27%
2013	32.6	32.8	32.5	6.2	2.5	7.6%	-0.188	38.2	32.6	25.7	362	99.18%	3	0.82%
2014	32.7	32.7	32.7	5.0	2.2	6.8%	-0.070	38.0	32.7	27.3	365	100.00%	0	0.00%
2015	32.5	32.5	32.4	6.0	2.4	7.5%	-0.340	37.4	32.5	25.5	243	66.58%	122	33.42 %
2016	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.0	0	0.00%	366	100.00 %
2017	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.0	0	0.00%	365	100.00 %
2018	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	0.0	0	0.00%	365	100.00 %
										TOTA L	13939	88.75%	1767	11.25 %
												TOTAL DE DIAS		15706

Con la estación ISER PAMPLONA se tiene 40644 valores (86.3%) de primer recibido y 26431 valores (56.1%) en el segundo recibido.

La estación CARMEN DE TONCHALA presenta 39550 valores (83.9%) en primer recibido y 26873 valores (57%) en segundo recibido.

CINERA-VILLA OLGA, tiene 31714 valores (67.3%) en primer recibido y en segundo recibido 22736 (48.3%).

TIBU, contiene 23629 valores (50.1%) de primer recibido y 14714 valores (31.2%) para el segundo recibido.

En ABREGO CENTRO ADMO se obtienen 36011 valores (76.4%) en primer recibido y 23890 valores (50.7%) en segundo recibido.

La estación ESCUELA AGRICOLA CACHIRA cuenta con 38622 valores (82%) para el primer recibido y 25114 valores (53.3%) para el segundo recibido.

Y como última estación del comparativo está TUNEBIA con 34531 valores (73.3%) en primer recibido y 34344 valores (72.9%) en segundo recibido.

Se muestra mediante Tabla 15 la diferencia de porcentajes entre las estaciones repetidas en recibidos 1 y 2.

Tabla 15. Comparación estaciones repetidas en recibidos 1 y 2

Estaciones	Datos		Datos	
	Recibido 1		Recibido 2	
	Valores	%	Valores	%
APTO CAMILO DAZA	28206	59.9%	25094	53.3%
ISER PAMPLONA	40644	86.3%	26431	56.1%
CARMEN DE TONCHALA	39550	83.9%	26873	57.0%
CINERA-VILLA OLGA	31714	67.3%	22736	48.3%
TIBU	23629	50.1%	14714	31.2%
ABREGO CENTRO ADMO	36011	76.4%	23890	50.7%
ESCUELA AGRICOLA CACHIRA	38622	82.0%	25114	53.3%
TUNEBIA	34531	73.3%	34344	72.9%

Con esta comparación se obtiene que porcentualmente es superior la información en el recibido 1 de las estaciones descritas, por lo cual, se decide continuar con este proyecto utilizando el primer recibido con las 8 estaciones que se muestran en el comparativo, las 2 estaciones no repetidas que son BLONAY y FRANCISCO ROMERO y de las 34 del segundo recibido se trabaja con 26 estaciones excluyendo las que se encuentran repetidas. Esta decisión se toma teniendo en cuenta que para el llenado de datos es conveniente tener mayor información para mejores resultados en la regresión lineal multivariada. De esta manera se programa la utilización total de 36 estaciones para continuar con el análisis de las series temporales de temperatura diaria.

Para efecto de referenciar las estaciones se denominaron estación 1 hasta estación 36 respectivamente como se muestra en Tabla 16.

Tabla 16. Denominación de estaciones

ESTACION	CODIGO	NOMBRE	MUNICIPIO	ABREV.
EST. 1	16010330	SAN FAUSTINO	Cúcuta	SF
EST. 2	16015010	AEROPUERTO CAMILO DAZA	Cúcuta	ACD
EST. 3	16015020	ISER PAMPLONA	Pamplona	IP
EST. 4	16015030	ESPERANZA LA	Ragonvalia	LESP
EST. 5	16015040	SANTA ISABEL	Cúcuta	SI
EST. 6	16015060	BLONAY	Chinácota	BLON
EST. 7	16015090	TAMA PARQUE NAL	Herrán	TAMAPN
EST. 8	16015100	RAGONVALIA	Ragonvalia	RAG
EST. 9	16015110	UNIV FRANCISO DE	Cúcuta	UFPSC
EST. 10	16015120	UNIV DE PAMPLONA	Pamplona	UP
EST. 11	16015130	ALCALDIA DE HERRAN	Herrán	HERR
EST. 12	16015140	FCA.LA PALMITA	Pamplonita	FP
EST. 13	16015501	AEROPUERTO CAMILO DAZA - AUT	Cúcuta	ACDAUT
EST. 14	16025010	CARMEN DE TONCHALA	Cúcuta	CT
EST. 15	16025020	RISARALDA	El Zulia	RIS
EST. 16	16025030	SALAZAR	Salazar	SAL
EST. 17	16025040	CINERA-VILLA OLGA	Cúcuta	CVO
EST. 18	16025060	FRANCISCO ROMERO	Salazar	FR
EST. 19	16025070	CEEZ ICA	Cúcuta	CICA
EST. 20	16025080	H.J.C GRANJA	Cucutilla	HJCG
EST. 21	16035010	TIBU	Tibú	TIBU
EST. 22	16035020	PETROLEA	Tibú	PET
EST. 23	16035030	SARDINATA	Sardinata	SAR
EST. 24	16035040	ESTANCIA LA	Tibú	LEST
EST. 25	16045020	EL TARRA	El Tarra	TARRA
EST. 26	16055010	AEROPUERTO AGUAS CLARAS	Ocaña	AAC
EST. 27	16055020	TEORAMA	Teorama	TEO
EST. 28	16055040	ABREGO CENTRO ADMO	Ábrego	ACA
EST. 29	16055060	PLAYA LA	La Playa	PLAYA
EST. 30	16055080	MARAVILLA LA	San Calixto	MAR
EST. 31	16055090	INSTITUCION AGRICOLA CONVENCION	Convención	IAC
EST. 32	16055100	UNIVERSIDAD FRANCISCO DE PAULA SANTANDER	Ocaña	UFPSC
EST. 33	16055120	AGUAS DE LA VIRGEN	Ocaña	AVIR
EST. 34	23195180	ESCUELA AGRICOLA CACHIRA	Cáchira	EAC
EST. 35	37015010	SILOS	Silos	SIL
EST. 36	37035010	TUNEBIA	Toledo	TUN

Continuando se genera el porcentaje de datos registrados (PDR) con el análisis de series de temperatura diaria, mostrándolo en mapa de calor, (Ver Figura 5). Esta actividad se adjunta en Anexo 12.

El PDR indica que si la estación cuenta con información entre 0 – 20% se determina de muy baja calidad de data, por ello exterioriza con color rojo, si tiene entre 20% a 40% de reporte se enmarca en rango relativamente baja y se demarca con color naranja, si cuenta entre 40% a 60% la expresa el color amarillo como data intermedia, para porcentajes de información entre 60% a 80% lo expresa el color verde claro como una estación relativamente buena y en data entre 80% a 100% la información se considera muy buena y se delimita con el color verde.

Tal mapa de calor adjunta información de las 36 estaciones objeto de análisis entre 1976 – 2018.

Tabla 17. Porcentaje de Datos Registrados (PDR) – Valores Máximos

% INFORMACION		ESTACION									
AÑO	EST. 1	EST. 2	EST. 3	EST. 4	EST. 5	EST. 6	EST. 7	EST. 8	EST. 9	EST. 10	
0-20											
20-40											
40-60											
60-80											
80-100											
1976	0.0%	99.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1977	0.0%	98.9%	7.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1978	0.0%	98.6%	83.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1979	0.0%	66.3%	89.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1980	0.0%	90.4%	89.3%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1981	0.0%	99.5%	97.5%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1982	0.0%	100.0%	99.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	
1983	17.5%	99.7%	97.5%	98.4%	95.3%	0.0%	0.0%	0.0%	0.0%	0.0%	
1984	0.0%	74.3%	98.6%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%	0.0%	
1985	0.0%	99.2%	98.6%	64.1%	99.7%	100.0%	0.0%	0.0%	0.0%	0.0%	
1986	0.0%	99.2%	8.5%	0.0%	0.0%	99.5%	0.0%	0.0%	0.0%	0.0%	
1987	0.0%	87.9%	98.1%	88.8%	80.8%	98.1%	0.0%	0.0%	0.0%	0.0%	
1988	0.0%	100.0%	99.5%	93.4%	97.8%	100.0%	0.0%	0.0%	0.0%	0.0%	
1989	0.0%	54.0%	96.7%	74.2%	81.9%	100.0%	69.9%	0.0%	0.0%	0.0%	
1990	0.0%	100.0%	92.3%	96.2%	52.6%	99.7%	94.5%	0.0%	0.0%	0.0%	
1991	0.0%	99.2%	95.6%	94.8%	74.5%	0.0%	95.3%	0.0%	0.0%	0.0%	
1992	0.0%	100.0%	97.0%	94.0%	47.5%	100.0%	93.7%	0.0%	0.0%	0.0%	
1993	0.0%	99.2%	95.9%	94.2%	69.6%	0.0%	95.3%	0.0%	0.0%	0.0%	
1994	0.0%	99.5%	92.6%	93.7%	95.3%	99.5%	93.2%	0.0%	0.0%	0.0%	
1995	0.0%	100.0%	86.3%	93.4%	37.0%	0.0%	92.1%	0.0%	0.0%	0.0%	
1996	0.0%	100.0%	38.5%	98.4%	69.1%	0.0%	99.5%	0.0%	0.0%	0.0%	
1997	0.0%	99.5%	0.0%	99.2%	77.3%	0.0%	71.2%	0.0%	0.0%	0.0%	
1998	0.0%	97.0%	79.2%	99.2%	48.5%	0.0%	85.8%	0.0%	0.0%	0.0%	
1999	0.0%	99.5%	99.7%	93.4%	69.9%	0.0%	85.2%	0.0%	0.0%	0.0%	
2000	0.0%	98.4%	99.5%	99.5%	33.9%	0.0%	87.2%	40.4%	0.0%	0.0%	
2001	0.0%	99.5%	99.7%	97.0%	0.0%	0.0%	59.2%	98.4%	0.0%	0.0%	
2002	0.0%	98.4%	100.0%	96.2%	14.0%	0.0%	35.1%	97.8%	0.0%	0.0%	
2003	0.0%	100.0%	98.9%	100.0%	0.0%	0.0%	0.0%	97.8%	0.0%	0.0%	
2004	0.0%	99.5%	98.4%	98.6%	0.0%	0.0%	0.0%	95.4%	0.0%	0.0%	
2005	0.0%	96.2%	94.5%	95.6%	0.0%	0.0%	19.5%	91.5%	2.7%	0.0%	
2006	0.0%	99.7%	96.7%	95.1%	0.0%	0.0%	57.3%	91.2%	77.0%	0.0%	
2007	0.0%	100.0%	100.0%	95.3%	0.0%	0.0%	93.7%	98.6%	95.6%	0.0%	
2008	0.0%	99.5%	99.7%	92.1%	0.0%	0.0%	75.1%	96.2%	95.4%	0.0%	
2009	0.0%	100.0%	99.2%	97.5%	0.0%	0.0%	55.9%	99.2%	80.5%	0.0%	
2010	0.0%	100.0%	99.2%	95.3%	0.0%	0.0%	92.1%	93.7%	84.9%	0.0%	
2011	0.0%	98.9%	99.7%	95.9%	0.0%	0.0%	86.6%	87.4%	98.4%	10.4%	
2012	0.0%	99.7%	100.0%	98.6%	0.0%	0.0%	93.4%	99.2%	91.0%	16.1%	
2013	0.0%	99.2%	97.0%	82.2%	0.0%	0.0%	86.3%	100.0%	70.7%	93.4%	
2014	0.0%	100.0%	97.5%	99.7%	0.0%	0.0%	84.7%	98.4%	88.5%	54.2%	
2015	0.0%	66.6%	99.5%	90.4%	0.0%	0.0%	62.7%	99.2%	97.0%	20.5%	
2016	0.0%	0.0%	99.7%	97.5%	0.0%	0.0%	88.3%	97.8%	38.0%	98.4%	
2017	0.0%	0.0%	47.7%	83.0%	0.0%	0.0%	64.1%	100.0%	93.2%	92.6%	
2018	0.0%	0.0%	0.0%	99.7%	0.0%	0.0%	60.8%	100.0%	97.3%	98.9%	

Entre los objetivos de este proyecto está dejar la información recibida en formatos de fácil manipulación. Para realizar la reconstrucción de series de tiempo de temperatura diaria, en el departamento de Norte de Santander, se crea inicialmente un formato en Excel el cual contiene un comentario con el código de la estación, nombre, la variable (TMAX, TMED, TMIN), y su unidad de temperatura (°C). Este se divide en columnas con la información del año, mes, día, día juliano y valor de la estación. Las variables se denominan: TMAX para temperatura máxima,

TMED para temperatura media y TMIN para temperatura mínima. Esta información (Ver Figura 4) está comprendida entre los años 1976 y 2018, adjunta en Anexo 13.

Estos archivos de columna se convierten a bloc de notas, para el uso futuro de esta información mediante otras herramientas tecnológicas. (Ver Anexo 14).

Los archivos de columna se unifican en una hoja de Excel, que es de utilidad para el análisis general estación a estación en un periodo de 40 años (Ver Tabla 18), y determinación de los parámetros estadísticos, comprendidos en Anexo 15, y también para la generación de box plot.

Tabla 18. Tmax16015010 - Temperatura Máxima 1976-2018 Apto Camilo Daza

№	16015010	APTO	CAMILO	DAZA	TMAX	°C
1976	1	1	1	1	28.0	
1976	1	2	2	2	27.7	
1976	1	3	3	3	27.3	
1976	1	4	4	4	28.4	
1976	1	5	5	5	27.2	
1976	1	6	6	6	23.3	
1976	1	7	7	7	29.2	
1976	1	8	8	8	28.6	
1976	1	9	9	9	27.8	
1976	1	10	10	10	26.7	
1976	1	11	11	11	27.4	
1976	1	12	12	12	28.7	
1976	1	13	13	13	28.6	
1976	1	14	14	14	27.1	
1976	1	15	15	15	26.2	
1976	1	16	16	16	30.0	
1976	1	17	17	17	28.8	
1976	1	18	18	18	29.8	
1976	1	19	19	19	27.0	
1976	1	20	20	20	28.4	
1976	1	21	21	21	27.0	
1976	1	22	22	22	28.8	
1976	1	23	23	23	28.8	
1976	1	24	24	24	29.3	
1976	1	25	25	25	30.0	
1976	1	26	26	26	29.0	
1976	1	27	27	27	24.6	
1976	1	28	28	28	26.3	
1976	1	29	29	29	26.4	
1976	1	30	30	30	26.8	
1976	1	31	31	31	25.2	

Tabla 19. Análisis general de series de temperatura diaria – Valores Máximos

Est	Punto medio (°c)			Variabilidad			Simetria	Valores (°c)			Datos			
	Med.aritme	Mediana	Med.geomet.	Varianza	Desv.estan.	Coef.varia.	Coef.asime.	Maximos	Medios	Minimos	Valores	%	Vacios	%
EST. 1	30.8	31.0	30.7	5.7	2.4	7.8%	-0.027	37.4	30.8	24.6	64	0.4%	15642	99.6 %
EST. 2	32.0	32.1	31.9	6.5	2.6	8.0%	-0.381	38.8	32.0	21.6	13939	88.7%	1767	11.3 %
EST. 3	19.4	19.5	19.4	2.9	1.7	8.7%	-0.276	25.8	19.4	12.0	13031	83.0%	2675	17.0 %
EST. 4	20.2	20.2	20.1	3.7	1.9	9.5%	0.117	30.6	20.2	13.8	11632	74.1%	4074	25.9 %
EST. 5	30.8	31.0	30.7	4.0	2.0	6.5%	-0.471	37.2	30.8	22.8	4181	26.6%	11525	73.4 %
EST. 6	26.0	26.2	25.9	5.2	2.3	8.7%	-0.407	33.1	26.0	17.6	3276	20.9%	12430	79.1 %
EST. 7	16.5	16.4	16.4	3.1	1.8	10.7%	0.149	24.8	16.5	10.8	7953	50.6%	7753	49.4 %
EST. 8	21.9	22.0	21.7	4.3	2.1	9.5%	-0.504	29.2	21.9	14.0	6509	41.4%	9197	58.6 %
EST. 9	32.9	32.9	32.7	8.7	3.0	9.0%	-0.312	39.7	32.9	20.9	4054	25.8%	11652	74.2 %
EST. 10	20.4	20.5	20.4	3.1	1.8	8.7%	-0.252	26.9	20.4	13.7	1770	11.3%	13936	88.7 %
EST. 11	23.6	23.7	23.5	4.4	2.1	8.9%	-0.358	30.1	23.6	14.7	3003	19.1%	12703	80.9 %
EST. 12	24.7	24.8	24.6	5.3	2.3	9.3%	-0.262	30.6	24.7	17.4	1470	9.4%	14236	90.6 %
EST. 13	32.0	32.1	31.9	6.4	2.5	7.9%	-0.387	37.7	32.0	23.0	1124	7.2%	14582	92.8 %
EST. 14	32.4	32.8	32.3	6.7	2.6	8.0%	-0.560	39.0	32.4	22.4	12128	77.2%	3578	22.8 %
EST. 15	30.8	31.1	30.7	5.1	2.3	7.3%	-0.504	39.6	30.8	21.4	3788	24.1%	11918	75.9 %
EST. 16	26.7	27.0	26.6	5.6	2.4	8.9%	-0.418	36.8	26.7	18.2	11963	76.2%	3743	23.8 %
EST. 17	32.3	32.4	32.3	3.4	1.9	5.7%	-0.074	39.6	32.3	23.4	9446	60.1%	6260	39.9 %
EST. 18	27.6	28.0	27.5	4.7	2.2	7.9%	-0.605	33.8	27.6	19.4	3286	20.9%	12420	79.1 %
EST. 19	29.8	30.0	29.7	3.8	1.9	6.5%	-0.478	34.4	29.8	24.0	686	4.4%	15020	95.6 %
EST. 20	25.0	25.3	24.9	4.6	2.2	8.6%	-0.456	29.8	25.0	18.2	963	6.1%	14743	93.9 %
EST. 21	32.9	33.0	32.8	3.8	2.0	6.0%	-0.330	39.6	32.9	24.2	6625	42.2%	9081	57.8 %

Est	Punto medio (°c)			Variabilidad			Simetria	Valores (°c)			Datos			
	Med.aritme	Mediana	Med.geomet.	Varianza	Desv.estan.	Coef.varia.	Coef.asime.	Maximos	Medios	Minimos	Valores	%	Vacios	%
EST. 22	32.1	32.0	32.1	5.1	2.3	7.0%	0.189	39.6	32.1	23.4	5825	37.1%	9881	62.9%
EST. 23	31.0	31.2	30.9	7.5	2.7	8.8%	-0.326	39.4	31.0	21.4	10445	66.5%	5261	33.5%
EST. 24	29.6	29.6	29.5	3.3	1.8	6.1%	-0.399	33.6	29.6	25.2	90	0.6%	15616	99.4%
EST. 25	30.9	31.2	30.8	2.5	1.6	5.2%	-0.909	38.2	30.9	22.2	5452	34.7%	10254	65.3%
EST. 26	25.4	25.6	25.3	3.8	2.0	7.7%	-0.370	32.0	25.4	17.0	7695	49.0%	8011	51.0%
EST. 27	24.3	24.6	24.3	3.6	1.9	7.8%	-0.332	34.4	24.3	17.0	9176	58.4%	6530	41.6%
EST. 28	27.2	27.2	27.1	2.9	1.7	6.2%	-0.111	33.4	27.2	20.0	11220	71.4%	4486	28.6%
EST. 29	24.5	24.6	24.4	5.3	2.3	9.4%	-0.419	31.4	24.5	16.0	11142	70.9%	4564	29.1%
EST. 30	20.8	20.8	20.7	4.3	2.1	10.0%	0.053	28.8	20.8	14.8	5480	34.9%	10226	65.1%
EST. 31	25.3	25.6	25.2	4.5	2.1	8.4%	-0.631	32.4	25.3	16.2	9057	57.7%	6649	42.3%
EST. 32	25.6	25.8	25.5	3.2	1.8	7.0%	-0.270	32.2	25.6	18.2	9115	58.0%	6591	42.0%
EST. 33	23.3	23.4	23.2	4.3	2.1	8.9%	-0.343	28.7	23.3	14.7	1855	11.8%	13851	88.2%
EST. 34	22.3	22.2	22.2	2.4	1.6	7.0%	0.028	27.4	22.3	16.0	13110	83.5%	2596	16.5%
EST. 35	16.0	16.0	15.8	4.4	2.1	13.1%	0.013	24.2	16.0	10.0	9871	62.8%	5835	37.2%
EST. 36	29.5	30.0	29.4	4.9	2.2	7.5%	-0.637	36.0	29.5	21.6	11562	73.6%	4144	26.4%

Con formato columna se genera diagrama de cajas o box plot (Ver Figura 5). De esta manera se muestra la serie de datos numéricos que evidencia la variación que se da en temperaturas, valor máximo, mínimo, la dispersión y simetría. (Anexo 16).

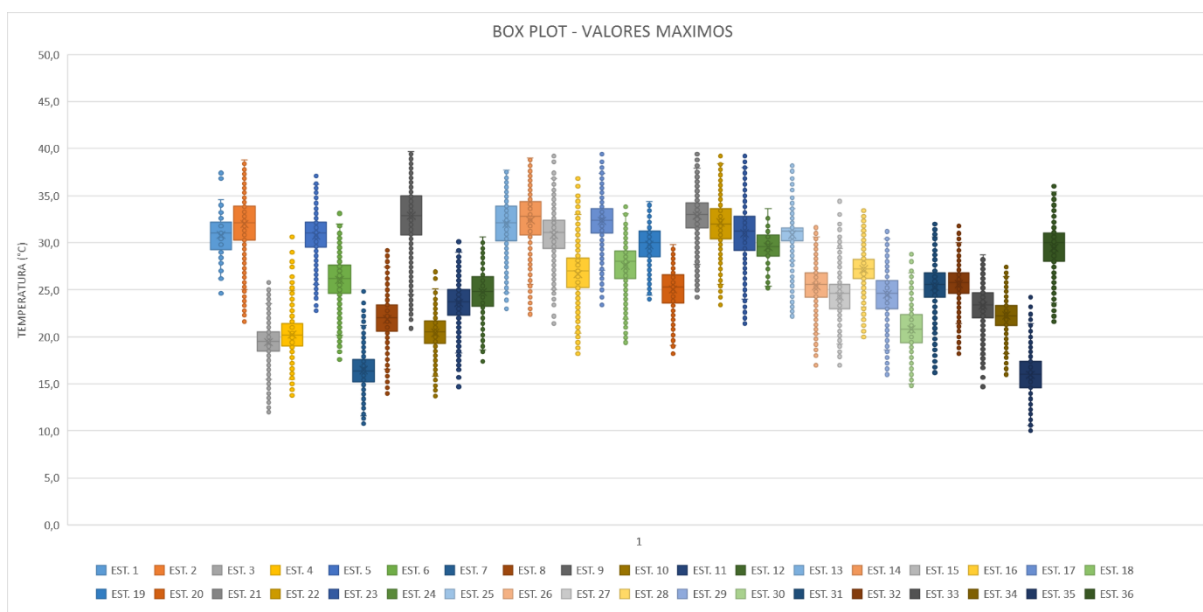


Figura 5. Box plot - Valores Máximos

Con ayuda del porcentaje de datos registrados (PDR) y el análisis general se obtiene tanto un estimado entre data buena y data baja, como también, parámetros estadísticos, el valor máximo, medio y mínimo de la variable analizada, cantidad de valores y vacíos con sus respectivos porcentajes.

Entonces, como ejemplo se tienen los valores máximos de la EST. 1 - SAN FAUSTINO; en PDR muestra entre 0 – 20% demarcada con color rojo su data. En su análisis general se caracteriza así, la media aritmética es la sumatoria de todos los valores dividida entre la cantidad de estos, dándose un promedio de los mismos que para esta estación es 30.8 °C. La mediana detecta la posición central, siendo para esta estación 31.0 °C. La media geométrica es la raíz a la

n (cantidad de valores reportada) del producto de los valores, se obtiene para la estación ejemplo 30.7°C . La varianza es la diferencia al cuadrado entre los valores reales y la media aritmética, dividido n veces (cantidad de valores), se genera para esta estación 5.7°C^2 . Este dato se da en $^{\circ}\text{C}^2$ e indica un rango por encima o debajo de la media aritmética. La desviación estándar es la raíz cuadrada de la varianza, esto reduce el rango y además tiende a ser menor. Este valor va acorde a los valores reales debido a la unidad que maneja ($^{\circ}\text{C}$) y se encuentra para esta estación en 2.4°C . El coeficiente de variación es la equivalencia porcentual de la desviación estándar, dado para esta estación es de 7.8%. El coeficiente de asimetría para esta estación es de -0.027 , lo que expresa que tiene una tendencia asimétrica negativa, que indica que hay más cantidad de datos por debajo de la media aritmética. Seguido, el valor máximo es de 37.4°C , valor medio de 30.8°C y valor mínimo de 24.6°C ; estos tres últimos valores se condensan para exclusivamente temperaturas máximas obtenidas en EST. 1. Ahora, los datos se expresan en cantidad y porcentaje; obteniéndose 64 valores y 15642 vacíos, con 0.4% y 99.6% respectivamente.

Referidos a cantidad de valores y vacíos se determinan con un total de valores de 15706, por el rango de años utilizados.

Analizando exclusivamente temperaturas máximas, se tiene como el ejemplo dado, entre 0 – 20% de valores, se encuentran 9 estaciones que son, EST. 1, EST. 24, EST. 19, EST. 20, EST. 13, EST. 12, EST. 10, EST. 33 y EST. 11 con porcentaje de valores concernientes de 0.4%, 0.6%, 4.4%, 6.1%, 7.2%, 9.4%, 11.3%, 11.8% y 19.1%. Las anteriores estaciones se colocan así con la denominación de muy baja calidad de data.

Entre 20% – 40% de valores máximos se encuentran 8 estaciones que son, EST. 6, EST. 18, EST. 15, EST. 9, EST. 5, EST. 25, EST. 30 y EST. 22 con porcentaje de valores atribuidos de

20.9%, 20.9%, 24.1%, 25.8%, 26.6%, 34.7%, 34.9% y 37,1%. Las anteriores estaciones se designan con calidad de data relativamente baja.

Entre 40% – 60% de valores máximos se encuentran 7 estaciones que son, EST. 8, EST. 21, EST. 26, EST. 7, EST. 31, EST. 32 y EST. 27 con porcentaje de valores de 41.4%, 42.2%, 49.0%, 50.6%, 57.7%, 58.0% y 58.4% respectivos y enmarcados en calidad de data intermedia.

Entre 60% – 80% de valores máximos se encuentran 9 estaciones que son, EST. 17, EST. 35, EST. 23, EST. 29, EST. 28, EST. 36, EST. 4, EST. 16 y EST. 14 con porcentaje de valores correspondientes de 60.1%, 62.8%, 66.5%, 70.9%, 71.4%, 73.6%, 74.1%, 76.2% y 77.2%. estas estaciones se delimitan en calidad de data relativamente buena.

Las estaciones con calidad de información denominada muy buena se encuentran entre 80% – 100% y son, EST. 3, EST. 34 y EST. 2. La EST. 3 tiene 13031 valores que corresponden a un 83.0%, la EST. 34 con 13110 valores representado en 83.5% y La EST. 2 reporta 13939 valores con 88.7%, lo que la hace la estación con mayor información en temperaturas máximas.

Se reportan temperaturas entre valor máximo y valor mínimo así: Estación 1 entre 24.6 °C y 37.4 °C, estación 2 entre 21.6 °C y 38.8 °C, estación 3 entre 12.0 °C y 25.8 °C, estación 4 entre 13.8 °C y 30.6 °C, estación 5 entre 22.8 °C y 37.2 °C, estación 6 entre 17.6 °C y 33.1 °C, estación 7 entre 10.8 °C y 24.8 °C, estación 8 entre 14.0 °C y 29.2 °C, estación 9 entre 20.9 y 39.7 °C, estación 10 entre 13.7 y 26.9 °C, estación 11 entre 14.7 °C y 30.1 °C, estación 12 entre 17.4% y 30.6%, estación 13 entre 23 °C y 37.7 °C, estación 14 entre 22.4 °C y 39.0 °C, estación 15 entre 21.4 °C y 39.6 °C, estación 16 entre 18.2 °C y 36.8 °C, estación 17 entre 23.4 °C y 39.6 °C, estación 18 entre 19.4 y 33.8%, estación 19 entre 24.0 °C y 34.4 °C, estación 20 entre 18.2 °C y 29.8 °C, estación 21 entre 24.2 °C y 39.6 °C, estación 22 entre 23.4 °C y 39.6 °C, estación

23 entre 21.4 °C y 39.4 °C, estación 24 entre 25.2 °C y 33.6 °C, estación 25 entre 22.2 °C y 38.2 °C, estación 26 entre 17.0 °C y 32.0 °C, estación 27 entre 17.0 °C y 34.4 °C, estación 28 entre 20.0 °C y 33.4 °C, estación 29 entre 16.0 °C y 31.4 °C, estación 30 entre 14.8 °C y 28.8 °C, estación 31 entre 16.2 °C y 32.4 °C, estación 32 entre 18.2 °C y 32.2 °C, estación 33 entre 14.7 °C y 28.7 °C, estación 34 entre 16.0 y 27.4 °C, estación 35 entre 10.0 °C y 24.2 °C y estación 36 entre 21.6 °C y 36.0 °C

Llevando el análisis a temperaturas medias se tiene que entre 0 – 20% de valores y denominadas de muy baja calidad de data, se encuentran 23 estaciones que son, EST. 31, EST. 30, EST. 25, EST. 20, EST. 24, EST. 7, EST. 32, EST. 19, EST. 29, EST. 1, EST. 26, EST.15, EST. 13, EST. 4, EST. 12, EST. 5, EST. 10, EST. 33, EST. 23, EST. 27, EST. 22, EST. 16, EST. 11, con porcentaje de valores correspondientes de 0%, 0 %, 0%, 0%, 0.3%, 1.6%, 2.2%, 2.3%, 4.7%, 5.4%, 6.0%, 7.1%, 7.1%, 8.3%, 9.4%, 11.1%, 11.2%, 11.6%, 12.5%, 12.5%, 13.7%, 13.9%, 18.6%, lo que se estima que en calidad de data muy baja se encuentra el 63.9% de las estaciones reportadas en esta variable.

Con data relativamente baja entre 20% - 40% de valores, se encuentran 3 estaciones que son, EST. 6, EST.18 y EST.9 con porcentaje de valores referentes de 20.9% 20.9 y 24.9%.

En la data intermedia entre 40% - 60% de valores, están 2 estaciones que son, EST. 8 y EST. 21 con porcentaje de valores de 41.4% y 54.6% respectivamente.

Para información relativamente buena, entre 60% - 80% hay 2 estaciones que son EST. 35 y EST. 17 con 62.8% y 72.7% respectivamente.

Entre 80 – 100 % consideradas de muy buena calidad de data, se tiene 6 estaciones y son, EST. 28, EST. 36, EST.34, EST. 3, EST. 14 y EST. 2 con porcentaje de valores de 80.3%, 81,0%, 82.3%, 87.7%, 87.8% y 90.8%, siendo en temperaturas medias, también, la EST. 2 la que cuenta con mayor cantidad de información.

Se reportan temperaturas entre valor máximo y valor mínimo así: La estación 1 entre 21.5 °C y 32.1 °C, estación 2 entre 20.4 °C y 32.7 °C, estación 3 entre 9.9 °C y 19.5 °C, estación 4 entre 13.0 °C y 21.3 °C, estación 5 entre 22.7 °C y 30.4 °C, estación 6 entre 15.3 °C y 25.8 °C, estación 7 entre 10.9 °C y 16.3 °C, estación 8 entre 13.7 °C y 23.7 °C, estación 9 entre 21.0 °C y 33.4 °C, estación 10 entre 10.7 °C y 17.7 °C, estación 11 entre 12.9 °C y 21.3 °C, estación 12 entre 16.3 °C y 26.0 °C, estación 13 entre 21.8 °C y 32.4 °C, estación 14 entre 20.9 °C y 33.6 °C, estación 15 entre 19.9 °C y 33.1 °C, estación 16 entre 18.3 °C y 26.8 °C, estación 17 entre 21.8 °C y 33.3 °C, estación 18 entre 17.6 °C y 26.0 °C, estación 19 entre 21.6 °C y 31.2 °C, estación 21 entre 22.0 °C y 33.0 °C, estación 22 entre 22.4 °C y 33.3 °C, estación 23 entre 21.3 °C y 31.1 °C, estación 24 entre 21.3 °C y 27.1 °C, estación 26 entre 17.4 °C y 24.7 °C, estación 27 entre 17.0 °C y 26.5 °C, estación 28 entre 15.0 °C y 25.8 °C, estación 29 entre 16.2 °C y 23.6 °C, estación 32 entre 18.5 °C y 24.1 °C, estación 33 entre 12.9 °C y 21.3 °C, estación 34 entre 11.7 °C y 22.9 °C, estación 35 entre 8.5 °C y 18.2 °C y estación 36 entre 19.8 °C y 31.4 °C.

En temperaturas medias las estaciones 20, 25, 30 y 31 no cuentan con información.

Llevando el análisis a temperaturas mínimas se tiene que entre 0 – 20% de valores y denominadas de muy baja calidad de data, se encuentran 10 estaciones que son, EST. 2, EST. 1, EST. 24, EST. 19, EST. 20, EST. 13, EST. 12, EST. 10, EST. 33, EST. 11 con porcentaje de valores de respectivos de 0%, 0.4%, 0.6%, 4.4%, 6.1%, 7.2%, 9.4%, 11.4%, 11.7%, 17.8%.

Con data relativamente baja entre 20% - 40% de valores se encuentran 8 estaciones que son, EST. 6, EST. 18, EST. 15, EST. 9, EST. 5, EST. 25, EST. 30, EST. 22 con porcentaje de valores de respectivos de 20.9%, 20.9%, 24.1%, 24.9%, 26.6%, 34.7%, 34.9%, 37.1%.

Entre 40% - 60% de valores se encuentran 7 estaciones que son, EST. 8, EST. 26, EST. 7, EST. 21, EST. 31, EST. 32, EST. 27, porcentajes 41.4%, 49.0%, 50.6%, 53.7%, 57.7%, 58%, 58.4%, enmarcadas estas estaciones dentro de data intermedia.

Entre 60% - 80% de valores se encuentran 8 estaciones que son, EST. 35, EST. 36, EST. 23, EST. 17, EST. 29, EST. 4, EST. 16, EST. 28 con porcentaje de valores de 62.8%, 65.3%, 66.5%, 69.0%, 70.9%, 74.1%, 76.2%, 77.5% respectivamente, estas estaciones se enmarcan entre data relativamente buena.

En data muy buena entre 80% - 100% de valores se encuentran 3 estaciones que son, EST. 34, EST. 14, EST. 3, con porcentaje de valores correspondientes de 80.1%, 86.8%, 88.1%.

Se reportan temperaturas entre valor máximo y valor mínimo así: Estación 1 entre 21.0 °C y 26.4 °C, estación 3 entre -1.0 °C y 16.0 °C, estación 4 entre 7.2 °C y 19.6 °C, estación 5 entre 18.0 °C y 30.3 °C, estación 6 entre 10.0 °C y 20.5 °C, estación 7 entre 1.5 °C y 16.8 °C, estación 8 entre 11.0 °C y 21 °C, estación 9 entre 17.3 °C y 29.8 °C, estación 10 entre 2.0 °C y 15.7 °C, estación 11 entre 9.2 °C y 17.8 °C, estación 12 entre 10.8 °C y 21.8 °C, estación 13 entre 18.7 °C y 29.1 °C, estación 14 entre 14.6 °C y 28.0 °C, estación 15 entre 19.0 °C y 30.2 °C, estación 16 entre 15.0 °C y 25.6 °C, estación 17 entre 13.2 °C y 28.0 °C, estación 18 entre 11.8 °C y 20.8 °C, estación 19 entre 19.4 °C y 29.4 °C, estación 20 entre 11.1 °C y 21.1 °C, estación 21 entre 15.2 °C y 26.4 °C, estación 22 entre 18.6 °C y 29.8 °C, estación 23 entre 16.0 °C y 30.6 °C, estación 24 entre 18.8 °C y 24.2 °C, estación 25 entre 17.0 °C y 29.2 °C, estación 26 entre 11.2 °C y 24.6

°C, estación 27 entre 14.2 °C y 24.6 °C, estación 28 entre 7.0 °C y 19.8 °C, estación 29 entre 8.6 °C y 22.0 °C, estación 30 entre 11.4 °C y 22.6 °C, estación 31 entre 15.6 °C y 24.4 °C, estación 32 entre 9.8 °C y 21.4 °C, estación 33 entre 11.4 °C y 17.8 °C, estación 34 entre 4.6 °C y 17.3 °C, estación 35 entre 2.4 °C y 14.8 °C y estación 36 entre 12.2 °C y 24.6 °C.

En temperaturas mínimas la estación 2 no cuenta con información.

4.4 Selección de Estaciones para Llenado de Datos

Basados tanto en PDR como en el análisis general, se seleccionan las estaciones que tienen más del 40% de su data (Ver Tabla 20), independientemente para cada variable, máximos, medios o mínimos.

Tabla 20. Estaciones seleccionadas para llenado de datos

Valores máximos	Valores medios	Valores mínimos
EST. 2	EST. 2	EST. 3
EST. 34	EST. 14	EST. 14
EST. 3	EST. 3	EST. 34
EST. 14	EST. 34	EST. 28
EST. 16	EST. 36	EST. 16
EST. 4	EST. 28	EST. 4
EST. 36	EST. 17	EST. 29
EST. 28	EST. 35	EST. 17
EST. 29	EST. 21	EST. 23
EST. 23	EST. 8	EST. 36
EST. 35		EST. 35
EST. 17		EST. 27
EST. 27		EST. 32
EST. 32		EST. 31
EST. 31		EST. 21
EST. 7		EST. 7
EST. 26		EST. 26
EST. 21		EST. 8
EST. 8		

4.5 Aplicación de Procedimientos Estándar para Determinar Datos Faltantes en las Series Seleccionadas

Se genera matriz de coeficiente de correlación lineal, para las estaciones seleccionadas anteriormente y para cada variable de temperatura. Esta matriz se rige en la misma gama de colores del PDR.

El coeficiente de correlación (R) toma valores entre -1 y 1, y su signo indica el sentido de la relación, entre más cerca a estos valores de -1 o +1, mejor correlación lineal presentan las estaciones y cuando R es cero, no existe relación lineal. Cuando R es igual a 1 la correlación es positiva perfecta, esto indica que cuando una estación aumenta, la otra también lo hace en proporción constante, caso que se da cuando la correlación se realiza matricialmente entre la misma estación.

Para realizar la correlación se utilizan las estaciones seleccionadas para el llenado de datos, pues, estas presentan la mayor información reportada, así sus correlaciones son las mejores que se pueden obtener.

	EST. 2	EST. 3	EST. 4	EST. 7	EST. 8	EST. 14	EST. 16	EST. 17	EST. 21	EST. 23	EST. 26	EST. 27	EST. 28	EST. 29	EST. 31	EST. 32	EST. 34	EST. 35	EST. 36
EST. 2	1.00	0.59	0.65	0.20	0.66	0.92	0.93	0.72	0.67	0.84	0.61	0.66	0.56	0.70	0.72	0.60	0.39	0.10	0.29
EST. 3		1.00	0.65	0.44	0.67	0.62	0.62	0.41	0.41	0.54	0.43	0.41	0.44	0.46	0.47	0.41	0.38	0.41	0.38
EST. 4			1.00	0.34	0.75	0.66	0.69	0.48	0.48	0.62	0.49	0.47	0.50	0.55	0.53	0.48	0.45	0.28	0.37
EST. 7				1.00	0.27	0.24	0.25	0.13	0.17	0.15	0.20	0.18	0.19	0.14	0.18	0.14	0.18	0.45	0.35
EST. 8					1.00	0.74	0.75	0.56	0.56	0.62	0.47	0.47	0.47	0.56	0.55	0.49	0.41	0.26	0.36
EST. 14						1.00	0.84	0.69	0.64	0.81	0.60	0.62	0.54	0.69	0.70	0.59	0.39	0.13	0.33
EST. 16							1.00	0.68	0.62	0.81	0.59	0.62	0.53	0.68	0.70	0.59	0.42	0.18	0.35
EST. 17								1.00	0.68	0.72	0.59	0.62	0.54	0.66	0.68	0.59	0.35	0.05	0.13
EST. 21									1.00	0.67	0.64	0.63	0.50	0.64	0.71	0.63	0.34	0.13	0.14
EST. 23										1.00	0.62	0.65	0.54	0.73	0.72	0.62	0.38	0.12	0.24
EST. 26											1.00	0.59	0.65	0.74	0.66	0.75	0.45	0.15	0.21
EST. 27												1.00	0.50	0.66	0.76	0.58	0.25	0.08	0.19
EST. 28													1.00	0.70	0.54	0.68	0.56	0.11	0.15
EST. 29														1.00	0.74	0.76	0.48	0.06	0.17
EST. 31															1.00	0.65	0.28	0.10	0.20
EST. 32																1.00	0.53	0.12	0.15
EST. 34																	1.00	0.21	0.16
EST. 35																		1.00	0.35
EST. 36																			1.00

Figura 6. Matriz de coeficiente de correlación lineal - Valores Máximos

	EST. 2	EST. 3	EST. 8	EST. 14	EST. 17	EST. 21	EST. 28	EST. 34	EST. 35	EST. 36
EST. 2	1.00	0.66	0.65	0.90	0.80	0.73	0.57	0.47	0.22	0.24
EST. 3		1.00	0.67	0.65	0.63	0.59	0.59	0.56	0.34	0.24
EST. 8			1.00	0.69	0.65	0.62	0.51	0.47	0.38	0.40
EST. 14				1.00	0.75	0.62	0.56	0.46	0.19	0.24
EST. 17					1.00	0.71	0.54	0.50	0.21	0.25
EST. 21						1.00	0.52	0.47	0.29	0.28
EST. 28							1.00	0.55	0.26	0.22
EST. 34								1.00	0.25	0.19
EST. 35									1.00	0.34
EST. 36										1.00

Figura 7. Matriz de coeficiente de correlación lineal - Valores Medios

	EST. 3	EST. 4	EST. 7	EST. 8	EST. 14	EST. 16	EST. 17	EST. 21	EST. 23	EST. 26	EST. 27	EST. 28	EST. 29	EST. 31	EST. 32	EST. 34	EST. 35	EST. 36
EST. 3	1.00	0.60	0.52	0.35	0.49	0.37	0.25	0.22	0.47	0.42	0.37	0.31	0.44	0.39	0.45	0.65	0.65	0.43
EST. 4		1.00	0.52	0.57	0.49	0.51	0.29	0.29	0.47	0.52	0.47	0.29	0.49	0.50	0.56	0.58	0.63	0.37
EST. 7			1.00	0.36	0.38	0.40	0.25	0.24	0.36	0.45	0.42	0.25	0.41	0.43	0.45	0.49	0.53	0.45
EST. 8				1.00	0.36	0.53	0.34	0.29	0.35	0.36	0.47	0.25	0.34	0.52	0.41	0.41	0.41	0.30
EST. 14					1.00	0.35	0.39	0.29	0.57	0.43	0.45	0.14	0.45	0.41	0.48	0.46	0.49	0.41
EST. 16						1.00	0.35	0.32	0.44	0.41	0.54	0.24	0.39	0.61	0.41	0.45	0.41	0.24
EST. 17							1.00	0.39	0.40	0.39	0.31	0.27	0.36	0.43	0.35	0.35	0.30	0.35
EST. 21								1.00	0.34	0.41	0.35	0.20	0.37	0.44	0.37	0.32	0.31	0.28
EST. 23									1.00	0.49	0.41	0.29	0.59	0.45	0.49	0.50	0.48	0.32
EST. 26										1.00	0.40	0.40	0.67	0.51	0.70	0.60	0.53	0.31
EST. 27											1.00	0.15	0.37	0.73	0.39	0.40	0.40	0.18
EST. 28												1.00	0.38	0.21	0.30	0.42	0.29	0.22
EST. 29													1.00	0.45	0.65	0.56	0.53	0.34
EST. 31														1.00	0.45	0.44	0.42	0.30
EST. 32															1.00	0.61	0.60	0.45
EST. 34																1.00	0.66	0.41
EST. 35																	1.00	0.47
EST. 36																		1.00

Figura 8. Matriz de coeficiente de correlación lineal - Valores mínimos

Con esta correlación se determina como se relaciona cada estación con las demás, ordenándose de mayor a menor para cada una de ellas y así poder iniciar con la regresión.

Para definir la cantidad de variables a aplicar en el llenado de datos y por ende la cantidad de regresiones a generar para una sola estación, se realiza el cálculo del error medio cuadrático (EMC) en 3 estaciones modelo, las cuales son EST. 3, EST. 14, EST. 34, y sólo para estas 3, se determina regresión hasta con 8 variables, esto sólo para observar cómo se comporta el error a medida que aumenta la cantidad de variables, entre más variables menor es el error, estableciendo así qué cantidad de variables es la óptima a utilizar.

Inicialmente se hace regresión lineal univariada, lo que indica que se trabaja la estación a simular y la mejor estación de correlación. Hallando la regresión lineal univariable, se obtiene datos simulados de la estación evaluada, restando lo simulado con los valores reales de la estación se obtiene el error, elevando este al cuadrado y calculando su promedio se obtiene el EMC. Para conocer su porcentaje, el EMC se divide entre el promedio de la estación evaluada.

Ecuación regresión lineal univariada:

$$Y = C_0 + C_1X_1$$

Continuando se efectúa regresión lineal multivariada, esta permite evaluar más de una variable, lo que se realiza así:

Se ingresa la estación a simular en conjunto con las que mejor correlaciona y se genera regresión lineal multivariada con 2, 3, 4 variables, o las que se determine utilizar. Se realiza el mismo procedimiento que con una variable y se generan datos simulados para la estación evaluada.

Ecuación regresión lineal multivariada:

$$Y = C_0 + C_1X_1 + C_2X_2 + \dots + C_nX_n$$

Donde,

C: Contantes

Y: Variable dependiente

X: Variable independiente

Se ejemplifica la generación de EMC en Figura 8, mostrando la EST. 3 - ISER PAMPLONA, con dos variables, donde su correlación se da para el ejemplo con la EST. 8 - RAGONVALIA y EST. 4 - LA ESPERANZA, generando en esta regresión lineal multivariada la EST. 3 SIMULADA con un error medio cuadrático de 7.2% que corresponde a 1.4 °C.

A cada EMC generado (Ver Anexo 17) se le crea tabla de dispersión entre la estación evaluada y la estación simulada, evidenciando su tendencia, (Ver Figura 9), se detalla el rango de temperatura real entre 13 °C hasta 25 °C y en los valores simulados temperaturas entre 15 °C hasta 24 °C, lo cual muestra una similitud con los valores reales. También se genera grafica de barras para mostrar la diferencia de error, donde se puede apreciar una tendencia de temperatura similar, superior o inferior al valor real, entre 2 °C y -2 °C, (Ver Figura 10). Además se adjunta grafica de barras representando el rango de error (Ver Figura 11), donde se observa que la mayor cantidad de datos se encuentra entre 0 °C hasta 3 °C, demostrando lo que expresa la gráfica anterior.

AÑO	MES	DIA	DATO N°	EST. 3	EST. 8	EST. 4	EST. 3	SIM	ERROR	ERROR^2
2000	8	1	214	8980	18.5	21.2	19.2	19.1	0.6	0.4
2000	8	2	215	8981	20.0	23.4	21.2	20.4	0.4	0.2
2000	8	3	216	8982	20.5	23.8	21.0	20.5	0.0	0.0
2000	8	4	217	8983	20.5	21.2	19.0	19.1	-1.4	2.1
2000	8	5	218	8984	19.0	19.4	18.0	18.1	-0.9	0.7
2000	8	6	219	8985	18.0	21.0	18.0	18.7	0.7	0.5
2000	8	7	220	8986	18.5	21.2	21.6	19.8	1.3	1.6
2000	8	8	221	8987	17.5	19.8	17.0	18.0	0.5	0.3
2000	8	9	222	8988	19.5	20.4	21.0	19.3	-0.2	0.0
2000	8	10	223	8989	19.5	21.8	21.0	19.8	0.3	0.1
2000	8	11	224	8990	20.5	23.4	22.0	20.6	0.1	0.0
2000	8	12	225	8991	20.0	20.8	20.2	19.2	-0.8	0.6
2000	8	14	227	8993	20.5	22.0	20.2	19.7	-0.8	0.7
2000	8	15	228	8994	20.5	21.2	19.0	19.1	-1.4	2.1
2000	8	16	229	8995	19.5	22.8	21.8	20.4	0.9	0.8
2000	8	17	230	8996	19.5	22.2	21.0	20.0	0.5	0.2
2000	8	18	231	8997	19.5	20.8	23.2	20.1	0.6	0.3
2000	8	19	232	8998	18.5	18.0	18.0	17.7	-0.8	0.7
2000	8	20	233	8999	20.5	23.2	22.0	20.6	0.1	0.0
2000	8	21	234	9000	19.5	22.0	23.0	20.4	0.9	0.9
2000	8	22	235	9001	21.5	23.0	23.2	20.8	-0.7	0.4
2000	8	23	236	9002	19.5	22.4	21.6	20.2	0.7	0.5
2000	8	24	237	9003	20.0	21.8	23.0	20.4	0.4	0.1
2000	8	25	238	9004	21.5	22.6	21.2	20.1	-1.4	1.8
2000	8	26	239	9005	21.0	23.0	24.0	21.1	0.1	0.0
2000	8	27	240	9006	18.5	19.4	20.0	18.7	0.2	0.0
2000	8	28	241	9007	21.5	22.0	20.4	19.7	-1.8	3.2
2000	8	29	242	9008	20.5	22.6	21.2	20.1	-0.4	0.1
2000	8	30	243	9009	22.0	22.6	20.4	19.9	-2.1	4.3
2000	8	31	244	9010	19.0	18.2	17.8	17.7	-1.3	1.8
2000	9	1	245	9011	19.5	22.4	20.0	19.7	0.2	0.1
2000	9	2	246	9012	19.5	20.4	17.0	18.2	-1.3	1.6
2000	9	3	247	9013	22.5	22.8	22.0	20.4	-2.1	4.3
2000	9	4	248	9014	19.5	19.2	18.8	18.3	-1.2	1.5
2000	9	5	249	9015	17.0	19.0	19.0	18.3	1.3	1.6
2000	9	6	250	9016	17.5	20.2	20.0	19.0	1.5	2.2
2000	9	7	251	9017	19.5	22.8	19.6	19.8	0.3	0.1
2000	9	8	252	9018	20.5	23.8	20.8	20.5	0.0	0.0
2000	9	9	253	9019	21.0	23.4	19.0	19.8	-1.2	1.4
2000	9	10	254	9020	20.5	22.4	19.0	19.5	-1.0	1.0
2000	9	11	255	9021	21.0	21.0	20.4	19.4	-1.6	2.7
2000	9	12	256	9022	19.5	21.4	21.6	19.8	0.3	0.1
2000	9	13	257	9023	22.0	24.0	21.8	20.8	-1.2	1.4
2000	9	14	258	9024	21.0	24.0	21.0	20.6	-0.4	0.2
2000	9	15	259	9025	21.0	22.6	22.2	20.4	-0.6	0.3
2000	9	16	260	9026	20.5	22.0	18.0	19.1	-1.4	2.1
2000	9	18	262	9028	21.0	20.6	17.0	18.3	-2.7	7.3
2000	9	19	263	9029	22.0	23.0	24.0	21.1	-0.9	0.9
2000	9	20	264	9030	21.0	22.4	21.0	20.0	-1.0	1.0
2000	9	21	265	9031	18.5	21.2	18.2	18.8	0.3	0.1
2000	9	22	266	9032	18.5	20.0	19.0	18.6	0.1	0.0
2000	9	23	267	9033	19.5	21.8	20.0	19.5	0.0	0.0
2000	9	24	268	9034	19.5	19.4	18.0	18.1	-1.4	1.8
2000	9	25	269	9035	18.5	20.8	18.4	18.7	0.2	0.1
2000	9	26	270	9036	21.0	22.4	20.6	19.9	-1.1	1.2
2000	9	27	271	9037	19.0	20.0	18.0	18.4	-0.6	0.4
2000	9	28	272	9038	21.0	22.4	21.0	20.0	-1.0	1.0
2000	9	29	273	9039	21.0	19.6	21.0	19.0	-2.0	3.9
2000	9	30	274	9040	20.5	22.0	21.0	19.9	-0.6	0.4
2000	10	1	275	9041	20.5	22.4	19.8	19.7	-0.8	0.7
2000	10	2	276	9042	19.0	20.0	19.0	18.6	-0.4	0.1
2000	10	3	277	9043	19.5	22.0	18.8	19.3	-0.2	0.0
2000	10	4	278	9044	20.0	21.8	20.6	19.7	-0.3	0.1
2000	10	5	279	9045	21.5	21.8	19.8	19.5	-2.0	4.1
2000	10	6	280	9046	21.0	23.4	21.0	20.4	-0.6	0.4

ERROR MEDIO CUADRATICO		
EMC	1.4	7.2%
RANGOS DE ERROR		
>=4°	4	0.07%
>=3° - <4°	40	0.71%
>=2° - <3°	250	4.43%
>=1° - <2°	841	14.90%
>1° - <1°	3347	59.28%
>-2° - <=-1°	929	16.45%
>-3° - <=-2°	220	3.90%
>-4° - <=-3°	13	0.23%
<=-4°	2	0.04%
TOTAL	5646	100%
RANGO DE ERROR FINAL		
0 - 0,9	3347	59.28%
1 - 1,9	1770	31.35%
2 - 2,9	470	8.32%
3 - 3,9	53	0.94%
>=4	6	0.11%
TOTAL	5646	100%

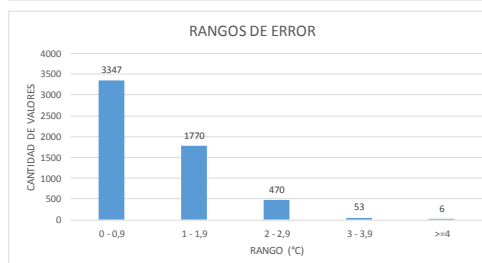
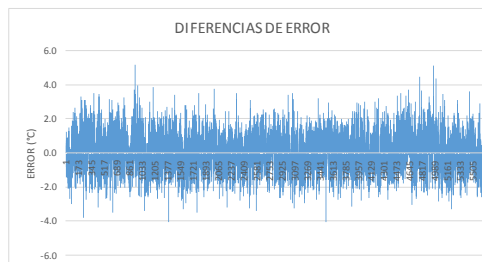
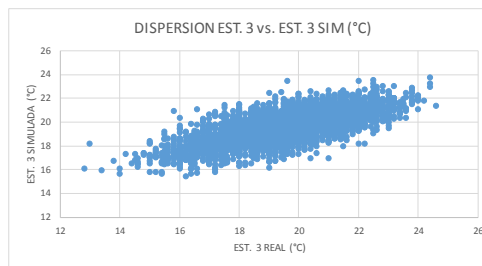


Figura 9. Generación de EMC de estación 3 simulada

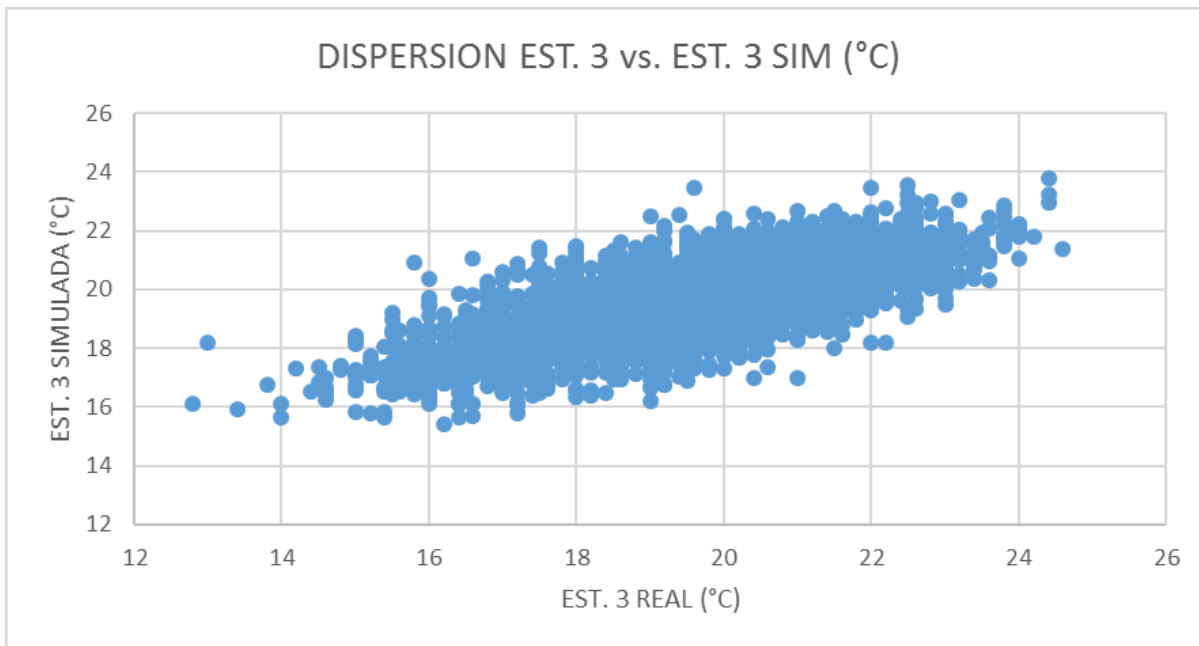


Figura 10. Dispersión EST. 3 vs. EST. 3 SIM

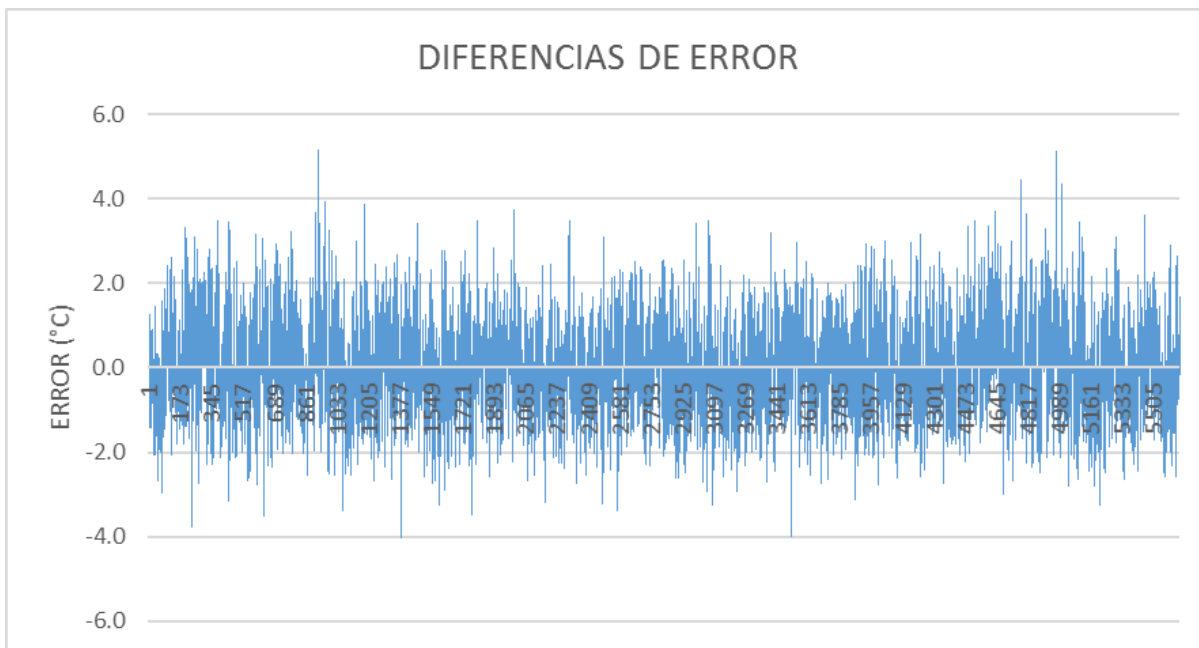


Figura 11. Diferencias de Error

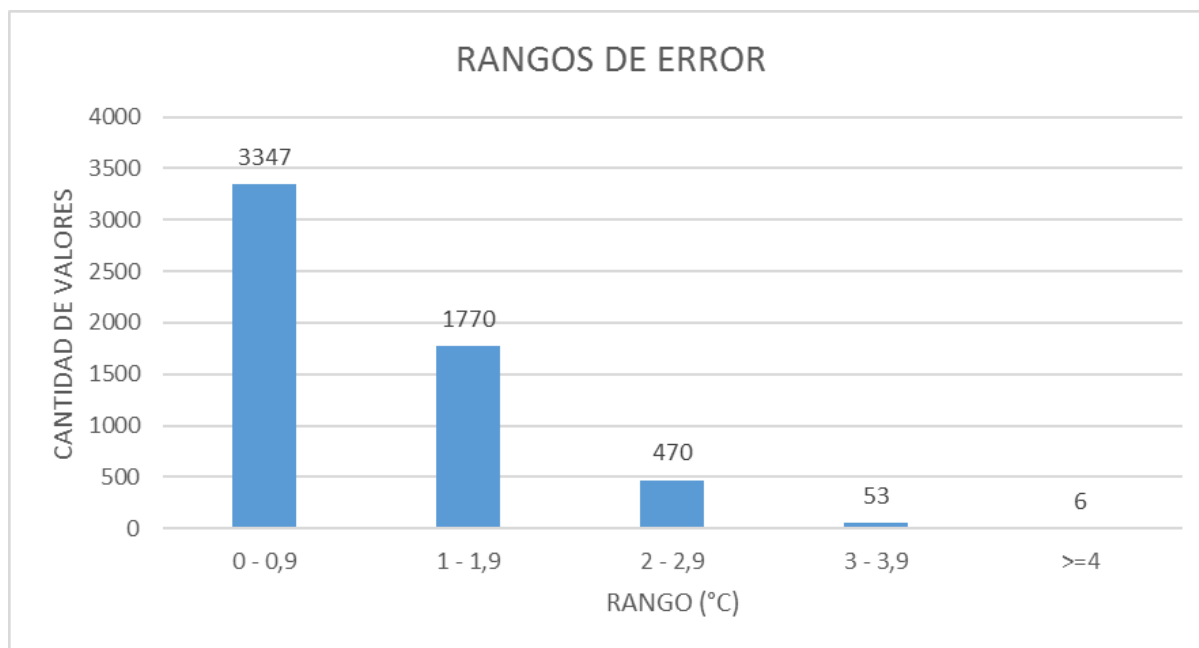


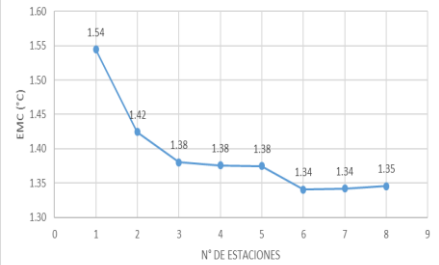
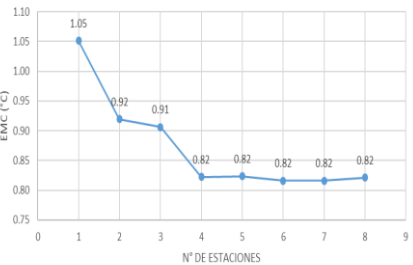
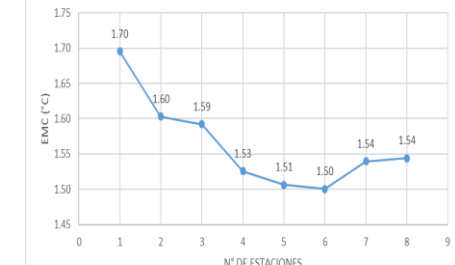
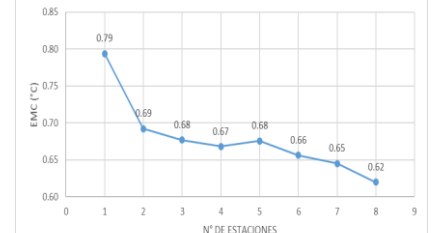
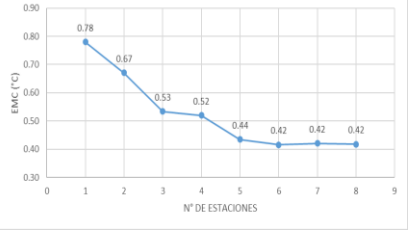
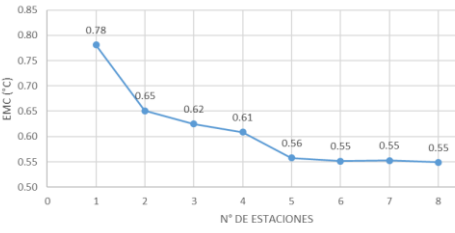
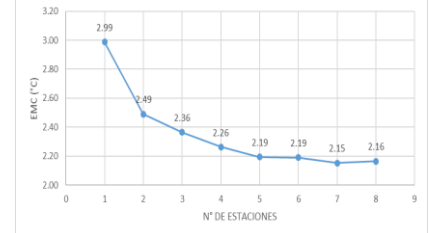
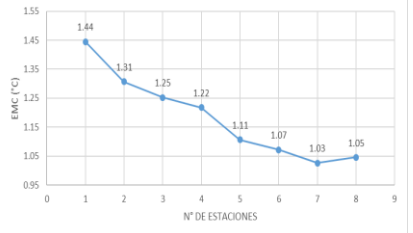
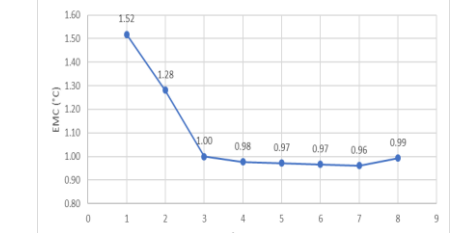
Figura 12. Rangos de Error

En la prueba de 3 estaciones con 8 variables se crea grafica EMC vs N° de estaciones (Ver Tabla 21).

Evaluada la Tabla 21, expone una tendencia general decreciente en el eje de las Y, y creciente en el eje de las X, lo que denota que a mayor cantidad de estaciones el EMC disminuye. Además, se demuestra que en las primeras tres variables está la mayor pendiente negativa del error, y en cambio indican que a partir de una cuarta variable la variación en el EMC no es muy notable.

Por generarse cambios considerables hasta una cuarta estación, se decide trabajar el llenado de datos con este límite de 4 variables.

Tabla 21. Gráfica Error Medio Cuadrático (EMC)

EST. 3	EST. 14	EST. 34																																																						
Valores Máximos																																																								
<p>EMC EST. 3 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>1.54</td></tr> <tr><td>2</td><td>1.42</td></tr> <tr><td>3</td><td>1.38</td></tr> <tr><td>4</td><td>1.38</td></tr> <tr><td>5</td><td>1.38</td></tr> <tr><td>6</td><td>1.34</td></tr> <tr><td>7</td><td>1.34</td></tr> <tr><td>8</td><td>1.35</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	1.54	2	1.42	3	1.38	4	1.38	5	1.38	6	1.34	7	1.34	8	1.35	<p>EMC EST. 14 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>1.05</td></tr> <tr><td>2</td><td>0.92</td></tr> <tr><td>3</td><td>0.91</td></tr> <tr><td>4</td><td>0.82</td></tr> <tr><td>5</td><td>0.82</td></tr> <tr><td>6</td><td>0.82</td></tr> <tr><td>7</td><td>0.82</td></tr> <tr><td>8</td><td>0.82</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	1.05	2	0.92	3	0.91	4	0.82	5	0.82	6	0.82	7	0.82	8	0.82	<p>EMC EST. 34 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>1.70</td></tr> <tr><td>2</td><td>1.60</td></tr> <tr><td>3</td><td>1.59</td></tr> <tr><td>4</td><td>1.53</td></tr> <tr><td>5</td><td>1.51</td></tr> <tr><td>6</td><td>1.50</td></tr> <tr><td>7</td><td>1.54</td></tr> <tr><td>8</td><td>1.54</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	1.70	2	1.60	3	1.59	4	1.53	5	1.51	6	1.50	7	1.54	8	1.54
Nº DE ESTACIONES	EMC (°C)																																																							
1	1.54																																																							
2	1.42																																																							
3	1.38																																																							
4	1.38																																																							
5	1.38																																																							
6	1.34																																																							
7	1.34																																																							
8	1.35																																																							
Nº DE ESTACIONES	EMC (°C)																																																							
1	1.05																																																							
2	0.92																																																							
3	0.91																																																							
4	0.82																																																							
5	0.82																																																							
6	0.82																																																							
7	0.82																																																							
8	0.82																																																							
Nº DE ESTACIONES	EMC (°C)																																																							
1	1.70																																																							
2	1.60																																																							
3	1.59																																																							
4	1.53																																																							
5	1.51																																																							
6	1.50																																																							
7	1.54																																																							
8	1.54																																																							
Valores Medios																																																								
<p>EMC EST. 3 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.79</td></tr> <tr><td>2</td><td>0.69</td></tr> <tr><td>3</td><td>0.68</td></tr> <tr><td>4</td><td>0.67</td></tr> <tr><td>5</td><td>0.68</td></tr> <tr><td>6</td><td>0.66</td></tr> <tr><td>7</td><td>0.65</td></tr> <tr><td>8</td><td>0.62</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	0.79	2	0.69	3	0.68	4	0.67	5	0.68	6	0.66	7	0.65	8	0.62	<p>EMC EST. 14 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.78</td></tr> <tr><td>2</td><td>0.67</td></tr> <tr><td>3</td><td>0.53</td></tr> <tr><td>4</td><td>0.52</td></tr> <tr><td>5</td><td>0.44</td></tr> <tr><td>6</td><td>0.42</td></tr> <tr><td>7</td><td>0.42</td></tr> <tr><td>8</td><td>0.42</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	0.78	2	0.67	3	0.53	4	0.52	5	0.44	6	0.42	7	0.42	8	0.42	<p>EMC EST. 34 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.78</td></tr> <tr><td>2</td><td>0.65</td></tr> <tr><td>3</td><td>0.62</td></tr> <tr><td>4</td><td>0.61</td></tr> <tr><td>5</td><td>0.56</td></tr> <tr><td>6</td><td>0.55</td></tr> <tr><td>7</td><td>0.55</td></tr> <tr><td>8</td><td>0.55</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	0.78	2	0.65	3	0.62	4	0.61	5	0.56	6	0.55	7	0.55	8	0.55
Nº DE ESTACIONES	EMC (°C)																																																							
1	0.79																																																							
2	0.69																																																							
3	0.68																																																							
4	0.67																																																							
5	0.68																																																							
6	0.66																																																							
7	0.65																																																							
8	0.62																																																							
Nº DE ESTACIONES	EMC (°C)																																																							
1	0.78																																																							
2	0.67																																																							
3	0.53																																																							
4	0.52																																																							
5	0.44																																																							
6	0.42																																																							
7	0.42																																																							
8	0.42																																																							
Nº DE ESTACIONES	EMC (°C)																																																							
1	0.78																																																							
2	0.65																																																							
3	0.62																																																							
4	0.61																																																							
5	0.56																																																							
6	0.55																																																							
7	0.55																																																							
8	0.55																																																							
Valores Mínimos																																																								
<p>EMC EST. 3 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>2.99</td></tr> <tr><td>2</td><td>2.49</td></tr> <tr><td>3</td><td>2.36</td></tr> <tr><td>4</td><td>2.26</td></tr> <tr><td>5</td><td>2.19</td></tr> <tr><td>6</td><td>2.19</td></tr> <tr><td>7</td><td>2.15</td></tr> <tr><td>8</td><td>2.16</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	2.99	2	2.49	3	2.36	4	2.26	5	2.19	6	2.19	7	2.15	8	2.16	<p>EMC EST. 14 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>1.44</td></tr> <tr><td>2</td><td>1.31</td></tr> <tr><td>3</td><td>1.25</td></tr> <tr><td>4</td><td>1.22</td></tr> <tr><td>5</td><td>1.11</td></tr> <tr><td>6</td><td>1.07</td></tr> <tr><td>7</td><td>1.03</td></tr> <tr><td>8</td><td>1.05</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	1.44	2	1.31	3	1.25	4	1.22	5	1.11	6	1.07	7	1.03	8	1.05	<p>EMC EST. 34 vs. CANTIDAD DE ESTACIONES</p>  <table border="1"> <thead> <tr> <th>Nº DE ESTACIONES</th> <th>EMC (°C)</th> </tr> </thead> <tbody> <tr><td>1</td><td>1.52</td></tr> <tr><td>2</td><td>1.28</td></tr> <tr><td>3</td><td>1.00</td></tr> <tr><td>4</td><td>0.98</td></tr> <tr><td>5</td><td>0.97</td></tr> <tr><td>6</td><td>0.97</td></tr> <tr><td>7</td><td>0.96</td></tr> <tr><td>8</td><td>0.99</td></tr> </tbody> </table>	Nº DE ESTACIONES	EMC (°C)	1	1.52	2	1.28	3	1.00	4	0.98	5	0.97	6	0.97	7	0.96	8	0.99
Nº DE ESTACIONES	EMC (°C)																																																							
1	2.99																																																							
2	2.49																																																							
3	2.36																																																							
4	2.26																																																							
5	2.19																																																							
6	2.19																																																							
7	2.15																																																							
8	2.16																																																							
Nº DE ESTACIONES	EMC (°C)																																																							
1	1.44																																																							
2	1.31																																																							
3	1.25																																																							
4	1.22																																																							
5	1.11																																																							
6	1.07																																																							
7	1.03																																																							
8	1.05																																																							
Nº DE ESTACIONES	EMC (°C)																																																							
1	1.52																																																							
2	1.28																																																							
3	1.00																																																							
4	0.98																																																							
5	0.97																																																							
6	0.97																																																							
7	0.96																																																							
8	0.99																																																							

Partiendo del límite de variables establecido, se genera regresión en las estaciones seleccionadas para el llenado de datos.

Se muestran las regresiones de los valores máximos de la estación APTO CAMILO DAZA y se efectúa desde 1 hasta 4 variables (Ver tabla 22, tabla 23, tabla 24, tabla 25), que se emplean para la elaboración de las ecuaciones respectivas para el llenado de datos.

Tabla 22. Regresión 1 Variable EST. 2 - Valores Máximos

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.918767235
Coefficiente de determinación R ²	0.844133231
R ² ajustado	0.844119183
Error típico	1.016487461
Observaciones	11097

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	1	62085.30572	62085.30572	60087.58819	0
Residuos	11095	11463.87278	1.033246758		
Total	11096	73549.1785			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	2.641417709	0.120513186	21.91808053	2.797E-104	2.405190435	2.877644983	2.405190435	2.877644983
Variable X 1	0.910766576	0.003715478	245.1276977	0	0.903483579	0.918049574	0.903483579	0.918049574

Tabla 23. Regresión 2 Variables EST. 2 - Valores Máximos

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.930822007
Coefficiente de determinación R ²	0.866429609
R ² ajustado	0.866402722
Error típico	0.933609837
Observaciones	9939

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	2	56177.90088	28088.95044	32225.8717	0
Residuos	9936	8660.48913	0.871627328		
Total	9938	64838.39001			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	1.919294386	0.119324113	16.08471528	1.73175E-57	1.68539493	2.153193843	1.68539493	2.153193843
Variable X 1	0.700387361	0.006253549	111.9983865	0	0.688129137	0.712645584	0.688129137	0.712645584
Variable X 2	0.24321422	0.005860504	41.50056221	0	0.231726444	0.254701997	0.231726444	0.254701997

Tabla 24. Regresión 3 Variables EST. 2 - Valores Máximos

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.931657686
Coefficiente de determinación R ²	0.867986044
R ² ajustado	0.867944473
Error típico	0.927377932
Observaciones	9531

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	3	53871.93504	17957.31168	20879.87076	0
Residuos	9527	8193.504182	0.860029829		
Total	9530	62065.43922			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	1.835596693	0.121264443	15.13713875	3.61817E-51	1.597892554	2.073300833	1.597892554	2.073300833
Variable X 1	0.650955167	0.00742181	87.70840823	0	0.636406838	0.665503496	0.636406838	0.665503496
Variable X 2	0.206447187	0.006523186	31.64821592	4.1323E-209	0.193660353	0.21923402	0.193660353	0.21923402
Variable X 3	0.105734988	0.008148073	12.97668661	3.49185E-38	0.089763029	0.121706946	0.089763029	0.121706946

Tabla 25. Regresión 4 Variables EST. 2 - Valores Máximos

Resumen

<i>Estadísticas de la regresión</i>	
Coefficiente de correlación múltiple	0.932349154
Coefficiente de determinación R ²	0.869274946
R ² ajustado	0.869207743
Error típico	0.913881071
Observaciones	7786

ANÁLISIS DE VARIANZA

	<i>Grados de libertad</i>	<i>Suma de cuadrados</i>	<i>Promedio de los cuadrados</i>	<i>F</i>	<i>Valor crítico de F</i>
Regresión	4	43212.86988	10803.21747	12935.21808	0
Residuos	7781	6498.524774	0.835178611		
Total	7785	49711.39465			

	<i>Coefficientes</i>	<i>Error típico</i>	<i>Estadístico t</i>	<i>Probabilidad</i>	<i>Inferior 95%</i>	<i>Superior 95%</i>	<i>Inferior 95,0%</i>	<i>Superior 95,0%</i>
Intercepción	1.351069575	0.140387293	9.623873704	8.38275E-22	1.07587273	1.626266421	1.07587273	1.626266421
Variable X 1	0.618151379	0.008207805	75.31263257	0	0.602061874	0.634240883	0.602061874	0.634240883
Variable X 2	0.177601211	0.007343221	24.18573685	1.1509E-124	0.163206523	0.191995899	0.163206523	0.191995899
Variable X 3	0.095792373	0.008923974	10.73427346	1.07993E-26	0.078298985	0.113285761	0.078298985	0.113285761
Variable X 4	0.106621664	0.007520061	14.17829777	4.5313E-45	0.091880322	0.121363005	0.091880322	0.121363005

Realizadas las regresiones en todas las estaciones seleccionadas, se desarrolla el llenado de datos (Anexo 18) que comprime su información en tabla denominada Comparación EST. Real vs EST. Simulada, y especifica el total de datos llenados en la estación y el porcentaje al que corresponde este llenado respecto de los vacíos iniciales (Ver tabla).

La prioridad al efectuar el respectivo llenado de datos a cada estación es aplicar la ecuación generada a partir de cuatro variables, ya que ésta genera menor error, en el caso de no ejecutarse dicha ecuación, se le da prioridad a la ecuación generada con tres variables y así sucesivamente, hasta llegar a la ecuación generada con una variable. Si ejecutado este llenado inicial, se encuentran vacíos únicos entre dos valores, entonces, éste se promedia entre su dato anterior y posterior, obteniendo así, la estación simulada.

Tabla 26. EST. 2 Real vs EST. 2 Simulada Valores Máximos

AÑO	MES	DIA	DATO N°	EST. 2	EST. 2 SIM
1976	1	1	1	28.0	28.0
1976	1	2	2	27.7	27.7
1976	1	3	3	27.3	27.3
1976	1	4	4	28.4	28.4
1976	1	5	5	27.2	27.2
1976	1	6	6	23.3	23.3
1976	1	7	7	29.2	29.2
1976	1	8	8	28.6	28.6
1976	1	9	9	27.8	27.8
1976	1	10	10	26.7	26.7
1976	1	11	11	27.4	27.4
1976	1	12	12	28.7	28.7
1976	1	13	13	28.6	28.6
1976	1	14	14	27.1	27.1
1976	1	15	15	26.2	26.2
1976	1	16	16	30.0	30.0
1976	1	17	17	28.8	28.8
1976	1	18	18	29.8	29.8
1976	1	19	19	27.0	27.0
1976	1	20	20	28.4	28.4
1976	1	21	21	27.0	27.0
1976	1	22	22	28.8	28.8
1976	1	23	23	28.8	28.8
1976	1	24	24	29.3	29.3
1976	1	25	25	30.0	30.0
1976	1	26	26	29.0	29.0
1976	1	27	27	24.6	24.6
1976	1	28	28	26.3	26.3
1976	1	29	29	26.4	26.4
1976	1	30	30	26.8	26.8
1976	1	31	31	25.2	25.2

COMPARACION EST. REAL VS. EST. SIMULADA				
	EST. 2		EST. 2 SIMULADA	
VALORES	13939	88.7%	15026	95.7%
VACIOS	1767	11.3%	680	4.3%
TOTAL	15706	100%	15706	100%
DATOS OBTENIDOS				
TOTAL	1087		6.9%	

En la tabla correspondiente se extrae la tabla de comparación, donde evidencia que la EST. 2 inicialmente cuenta con 13939 datos (88.7%), llenando 1087 vacíos (6.9% del total de valores) de los 1767 vacíos iniciales, quedando con 15026 datos (95.7%) y 680 vacíos (4.3%) en la EST. 2 SIMULADA.

Tabla 27. Comparación EST.2 Real vs EST.2 Simulada Valores Máximos

COMPARACION EST. REAL VS. EST. SIMULADA				
	EST. 2		EST. 2 SIMULADA	
VALORES	13939	88.7%	15026	95.7%
VACIOS	1767	11.3%	680	4.3%
TOTAL	15706	100%	15706	100%
DATOS OBTENIDOS				
TOTAL	1087		6.9%	

4.6 Presentación de la Información Recolectada Usando una Herramienta Tecnológica que Permita Reconstruir las Series de Temperatura

La información recolectada se encuentra adjunta en el presente informe y en anexos descritos en el contenido del mismo.

Adicionalmente se realiza el análisis de las estaciones seleccionadas con el llenado de datos generado, donde se transforma ésta información a formato propio, (Ver Anexo 19). Como actividad complementaria se convierte a formato columna (Ver Anexo 20) y formato columna (bloc de notas) (Ver Anexo 21), proveyendo de esta manera, información simulada para posible utilización.

Se relaciona, Anexo 22, mapa de localización de estaciones mediante Google Earth, que muestra la ubicación, nombre y código de las estaciones trabajadas en la ejecución del proyecto (Ver tabla).

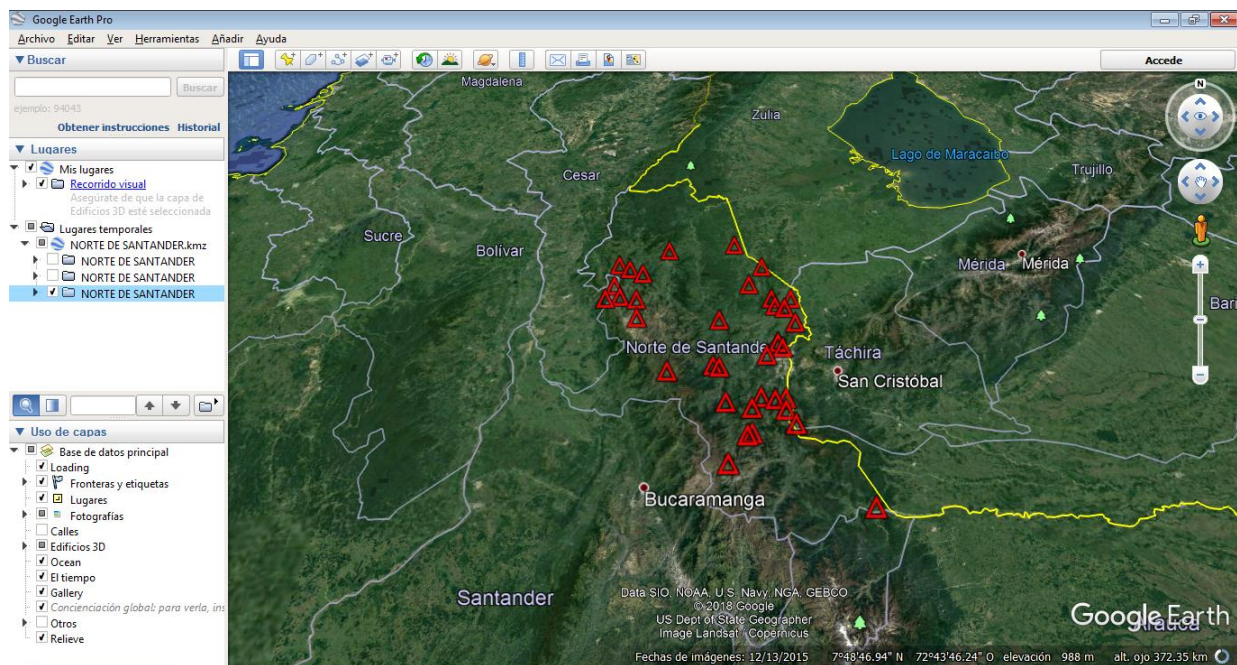


Figura 13. Localización de Estaciones mediante Google Earth

El Anexo 23, comprende la información característica de las estaciones que ofrece el Catálogo Nacional de Estaciones del IDEAM y otras entidades. El anexo consiste en un documento en formato Excel que contiene tres hojas de cálculo que se han mostrado previamente en el video del anexo 2. En hoja 1, denominada ESTACIONES NORTE DE SANTANDER, se muestra la totalidad de las estaciones del Departamento, la hoja 2, denominada PRESELECCIÓN, contiene el listado de las estaciones seleccionadas previamente de acuerdo a las fechas de instalación del Catálogo, la hoja 3, denominada RECIBIDO, con las treinta y seis estaciones recibidas para la realización del proyecto.

El Anexo 24, corresponde a los archivos originales suministrados por el IDEAM. Los anteriores se adjuntan con el fin de ofrecer ésta información a quien le sea de utilidad la documentación.

Lo obtenido mediante este proyecto, lo conservará el Grupo de Investigación HYDROS – UFPS liderado por el Ingeniero Gustavo Adolfo Carrillo Soto.

5. Conclusiones

Al finalizar el desarrollo del proyecto se cumplieron los objetivos propuestos, al igual que se identificaron limitaciones propias del método implementado.

Gracias a la ejecución de este proyecto, la Institución contará con información de la Variable Temperatura, adjunta en los anexos, que forman parte integral de este trabajo y comprende la compilación de los datos del Ideam reorganizados, conjunto con el llenado de datos faltantes, siguiendo la metodología de la regresión lineal multivariada en los casos donde se pudo aplicar.

La recolección de la información se torna de difícil acceso, utilizar las diferentes plataformas es relativamente dispendioso, además el recibir diferente información en diferentes pedidos genera complicación en el análisis de la misma.

Al recopilar la información, originalmente enviada por el IDEAM, esta se reorganizó en formatos de Hoja de Cálculo y formatos de Texto plano, y se encuentra a disposición del público a través del Grupo de Investigación HYDROS-UFPS.

Las estaciones que presentan datos dudosos son apto camilo daza (est. 2), iser pamplona (est. 3), santa isabel (est. 5) blonay (est. 6), tama parque nal (est. 7), ragonvalia (est.8), univ franciso de (est. 9), univ de pamplona (est. 10), alcaldia de herran (est. 11), aeropuerto camilo daza – aut (est. 13), risaralda (est. 15), salazar (est. 16), ceez ica (est. 19), hjc granja (est. 20), tibu (est. 21), petrolea (est. 22), sardinata (est. 23), la estancia (est. 24), teorama (est. 27), abrego centro admo (est. 28), maravilla la (est. 30), universidad francisco de paula santander (est. 32), aguas de la virgen (est. 33), escuela agricola cachira (est. 34) y silos (est. 35).

Los valores de temperaturas medias presentan muy baja información, mientras que los máximos y mínimos presentan una data considerable.

Las estaciones con calidad de información denominada muy buena son, para valores de Temperatura Máxima, Iser Pamplona (est. 3), escuela Agrícola Cachira (est. 34) y aeropuerto camilo daza (EST. 2) con porcentajes respectivos de 83.0%, 83.5% y 88.7%, lo que aduce que Aeropuerto Camilo Daza tiene la mayor información en este parámetro. En valores de Temperatura Media se tiene 6 estaciones y son, Abrego centro admo (est. 28), Tunebia (est. 36), escuela Agrícola Cachira (est.34), Iser Pamplona (est. 3), Carmen de Tonchala (est. 14) y aeropuerto camilo daza (EST. 2) con porcentaje de valores de 80.3%, 81%, 82.3%, 87.7%, 87.8% y 90.8%, siendo en temperaturas medias, también, Aeropuerto Camilo Daza la que cuenta con mayor cantidad de información. Finalmente, en valores de Temperatura Mínima se encuentran 3 estaciones que son, Escuela Agrícola Cachira (EST.34), Carmen de Tonchala (EST.14), Iser Pamplona (EST.3), con porcentaje de valores correspondientes de 80.1%, 86.8%, 88.1%, donde Iser Pamplona proporciona la mayor cantidad de datos.

Se genera disminución significativa del error medio cuadrático (EMC) hasta una cuarta estación, por lo cual se trabaja el llenado de datos con este límite de cuatro variables. Lo anterior aduce que la regresión lineal multivariada tiene limitación, ya que depende de los valores de otras cuatro estaciones (variables).

La regresión lineal multivariada fue aplicada en las estaciones con data por encima del 40%, ya que por debajo de este porcentaje la confiabilidad del método se vería afectada, puesto que una gran cantidad de vacíos dispersaría la realidad de los datos simulados.

El llenado de datos se restringe si el valor de la primera estación (variable) no tiene dato, por tanto, no se genera información.

El método de regresión lineal multivariada se ajusta a la variable temperatura, ya que ésta es constante y muestra un rango de variabilidad relativamente lineal.

6. Recomendaciones

Se recomienda continuar la reconstrucción de series de tiempo de las variables principales de modelación hidrológica.

Recomendamos que en futuras investigaciones se utilicen otros métodos alternos, para verificar si hay una disminución del error o si hay un mejor comportamiento en el llenado de datos faltantes, métodos tales como Distancia Inversa Ponderada (Inverse Distance Weighting - IDW) o Redes Neuronales Artificiales (RNA), que podría ser trabajado por el Grupo de Investigación HYDROS-UFPS con el apoyo del grupo de investigación de Ingeniería de Sistemas.

Se recomienda hacer un análisis exhaustivo de datos dudosos, pues estos son difíciles de detectar, la visualización de graficas de dispersión y box plot, muestran fácilmente comportamientos de temperatura anormales, además se deben conocer características de la estación como su elevación y localización para una posible comparación con otras similares.

Es viable continuar con la actualización de las series de temperatura, utilizando la información comprendida en este proyecto y anexando la que se obtendrá en adelante a la presentación del mismo.

Se recomienda el mantenimiento de las estaciones en Norte de Santander, ya que contar con información de 36 estaciones entre 294 instaladas en el departamento deja entrever una falencia en esta situación.

Es recomendable la utilización de otros métodos para el llenado de datos de la variable temperatura, para hacer una comparación entre estos y tener claridad sobre cuál es el método más exacto y viable a utilizar.

Referencias Bibliográficas

- Área Metropolitana de Cúcuta. (2019). San José de Cúcuta. Recuperado de: https://amc.gov.co/amc/index.php/san-jose-de-cucuta/?fbclid=iwar2hpblbndtf1imdnb5adphd0v4mc6bcuaubjmcyoym33b4hancled_r_f4
- Arias, F. (2006). El Proyecto de Investigación. Caracas: Episteme.
- Blog Fibras y Normas de Colombia S.A.S. (2019). Blog Fibras y Normas de Colombia. Recuperado de: <https://www.fibrasynormasdecolombia.com/terminos-definiciones/caudal-definicion-y-metodos-de-medicion/>
- Castro, I. (2009). Metodología de validación de productos Modis para la estimación de temperatura de la superficie en zonas heterogéneas y homogéneas de Colombia. Universidad Nacional de Colombia. Bogotá, Colombia.
- Chow, V., Maidment, D. & Mays, L. (2010). Hidrología Aplicada. Bogotá: Nomos S.A.
- Concepto Definición De. (2019). Concepto de Climatología. Recuperado de: <https://conceptodefinicion.de/climatologia>
- Definición.De. (2019). Definición de temperatura. Recuperado de: <https://definicion.de/temperatura/>
- Economipedia. (2019). Diccionario económico. Recuperado de: <https://economipedia.com/definiciones>
- Ecured. (2011). Medidas de dispersión. Recuperado de: https://www.ecured.cu/Medidas_de_dispersi%C3%B3n

Geoenciclopedia. (2019). Precipitación. Recuperado de:
<https://www.geoenciclopedia.com/precipitacion/?fbclid=iwar0tygysmsbsp3evhzt5kx8n5lhp0qxbeg3ob6ht1adq6icvgrk5gdp6i>

Gobernación de Norte de Santander. (2019). Información General. Recuperado de:
<http://www.nortedesantander.gov.co/gobernaci%3%b3n/nuestro-departamento/informaci%3%b3n-general-norte-de-santander>

Grupo Camaleon. (2017). Glosario Términos estadísticos / Término. Recuperado de:
<https://glosarios.servidor-alicante.com/terminos-estadistica/coeficiente-de-asimetria-de-fisher>

Grupo Camaleon. (2018). Glosario Topografía, Geodesia y GPS / Término. Recuperado de:
<https://glosarios.servidor-alicante.com> Obtenido de <https://glosarios.servidor-alicante.com/topografia-geodesia-gps/error-medio-cuadratico-emc>

Guardiola, P. (2019). Docencia. Recuperado de:
<https://www.um.es/docencia/pguardio/documentos/Tec2.pdf>

HidroChep. (2019). Cuenca Río Pamplonita. Recuperado de:
<https://sites.google.com/site/hydrotoolbox/casos-de-estudio/cuenca-rio-pamplonita>

Instituto de Hidrología, Meteorología y Estudios Ambientales. (2019). Modelación Hidrológica. Recuperado de:
http://www.ideam.gov.co/web/agua/modelacion-hidrologica?fbclid=IwAR1BU71uUaCv7tI0m7hRhlo2MlbbnZXdKs_zAspqWSNWs3l5D8-gjomlfH4

Instituto de Hidrología, Meteorología y Estudios Ambientales. (2019). Tiempo y clima. Recuperado de http://www.ideam.gov.co/web/tiempo-y-clima/seguimiento-tiempo?fbclid=iwar2w8xqzbnctfbnadijep_8ns69byn6hqhdzal4d5iw1h3yrjbyoc913vjo

Medina, R. (2008). Estimación Estadística de Valores faltantes en series históricas de lluvias. Universidad Tecnológica de Pereira. Pereira, Colombia.

Mijares, F. (1992). Fundamentos de Hidrología de Superficie. México: Limusa.

Networkianos. (2019). Qué es la Regresión Lineal Múltiple y cómo analizarla. Recuperado de: <http://networkianos.com/regresion-lineal-multiple/>

Quevedo, F. (2011). Medidas de tendencia central y dispersión. Medwave, 3(11), p.6-10.

SPSSFree. (2019). Medidas de distribución, asimetría y curtosis. Recuperado de: <http://www.spssfree.com/curso-de-spss/analisis-descriptivo/medidas-de-distribucion-curtosis-asimetria.html>

Sangaku S.L. (2019). Teoría y ejercicios de matemáticas para secundaria, bachillerato y primeros cursos de carreras técnicas. Recuperado de: <https://www.sangakoo.com/>

Tutiempo.net. (2019). La Climatología. Recuperado de: https://blog.tutiempo.net/la-climatologia/?fbclid=IwAR0UGR4ehue6S8svCqONVMYVdzBRAG8WFsb42hjwt4Teey_1i5NzU_2v5aTk

Universidad Francisco de Paula Santander. (1996). Estatuto Estudiantil. U.F.P.S. Cúcuta, Colombia.

Universo Formulas. (2019). Asimetría y Curtosis. Recuperado de:
<https://www.universoformulas.com/estadistica/descriptiva/asimetria-curtosis/>