

# Entendimiento de fenómenos ambientales mediante análisis de datos

**Jorge Julián Vélez Upegui**  
**Mauricio Orozco Alzate**  
**Néstor Darío Duque Méndez**  
**Beatriz Helena Aristizábal Zuluaga**



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

SEDE MANIZALES  
FACULTAD DE ADMINISTRACIÓN



ENTENDIMIENTO DE  
FENÓMENOS AMBIENTALES  
MEDIANTE ANÁLISIS DE DATOS



# ENTENDIMIENTO DE FENÓMENOS AMBIENTALES MEDIANTE ANÁLISIS DE DATOS

Editores

Jorge Julián Vélez Upegui

Mauricio Orozco Alzate

Néstor Darío Duque Méndez

Beatriz Helena Aristizábal Zuluaga



UNIVERSIDAD NACIONAL DE COLOMBIA

SEDE MANIZALES

FACULTAD DE ADMINISTRACIÓN

Manizales, Colombia febrero de 2015

© Universidad Nacional de Colombia  
Sede Manizales  
Facultad de Administración  
Decano Juan Manuel Castaño Molano

© AUTORES

Jorge Julián Vélez Upegui  
Mauricio Orozco Alzate  
Néstor Darío Duque Méndez  
Beatriz Helena Aristizábal Zuluaga

Primera edición, 2015  
ISBN: 978-958-775-282-3

Revisión  
Gerard Olivar Tost  
Alfonso Pío Agudelo Salazar

Impresión  
Sección de Publicaciones e Imagen. Universidad Nacional de Colombia. Sede Manizales

Manizales, Colombia, 2015

Prohibida la reproducción total o parcial de esta obra sin autorización escrita del titular de los derechos patrimoniales.

---

Catalogación en la publicación Universidad Nacional de Colombia Sede Manizales

Vélez Upegui, Jorge Julián  
Entendimiento de fenómenos ambientales mediante el análisis de datos / Jorge Julián Vélez Upegui,  
Néstor Darío Duque Méndez, Mauricio Orozco Alzate, Beatriz Helena Aristizábal Zuluaga – Manizales:  
Universidad Nacional de Colombia.

132 páginas  
Incluye referencias bibliográficas  
ISBN: 9789587752823

1. Hidrometeorología – Procesamiento de datos – Caldas – Estadísticas 2. Hidrología – Caldas 3. Índice de  
humedad 4. Meteorología I.

551.57 Web Dewey 23



# Agradecimientos

Los editores y el equipo de investigación expresan sus agradecimientos:  
A la Vicerrectoría de Investigaciones de la Universidad Nacional de Colombia, pues este libro de investigación fue financiado parcialmente por el programa titulado: 'Fortalecimiento de capacidades en Análisis de Datos', con código 12677.

Y desde luego a los profesores y estudiantes del Grupo de Trabajo Académico en Ingeniería Hidráulica y Ambiental y el Grupo de Ambientes Inteligentes Adaptativos GAIA, por los esfuerzos orientados a aprovechar este programa para efectivamente fortalecer en forma conjunta las capacidades particulares.

Se agradece especialmente a la Corporación Autónoma Regional de Caldas CORPOCALDAS, al Instituto de Estudios Ambientales IDEA de la Universidad Nacional de Colombia Sede Manizales, a la CHEC S.A. E.S.P., al Centro Nacional de Investigaciones del Café CENICAFE, ISAGEN, GENSA y a la Alcaldía de Manizales a través de la Oficina de Prevención y Atención de Desastres, por suministrar los datos para esta investigación.



# Índice general

<b>PRÓLOGO</b>	<b>XIII</b>
<b>1. Análisis climatológico para el departamento de Caldas</b>	<b>1</b>
1.1. Introducción . . . . .	2
1.1.1. Clima y tiempo atmosférico . . . . .	2
1.1.2. Sistema climático . . . . .	3
1.1.3. Variables climáticas . . . . .	3
1.1.4. Variabilidad climática natural . . . . .	4
1.2. Clima y variabilidad climática en Colombia . . . . .	4
1.2.1. Variabilidad climática interdecadal . . . . .	5
1.2.2. Variabilidad climática interanual . . . . .	6
1.2.3. Variabilidad climática estacional . . . . .	7
1.2.4. Variabilidad climática intraestacional . . . . .	8
1.2.5. Ciclo diurno . . . . .	8
1.3. Caracterización de la zona de estudio . . . . .	9
1.4. Análisis de las variables climáticas de caldas . . . . .	11
1.4.1. Brillo solar . . . . .	12
1.4.2. Radiación solar . . . . .	14
1.4.3. Temperatura . . . . .	17
1.4.4. Humedad relativa . . . . .	24
1.4.5. Precipitación . . . . .	26
1.4.6. Velocidad y dirección del viento . . . . .	30
1.5. Clasificación climática . . . . .	35

1.6. Conclusiones . . . . .	37
<b>2. Predicción de caudales</b>	<b>45</b>
2.1. Introducción . . . . .	46
2.1.1. Hipótesis básicas en la modelación hidrológica . . . . .	47
2.1.2. Las RNA en hidrología . . . . .	47
2.1.3. Planteamiento general . . . . .	48
2.2. Caso de estudio, cuenca del río Risaralda . . . . .	49
2.3. Descripción de herramientas . . . . .	51
2.3.1. Modelo hidrológico TETIS . . . . .	51
2.3.2. Redes neuronales artificiales (RNA) . . . . .	52
2.3.3. NSGA-II . . . . .	55
2.4. Protocolo de experimentación . . . . .	56
2.4.1. Sistema de información de variables de entrada . . . . .	56
2.4.2. Modelo base de predicción . . . . .	56
2.4.2.1. Definición de los subconjuntos de entrenamiento y prueba . . . . .	57
2.4.2.2. Pre y post-procesamiento . . . . .	59
2.4.2.3. Arquitectura de la RNA . . . . .	59
2.4.3. Definición del espacio de entrada con NSGA-II . . . . .	60
2.4.3.1. Fase I: Evaluación del número de entradas . . . . .	60
2.4.3.2. Fase II: Evaluación de la selección final de entradas . . . . .	60
2.4.3.3. Elementos comunes en la configuración del AE en ambas fases . . . . .	61
2.5. Resultados y análisis . . . . .	62
2.5.1. Fase I: Evaluación del número de entradas . . . . .	62
2.5.2. Fase II: Evaluación de la selección final de entradas . . . . .	65
2.6. Conclusiones . . . . .	66
<b>3. Análisis de Datos de Calidad del Aire en la Ciudad de Manizales Utilizando Matlab y Openair</b>	<b>73</b>
3.1. Resumen . . . . .	73
3.2. Introducción . . . . .	74
3.3. Generalidades . . . . .	75

3.3.1.	El $PM_{10}$ y sus efectos en la salud . . . . .	75
3.3.2.	Meteorología y su influencia en el estudio de la calidad del aire . . . . .	77
3.3.3.	Herramientas para análisis de datos de calidad del aire . . . . .	78
3.3.4.	R y Openair. . . . .	78
3.3.5.	MATLAB . . . . .	81
3.4.	Metodología . . . . .	84
3.4.1.	Conformación de las bases de datos . . . . .	85
3.4.2.	Metodología para análisis con Openair . . . . .	86
3.4.3.	Metodología para análisis con MATLAB . . . . .	87
3.5.	Resultados y Análisis . . . . .	88
3.5.1.	Estudio de variaciones temporales de $PM_{10}$ y la influencia de variables meteorológicas utilizando Openair . . . . .	88
3.5.1.1.	Distribución de vientos para periodos diurno, nocturno y predominancia para registros de 24h (día completo). Uso de la herramienta windRose para la elaboración de la rosa de vientos. . . . .	88
3.5.1.2.	Resumen de estadísticos principales en las estaciones de monitoreo para el periodo 2010 - 2012 utilizando el gráfico de cajas y la función aqStats . . . . .	90
3.5.1.3.	Variación diaria, media mensual y semanal de $PM_{10}$ en las estaciones de monitoreo durante 2010-2012 . . . . .	92
3.5.1.4.	La rosa de contaminantes y su utilidad en el análisis de la influencia de variables meteorológicas individuales . . . . .	93
3.5.2.	Análisis de $PM_{10}$ y su correlación con las variables meteorológicas mediante la implementación de MATLAB . . . . .	97
3.5.2.1.	Técnica de visualización de datos empleando gráficas de dos y tres variables . . . . .	98
3.5.2.2.	Aplicación de regresiones lineales a los datos de $PM_{10}$ respecto a las variables meteorológicas (Técnica de regresión) . . . . .	102
3.5.2.3.	Aplicación del Análisis de Componentes Principales - PCA (Técnica de reducción de datos y proyección) . . . . .	103
3.6.	Conclusiones . . . . .	108
<b>4.</b>	<b>Análisis multidimensional de datos ambientales</b>	<b>115</b>
4.1.	Introducción . . . . .	116
4.2.	Almacenamiento y análisis de datos hidroclimatológicos . . . . .	117
4.2.1.	Diseño de la bodega de datos . . . . .	117

4.2.2.	Datos disponibles desde las estaciones hidroclimatológicas . . . . .	117
4.2.3.	Modelo multidimensional - bodegas de datos . . . . .	118
4.3.	Proceso KDD para estaciones hidroclimatológicas . . . . .	119
4.3.1.	Proceso ETL . . . . .	120
4.3.2.	Herramientas OLAP para datos hidroclimatológicos . . . . .	121
4.4.	Sistema <b>MADÉ</b> - Multidimensional Analysis of Data Environmental . . . . .	121
4.4.1.	Bodega de datos Made-DWH . . . . .	122
4.4.2.	Herramienta para OLAP - Made-OLAP . . . . .	122
4.5.	Análisis Multidimensional de datos hidroclimatológicos de Caldas . . . . .	122
4.6.	Conclusiones y trabajos futuros . . . . .	129

# Índice de figuras

1.1. Localización del departamento de Caldas. Fuente: propia . . . . .	9
1.2. Modelo de Elevación Digital del Terreno. Fuente: Adaptado (USGS) . . . . .	9
1.3. Mapa de pendientes. Fuente: Adaptado (USGS) . . . . .	10
1.4. Distritos o Provincias del departamento de Caldas. Fuente: Adaptado (IGAC) . .	10
1.5. Cuencas hidrográficas del departamento de Caldas. Fuente: propia . . . . .	11
1.6. Comportamiento estacional del brillo solar mensual 1981-2010. Fuente: propia . .	13
1.7. Mapa de brillo solar diario 1981-2010. Fuente: propia . . . . .	13
1.8. Mapa de brillo solar anual 1981-2010. Fuente: propia . . . . .	14
1.9. Variación Interdecadal del brillo solar diario- Estación Agronomía 1981-2010. Fuente: propia . . . . .	14
1.10. Ciclo diario de la radiación solar en la estación Posgrados 2002-2010. Fuente: propia . . . . .	15
1.11. Comportamiento de la radiación en la estación Posgrados 2002-2010. Fuente: propia	16
1.12. Comportamiento estacional de la radiación 1981-2010. Fuente: propia . . . . .	16
1.13. Mapa de radiación solar 1981-2010. Fuente: propia . . . . .	17
1.14. Comportamiento Interdecadal de la radiación solar Estación Cenicafé 1981-2010. Fuente: propia . . . . .	17
1.15. Ciclo diario de la Temperatura Media Estación Posgrados 2002-2010. Fuente: propia . . . . .	20
1.16. Comportamiento estacional Temperatura Media 1981-2010. Fuente: propia . . . .	20
1.17. Mapa de Temperatura Media 1981-2010. Fuente: propia . . . . .	21
1.18. Mapa de Temperatura Máxima 1981-2010. Fuente: propia . . . . .	22
1.19. Mapa de Temperatura Mínima 1981-2010. Fuente: propia . . . . .	22
1.20. Mapa de Temperatura Amplitud o Rango de Temperatura 1981-2010. Fuente: propia . . . . .	23

1.21. Comportamiento Interdecadal de la Temperatura Media Estación Agronomía. Fuente: propia . . . . .	23
1.22. Ciclo diurno de la humedad relativa estación Posgrados 2002-2010. Fuente: propia	24
1.23. Comportamiento estacional de la humedad relativa 1981-2010. Fuente: propia . .	25
1.24. Mapa humedad relativa media anual 1981-2010. Fuente: propia . . . . .	25
1.25. Comportamiento interdecadal de la humedad relativa estación CENICAFÉ. Fuen- te: propia . . . . .	26
1.26. Comportamiento del ciclo diurno de la precipitación estación Posgrados 2002- 2010. Fuente: propia . . . . .	28
1.27. Comportamiento estacional de la precipitación media 1981-2010. Fuente: propia .	29
1.28. Mapa de precipitación media anual 1981-2010. Fuente: propia . . . . .	29
1.29. Ciclo diurno de la velocidad media del viento estación Posgrados 2002-2010. Fuen- te: propia . . . . .	33
1.30. Comportamiento de estacional de la velocidad del viento. Fuente: propia . . . . .	34
1.31. Comportamiento de estacional de la velocidad del viento estación El Cisne Parque Nacional Natural Los Nevados. Fuente: propia . . . . .	34
1.32. Comportamiento de la velocidad y dirección del viento. . . . .	35
1.33. Pisos térmicos en el departamento de Caldas 1981-2010. Fuente: propia . . . . .	36
1.34. Clasificación climática de Lang departamento de Caldas 1981-2010. Fuente: propia	37
1.35. Clasificación climática de Lang-Caldas departamento de Caldas 1981-2010. Fuen- te: propia . . . . .	37
2.1. Ciclo de vida de una adopción tecnológica. . . . .	48
2.2. Zona de estudio. Fuente: propia. . . . .	49
2.3. Precipitación media anual modelada. Fuente: propia. . . . .	50
2.4. Modelo hidrológico TETIS. Fuente: [Vélez et al., 2009] . . . . .	52
2.5. Problema de generalización. Fuente: propia. . . . .	54
2.6. Frente de Pareto. Fuente: propia. . . . .	55
2.7. Observaciones y simulación hidrológica en el periodo de estudio. . . . .	58
2.8. Relación entre el número de entradas y el MSE. Fuente: propia. . . . .	63
2.9. Predicción con 50 RNA y esquema de 8 entradas. Fuente: propia. . . . .	64
2.10. Evaluación multi-objetivo de la selección de entradas. Fuente: propia. . . . .	65
2.11. Hidrogramas observados y predichos según el esquema de selección final de variables.	66

3.1. Esquema de gráfico de cajas elaborado en Openair (Tomado de [Follos, 2012]). . . . .	81
3.2. Distribución de vientos para periodo diurno (daylight) y nocturno (nighttime) en las estaciones Palogrande y Nubia durante 2010 - 2012. (a) Palogrande y (b) Nubia. . . . .	89
3.3. Distribución de vientos para periodo diario (24 h) en las estaciones de monitoreo durante 2010 - 2012. (a) Palogrande. (b) Nubia. (c) Liceo. . . . .	90
3.4. Distribución de concentraciones de $PM_{10}$ para tres estaciones de monitoreo en la ciudad de Manizales durante 2010 - 2012. . . . .	91
3.5. Distribución diaria del $PM_{10}$ en la ciudad de Manizales para las estaciones Palogrande (P), Nubia (N) y Liceo (L) durante 2010 - 2012. Las sombras representan los rangos de dispersión asociados a las líneas de tendencia calculada para cada estación. . . . .	92
3.6. Variación media mensual y concentración media por días de la semana del $PM_{10}$ en la ciudad de Manizales para las estaciones Palogrande (P), Nubia (N) y Liceo (L) durante 2010 - 2012. Las sombras representan los rangos de valores para el cálculo de cada promedio (mensual o por días de la semana). . . . .	93
3.7. Porcentaje de contribución proporcional a la media del contaminante $PM_{10}$ en función de la velocidad y dirección del viento para diferentes intervalos de precipitación en a) Liceo, b) Palogrande y c) Nubia. . . . .	95
3.8. Porcentaje de contribución proporcional a la media del contaminante $PM_{10}$ en función de la velocidad y dirección del viento para diferentes intervalos de humedad relativa en a) Liceo, b) Palogrande y c) Nubia. . . . .	96
3.9. Porcentaje de contribución proporcional a la media del contaminante $PM_{10}$ en función de la velocidad y dirección del viento para diferentes intervalos de temperatura en a) Liceo, b) Palogrande y c) Nubia. . . . .	97
3.10. Gráficas que relacionan el nivel de $PM_{10}$ con una variable meteorológica para las tres estaciones: a), c) y e) $PM_{10}$ respecto a la Precipitación Total; b), d) y f) $PM_{10}$ respecto al Porcentaje de Humedad Relativa. La escala de color hace referencia a los niveles de concentración de $PM_{10}$ : El color rojo indica mayor valor de concentración y el color azul indica menor valor de concentración. . . . .	99
3.11. Gráficas que relacionan el nivel de $PM_{10}$ con dos variables meteorológicas para las tres estaciones: a), c) y e) Relación del $PM_{10}$ respecto a la Radiación Solar y la Velocidad del Viento; b), d) y f) Relación del $PM_{10}$ respecto a la Precipitación y el Porcentaje de Humedad Relativa. Escala de color hace referencia a los niveles de concentración de $PM_{10}$ : El color rojo indica mayor valor de concentración y el color azul indica menor valor de concentración. . . . .	101
3.12. Esquema conceptual de la estabilidad atmosférica en Manizales . . . . .	102
3.13. Gráficas Scree para las tres estaciones de monitoreo . . . . .	105
3.14. Proyección de los datos en dos y tres componentes principales de las tres estaciones de monitoreo por clases. a) y b) Estación Liceo; Clase 1: Concentraciones mayores de $52 \mu\text{g}/\text{m}^3$ , Clase 2: Concentraciones entre $35$ y $52 \mu\text{g}/\text{m}^3$ , Clase 3: Concentraciones menores a $35 \mu\text{g}/\text{m}^3$ . c) y d) Estación Nubia; Clase 1: Concentraciones mayores de $39 \mu\text{g}/\text{m}^3$ , Clase 2: Concentraciones entre $25$ y $39 \mu\text{g}/\text{m}^3$ , Clase 3: Concentraciones menores a $25 \mu\text{g}/\text{m}^3$ . e) y f) Estación Palogrande; Clase 1: Concentraciones mayores de $34 \mu\text{g}/\text{m}^3$ , Clase 2: Concentraciones entre $22$ y $34 \mu\text{g}/\text{m}^3$ , Clase 3: Concentraciones menores a $22 \mu\text{g}/\text{m}^3$ . . . . .	106

4.1. Diseño de la Bodega de Datos - Modelo Estrella. Fuente: propia. . . . .	119
4.2. Fases de KDD con un modelo multidimensional de bodega de datos. [Zubcoff (2009)]	120
4.3. Sistema <b>MADE</b> - Multidimensional Analysis of Data Environmental. Fuente: propia. . . . .	121
4.4. Interfaz de consulta en la herramienta MADE-OLAP. Fuente: propia. . . . .	122
4.5. Relación entre flujo y nivel medio en tres estaciones. Fuente: propia. . . . .	123
4.6. Relación entre Flujo y Nivel medio para 15 lustros. Fuente: propia. . . . .	124
4.7. Radiación solar - Presión barométrica estación posgrados. Fecha: 01/01/2010 Intervalo: 5 Minutos. Fuente: propia. . . . .	124
4.8. Lluvia acumulada en año/mes para diferentes estaciones. Fuente: propia. . . . .	125
4.9. Promedio de lluvias en el periodo de los fenómenos de <i>El niño</i> y <i>La niña</i> . Fuente: propia. . . . .	125
4.10. Lluvias acumuladas para el periodo 1956-1961. Fuente: propia. . . . .	126
4.11. Diferentes escalas y variables para análisis. Fuente: propia. . . . .	126
4.12. Tendencia para la estación Cenicafé. Fuente: propia. . . . .	127
4.13. Climograma para la estación Java 1999-2010. Fuente: propia. . . . .	127
4.14. Datos faltantes en la medida brillo para la estación Agronomía en un período de tiempo. Fuente: propia. . . . .	128
4.15. Climograma de la estación Santagueda. Fuente: propia. . . . .	128
4.16. Medidas almacenadas de la estación La Argentina - Ausencia de datos. Fuente: propia. . . . .	129

# Índice de tablas

1.1. Zonas de Vida del departamento de Caldas. Fuente: [CORPOCALDAS, 2007] . . .	12
1.2. Localización geográfica de estaciones climáticas con registros de brillo solar. Fuente: propia . . . . .	12
1.3. Localización geográfica de estaciones climáticas con registros de radiación solar. Fuente: propia . . . . .	15
1.4. Localización geográfica de estaciones climáticas con registros de temperatura. Fuente: propia . . . . .	19
1.5. Significancia estadística análisis de varianza precipitación interdecadal. Fuente: propia . . . . .	30
1.6. Escala de Beaufort – Velocidad del viento. Fuente: [OMM, 2008] . . . . .	31
1.7. Localización geográfica de estaciones con registros de humedad relativa. Fuente: propia . . . . .	33
1.8. Clasificación Climática- Caldas. Adaptada: [IDEAM, 2005a] . . . . .	35
1.9. Clasificación de Lang. Adaptada: [IDEAM, 2005a] . . . . .	36
1.10. Clasificación de Caldas- Lang. Adaptada: [IDEAM, 2005a] . . . . .	36
2.1. Tabla de 71 entradas candidatas para la selección de variables. . . . .	57
2.2. Estadísticos subconjuntos de entrenamiento y prueba. . . . .	58
2.3. Configuración del NSGA-II para la selección de variables. Fuente: propia. . . . .	62
2.4. Subconjuntos de variables para el modelo de predicción con un conjunto de RNA. La denotación de las variables sigue la identificación dada en la Tabla 2.1. Adicionalmente las variables aparecen tabuladas para visualizar su participación en distintos esquemas de selección. . . . .	63
2.5. Esquema final de la selección de variables basada en el compromiso de tres funciones objetivo. . . . .	66
3.1. Límites máximos permisibles para niveles de inmisión de $PM_{10}$ en Colombia. . .	76
3.2. Algunos estudios de calidad del aire con manejo de datos mediante la aplicación de Openair. . . . .	79

3.3. Características principales de las estaciones de monitoreo analizadas. . . . .	84
3.4. Características principales de las estaciones de monitoreo analizadas. . . . .	85
3.5. Comandos de Openair utilizados en el procedimiento analítico. . . . .	86
3.6. Comandos de MATLAB utilizados en el procedimiento analítico. . . . .	87
3.7. Resumen de estadísticos principales para $PM_{10}$ en cada estación de monitoreo aplicando la función <code>aqStats</code> de Openair. . . . .	92
3.8. Promedios mensuales de $PM_{10}$ para las tres estaciones desde Noviembre de 2011 hasta Junio de 2012 . . . . .	100
3.9. Parámetros estadísticos para las regresiones lineales realizadas entre las concentraciones de $PM_{10}$ y cada variable meteorológica de las tres estaciones de monitoreo	103
3.10. Componentes Principales evaluados para los datos de las tres estaciones . . . . .	105
3.11. Cargas de las variables meteorológicas en cada Componente Principal (Matriz de valores propios). . . . .	107
4.1. Características de algunas Estaciones de la Red. Fuente: propia. . . . .	118

# PRÓLOGO

*... Los datos en bruto sólo son la recopilación de hechos, el procesamiento y análisis permiten obtener conocimiento y dan valor a lo almacenado. ...*

*Angie Vidal*

En primer lugar hay que precisar que los conceptos y propuestas incluidas en este libro son el resultado de trabajos de investigación de los grupos vinculados y de las personas (profesores, estudiantes de doctorado, de maestría y de pregrado) que participaron en el proceso. Por lo tanto, además de conceptos generales sobre los asuntos tratados, también hay una revisión del estado del arte y sobre todo es un libro propositivo, con una visión sobre la que se fundamentan los desarrollos realizados.

En la medida en que se dispone de información mejor y más confiable es posible mejorar nuestro conocimiento y entendimiento sobre el entorno, se hace factible realizar mejores modelos que nos permitan comprender mejor los fenómenos naturales que nos aquejan y podemos difundir este conocimiento para que esté a disposición de todos los interesados. Esto es lo que sucede actualmente con el Centro de Información que se está gestando al interior del IDEA de la Universidad Nacional de Colombia Sede Manizales, con el apoyo de CORPOCALDAS, la CHEC, la OMPAD y CENICAFÉ, quienes vienen apoyando este proyecto que permite mejorar y poner a disposición del público la información hidroclimatológica del departamento de Caldas mediante las redes de monitoreo hidroclimatológico, por lo que disponer de una línea base es de especial interés para la región y los planificadores.

Las herramientas como la minería de datos han sido empleadas para la predicción de caudales con la novedad que se hace uso de las redes neuronales y de algoritmos evolutivos en el marco de la exploración de la variables internas de un modelo hidrológico de tipo conceptual, por lo que este tipo de modelos híbridos abre un abanico de posibilidades en la predicción de caudales en tiempo real.



# Capítulo 1

## Análisis climatológico para el departamento de Caldas

*Olga Lucía Ocampo López*<sup>1,2</sup>

*Jorge Julián Vélez Upegui*<sup>1</sup>

<sup>1</sup>Universidad Nacional de Colombia Sede Manizales, Facultad de Ingeniería y Arquitectura. IDEA Campus Palogrande Manizales, Colombia.

olocampol@unal.edu.co, jjvelez@unal.edu.co

<sup>2</sup>Universidad Autónoma de Manizales, Antigua Estación del Ferrocarril, Manizales, Colombia.

olocampo@autonoma.edu.co

*... En el corazón de todos los inviernos vive una primavera palpitante, y detrás de cada noche, viene una aurora sonriente. Por muy larga que sea la tormenta, el sol siempre vuelve a brillar entre las nubes. Khalil Gibran (1883-1931) Ensayista, novelista y poeta libanés*

...

### Resumen

Este capítulo presenta una revisión de los factores que afectan la variabilidad climática en Colombia y en el departamento de Caldas en escalas temporales que incluyen la escala interdecadal, interanual, anual, estacional y el ciclo diario. Esta alta variabilidad es característica de las regiones tropicales y está influenciada no sólo por los fenómenos macroclimáticos, de mesoescala y de escala sinóptica, sino también por los factores locales en los cuales la orografía y la interacción suelo atmósfera cumplen un papel importante. El comportamiento de las variables climáticas y el análisis realizado en este capítulo se convierten en elementos de Planificación de cuenca y ordenación del territorio para quienes toman las decisiones. El estudio concluye que es necesario mejorar las redes de monitoreo de las variables climáticas, precipitación, temperatura, radiación solar, humedad relativa, velocidad y dirección de viento, en el departamento acorde con la complejidad del clima en la región.

**Palabras Clave:** Clima, variabilidad climática, precipitación, temperatura, radiación solar, humedad relativa, velocidad y dirección del viento.

## 1.1. Introducción

Las regiones tropicales exhiben condiciones climáticas complejas de alta variabilidad; generar conocimiento que permita comprender la dinámica del clima en estas regiones, es relevante para ser incluida en múltiples actividades, en especial para los procesos de ordenación del territorio y la gestión integrada de cuencas hidrográficas. Este capítulo presenta una visión general del clima en el departamento de Caldas, situado en los Andes tropicales, entre las cordilleras Occidental y Central, en la región Andina Colombiana. Describe el comportamiento medio de las variables climáticas para el período 1981-2010 y muestra una síntesis de los principales factores que influyen sobre la variabilidad climática en el territorio. Inicialmente, se describen los conceptos de clima, tiempo atmosférico y los factores que afectan la variabilidad climática, con énfasis en Colombia; luego, se exponen los análisis de las diferentes variables climatológicas: radiación solar, brillo solar, temperatura, humedad relativa, y precipitación para el departamento de Caldas. Finalmente, se presentan los mapas de la clasificación climática con base en clasificaciones internacionales.

### 1.1.1. Clima y tiempo atmosférico

El clima es el conjunto fluctuante de las condiciones atmosféricas, que se caracteriza por los estados del tiempo en una región durante un período de tiempo relativamente largo [IDEAM, 2005a]. Suele definirse como el estado promedio del tiempo durante muchos años (Organización Meteorológica Mundial –OMM-, 2011). Por otra parte, el tiempo atmosférico es la manifestación de la dinámica de la atmósfera en un lugar y momento determinados [IDEAM, 2005a]; describe las condiciones atmosféricas para ese momento y lugar específicos [FECYT., 2004] y es por tanto, la manifestación diaria del clima [OMM, 2013b].

El tiempo y el clima desempeñan una función decisiva en la producción de alimentos y la disponibilidad de agua; en la generación y uso de energía; en las actividades industriales, de transporte, ocio y socioeconómicas [OMM, 2002] [OMM, 2003]; influyen por tanto en el desarrollo y en el bienestar humano [OMM, 2005] [OMM, 2006] [OMM, 2013b]. Afectan incluso el humor de la gente, su carácter, hábitos y rigen su manera de pensar [OMM, 2003]. A lo largo de los años, las poblaciones han prosperado adaptándose al clima local, que configura el desarrollo socioeconómico y la identidad cultural. En el pasado, cuando las poblaciones migraban por condiciones climáticas como sequías, crecidas o condiciones adversas, provocadas por el clima, se adaptaban gradualmente a su nuevo entorno aprendiendo a conocer el tiempo y el clima y aplicando esos conocimientos para obtener seguridad y bienestar [OMM, 2011]. Hoy en día, la demanda de información sobre el clima es cada vez mayor por el cambio climático producto de la acción humana [IPCC., 2007] [IDEAM, 2010]. La observación y el seguimiento son uno de los pilares que busca ayudar a todos los sectores socioeconómicos a afrontar los impactos de la variabilidad del clima y del cambio climático, a mitigar los riesgos de desastres naturales y a proteger la seguridad alimentaria, la salud y los recursos hídricos [OMM, 2012]. Por tanto, las regiones y países deben contar con un sistema de monitoreo y vigilancia del clima [OMM, 2005] [OMM, 2006] [OMM, 2013a]. Este sistema es un mecanismo que sirve para interactuar con los usuarios y alertarlos de las anomalías climáticas y los fenómenos extremos importantes [OMM, 2010]. De acuerdo con la OMM [OMM, 2012] es necesario ampliar el grado de exactitud y cobertura de estos sistemas, mejorar la difusión y fomentar el uso de esta información climática en la toma de decisiones. En términos generales, la vigilancia climática se ha mejorado gracias a los avances en las tecnologías y en los sistemas de información que han permitido el registro de las variables climáticas en tiempo real [Wang et al., 2008]. La OMM [OMM, 2004] recomienda como mínimo la frecuencia diaria para vigilar y predecir eficazmente los fenómenos climáticos extremos. Sin embargo, las observaciones sistemáticas de largos períodos son de vital importancia para comprender la variabilidad natural y para detectar si las actividades humanas la están alterando [OMM, 2012].

### 1.1.2. Sistema climático

El Sistema climático es un sistema abierto, complejo e interactivo; constituido por la atmósfera, capa gaseosa que envuelve la tierra; la hidrosfera, océanos y agua dulce en estado líquido; la criosfera, agua en estado sólido; la litosfera, superficie de la tierra; la biosfera, seres vivos que habitan la tierra, y la radiación solar incidente [OMM, 2003] [OMM, 2005] [OMM, 2011]. Los componentes del sistema climático son alterados por factores naturales como las variaciones en la órbita y la inclinación de la tierra, la posición relativa de sus ejes, los cambios en la actividad solar y las erupciones volcánicas [OMM, 2003]. Por otra parte, como la mezcla de gases que conforman la atmósfera permite que entre parte de la radiación solar; los componentes de este sistema también son controlados por los factores forzantes como la concentración de gases de efecto invernadero (GEI) y los aerosoles [IDEAM, 2005a]. Los GEI como el vapor de agua, el dióxido de carbono, el metano, el óxido nitroso, entre otros, retienen parte de la energía que se refleja en la superficie del planeta, creando así el efecto invernadero que hace posible la vida en la Tierra [OMM, 2005]; si no existiera este efecto, la temperatura media mundial sería aproximadamente de  $-19^{\circ}\text{C}$ . Sin embargo, durante el último medio siglo se ha elevado la concentración de los GEI hasta superar los límites naturales [OMM, 2005] [IPCC., 2007], causando el calentamiento global. Los aerosoles son finas partículas de origen natural o antropogénico emitidas a la tropósfera que pueden reflejar o absorber radiación solar. Su presencia altera la cantidad y reflectividad de las nubes [IPCC., 2007] y por tanto, son un forzante radiativo negativo que tiende a enfriar el sistema climático. Existen además una serie de condiciones físicas y geográficas llamadas factores determinantes que influyen marcadamente en el clima como la latitud, la altitud, la distribución de las tierras y los mares, la continentalidad, la topografía y las corrientes oceánicas [Lowry., 1973] [OMM, 2003] [IDEAM, 2005a]. La latitud por ejemplo, define las grandes regiones latitudinales o sistemas de tiempo [Griffiths., 1994]: tropical, subtropical, sistemas de latitud media y polares. El trópico comprendido entre latitudes  $23.5^{\circ}\text{N}$  y  $23.5^{\circ}\text{S}$  presenta condiciones que lo diferencian de las demás regiones del mundo. Su clima y el tiempo atmosférico están influenciados por los fenómenos macro-climáticos, de meso-escala o de escala sinóptica; los mecanismos de convección que dominan la lluvia tropical y la interacción suelo-atmósfera [Poveda, 2004]. En esta zona los efectos de los ciclos día-noche son más importantes que los ciclos estacionales [Jaramillo, 2005]. Se caracteriza además por la óptima fijación de carbono y la alta capacidad fotosintética que globalmente corresponde a los bosques húmedos tropicales [Jaramillo, 2005]. El trópico es de interés especial, porque alberga la mayor biodiversidad que cualquier otra zona de la Tierra [Herzog et al., 2013]. Finalmente, la humedad del suelo tiene una marcada influencia en la intercambio de agua y energía y es por tanto una variable fundamental del sistema climático [OMM, 2013a]. Es necesario mejorar la comprensión de la interacción suelo-atmósfera y del papel que juega la humedad del suelo y la evapotranspiración en la precipitación tropical en las distintas escalas especiales y temporales [Poveda, 2004].

### 1.1.3. Variables climáticas

Como el componente atmosférico es el que más claramente caracteriza el clima [OMM, 2011], sus condiciones predominantes se describen a partir una amplia gama de variables como son temperatura, precipitación, presión atmosférica, duración de la insolación, dirección y velocidad del viento. Otros elementos pueden comprender humedad, nubosidad, condiciones meteorológicas extremas como tormentas e incluso el tipo de suelo [OMM, 2003]. En climatología se utilizan los valores promedios para definir y comparar el clima [IDEAM, 2005a]; por tanto, suele describirse en términos del valor medio y la variabilidad de una serie de variables meteorológicas -temperatura tales como, precipitación, presión atmosférica, insolación, viento, humedad, cubierta de nubes- para una región durante cierto período de tiempo [OMM, 2005] [OMM, 2011]. La normal climática, la cual representa el clima presente, es una medida utilizada con este propósito [IDEAM, 2005a]. La OMM la define como el promedio mensual de los datos climatológicos computados para un período de 30 años; incluye la distribución de frecuencias y otros estadísticos como cuantiles, intervalos de confianza e información de los valores extremos, para describir el comportamiento de la variable [WMO, 2007].

#### 1.1.4. Variabilidad climática natural

El clima cambia en diferentes escalas temporales lo que provoca impactos sobre numerosos aspectos sociales, económicos y medioambientales que afectan la agricultura, la salud, el turismo, la energía y el agua entre otros sectores [FAO, 2003] [Wang et al., 2008] [Stern and Nicholas., 2006] [IPCC., 2007]. En diferentes años, los valores de las variables climatológicas fluctúan por encima o por debajo de la normal climatológica. Las fluctuaciones observadas en el clima, durante períodos de tiempo relativamente cortos, se conocen como variabilidad climática. A la diferencia entre el valor registrado de la variable y su promedio se le conoce como anomalía [IDEAM, 2005a]. El clima varía naturalmente, por los cambios en los factores radiativos forzantes y las modificaciones en la forma de interacción de los distintos componentes del sistema climático [IDEAM, 2005a]. La variabilidad natural es por tanto, parte intrínseca del clima; obedece a los ciclos y tendencias de la órbita terrestre, la radiación solar incidente, la composición química de la atmósfera, la circulación de los océanos, la biosfera, las erupciones volcánicas y otros fenómenos naturales [OMM, 2013a]. Estas fluctuaciones se dan a lo largos de estaciones, decenios y siglos, en diferentes escalas temporales, entre las que se destacan [OMM, 2013a] [IDEAM, 2005a] [Montealegre, 2005]:

- **Intraestacional:** dentro de las estaciones se presentan oscilaciones que determinan el estado del tiempo, durante semanas o de uno a dos meses. Se destaca una señal de tipo ondulatorio denominada de 30-60 días, de gran importancia en el proceso de predicción climática.
- **Estacional:** corresponde a la fluctuación del clima a escala mensual. En el trópico se tienen temporadas lluviosas y secas en el ciclo anual, fundamental en la variabilidad climática a este nivel. La migración de la Zona de Confluencia Intertropical, ZCIT, es una de las más importantes fluctuaciones climáticas de la escala estacional en el trópico.
- **Interanual:** se manifiesta por las variaciones que se presentan en las variables climatológicas de año en año. En esta escala, la variabilidad climática podría estar relacionada con alteraciones en el balance global de radiación. Los fenómenos El Niño - La Niña - Oscilación del Sur y la Oscilación Cuasibienal son ejemplos típicos de variabilidad climática interanual.
- **Interdecadal:** determinada por las fluctuaciones del clima a nivel de décadas. Esta variabilidad influye notablemente en las actividades humanas y es clave para la determinación de posibles tendencias en las variables climáticas.

Muchos de estos ciclos aparentes no son oscilaciones desconocidas, sino que están relacionados con los procesos de interacción Tierra-Atmósfera-Océano. Entre los más importantes fenómenos a gran escala impulsores de la variabilidad estacional e interanual del clima mundial se tienen: El Niño/Oscilación del Sur (ENSO), el Dipolo del Océano Índico (DOI), la Oscilación Ártica (OA), la Oscilación del Atlántico Norte (OAN) y el Modo Anular del Sur (SAM) [OMM, 2004] [OMM, 2012]. Otras características meteorológicas afectan el clima al fortalecer o debilitar corrientes de aire denominadas corrientes en chorro [OMM, 2013a]. La vigilancia climática debe por tanto ofrecer la información indispensable para desarrollar una planificación de manera eficaz que permita responder a las variaciones climáticas en términos de frecuencia, intensidad y localización.

## 1.2. Clima y variabilidad climática en Colombia

En Colombia, el clima está determinado por la ubicación del país en la franja ecuatorial y por factores geográficos como la altitud, la distancia al mar y la continentalidad [Jaramillo, 2005] [IDEAM, 2005a]. El clima está influenciado por los procesos que ocurren en los océanos Atlántico y Pacífico; la dinámica de las cuencas de los ríos Amazonas y Orinoco y las

fuertes interacciones suelo-atmósfera [Poveda, 2004] [Poveda et al., 2007] [Poveda et al., 2011]. Estos aspectos crean un amplio espectro de climas y microclimas que van desde los más calurosos hasta lo más fríos en los picos de las montañas de las Cordilleras de los Andes [Jaramillo, 2005] [IDEAM, 2010]. El IDEAM [IDEAM, 2005a] (2005a) presenta una descripción del clima colombiano basado en las clasificaciones climáticas Internacionales. Otros autores [Eslava et al., 1986e] [Eslava et al., 1986c] [Eslava et al., 1986b] [Eslava et al., 1986d] [Eslava, 1993] [Mesa et al., 1997] [Pabón et al., 1998] [Pabón et al., 2001] [Jaramillo, 2005] han desarrollado también descripciones del clima colombiano. La diversidad climática es un recurso importante del país y juega un papel clave en el desarrollo socioeconómico. Las ecoregiones colombianas -Andina, Orinoquía, Amazonía, Pacífica y Caribe- poseen características climatológicas diversas y complejas. Sin embargo, predominan los climas cálidos y húmedos que representan un 60.54 % del territorio [IDEAM, 2005a]. La región Andina por sus características montañosas, presenta muy amplia diversidad de climas causada por la variación de los pisos térmicos [Poveda, 2004] [Jaramillo, 2005] [IDEAM, 2005a] [IDEAM, 2010]. Se destaca en esta región la cordillera de los Andes que se divide en las cordilleras Occidental, Central y Oriental, separadas por los valles longitudinales de los ríos Cauca y Magdalena [IDEAM, 2010]. En los Andes colombianos el clima es muy húmedo, con baja estacionalidad térmica y marcadas variaciones diurnas de temperatura. La variabilidad de la temperatura está influenciada por los gradientes altitudinales y la humedad del aire [Eslava, 1993] [Martínez et al., 2013]. De igual manera, la precipitación es muy diversa con amplia variabilidad espacial y temporal [Eslava, 1993] [Pabón et al., 2001] [Poveda, 2011]. Los principales factores que influyen el clima en la región Andina son la presencia del bosque amazónico al este, el desplazamiento de la Zona de Convergencia Intertropical (ZCIT), los vientos Alisios al Norte, las corrientes en chorro y el sistema permanente de alta presión del Pacífico Sur y el Atlántico Norte [Poveda, 2004] [Jaramillo, 2005] [Martínez et al., 2013]. Los factores locales tienen efectos significativos en la región Andina, por la diferencia entre el comportamiento físico de las superficies tierra y agua; valle y montaña; los cambios en el uso del suelo, y el desarrollo urbano acelerado. Con respecto a la variabilidad climática natural, depende de las condiciones atmosféricas, que presentan anomalías frente a la normal climatológica. Los fenómenos naturales que producen eventos extremos, son frentes fríos, estacionarios, secos, huracanes, perturbaciones tropicales y células con una humedad desproporcionada, los cuales pueden provocar precipitación excesiva o sequías prolongadas. En Colombia, al igual que en muchas partes del mundo, se considera que tales fenómenos están relacionados con la Oscilación del Sur-El Niño [CEPAL, 2009] [IPCC., 2007] [Trenberth and Thomas., 2003] [Adger et al., 2007]. Por su localización en el Trópico, el clima y el tiempo atmosférico en Colombia están influenciados por los fenómenos macro-climáticos, de meso-escala o de escala sinóptica y los mecanismos de convección [Poveda and Mesa, 1997] [Poveda et al., 2002] [Poveda, 2004]. Estos fenómenos que influyen en la variabilidad climática en las distintas escalas temporales se describen a continuación.

### 1.2.1. Variabilidad climática interdecadal

La variabilidad climática a escalas temporales decadales de interdecadales está controlada por fenómenos macro-climáticos que operan en esta escala temporal como la Oscilación Decadal del Pacífico (PDO) y la Oscilación del Atlántico Norte (NAO) que tienen influencia significativa sobre el clima de Colombia, según lo reportado por [Poveda et al., 1998] [Poveda et al., 2002] [Poveda, 2004]. La Oscilación Decadal del Pacífico (PDO) es un patrón exhibido en las aguas superficiales del Pacífico Norte a 20°N que pasa de las fases cálidas a frías [Mantua and Hare., 2002]; durante la fase cálida o positiva el Pacífico Occidental se enfría, mientras el Oriental se calienta; por otra parte, en la fase negativa se produce el efecto contrario. Las anomalías en la precipitación y temperatura relacionadas con la PDO en Suramérica son espacialmente similares al ENSO, pero su amplitud es la mitad de las equivalentes en ENSO [Garreaud et al., 2008]. Existen cada vez más pruebas sobre su influencia en la región Andina tropical [Mantua and Hare., 2002] [Poveda et al., 2002] [Martínez et al., 2013]. NAO es un modo de variabilidad a gran escala en el sistema climático relacionada con cambios en la velocidad del viento así como su dirección sobre el océano y en las trayectorias de las tormentas de invierno [Hurrell and

Loon., 1997] [Hurrell et al., 2003] [Jones et al., 2003] [Bojariu and Gimeno, 2003] que producen cambios en la distribución de precipitación y temperatura, en especial en el hemisferio norte. En Colombia, la correlación entre la NAO y la hidroclimatología no parece muy alta, pero es significativa [Poveda et al., 1998] [Poveda et al., 2002]. Los vínculos entre estos fenómenos y sus dinámicas con ENSO requieren de una mayor comprensión [Newman et al., 2003] [Poveda et al., 2002] [Newman et al., 2003] [Schneider and Cornuelle., 2005] [Martínez et al., 2013]. Por otra parte, es evidente la superposición de los efectos combinados de la variabilidad climática natural y del cambio climático antropogénico, tema central y actual de investigación [Poveda, 2004] [IPCC., 2007].

### 1.2.2. Variabilidad climática interanual

Los fenómenos más importantes que influyen en el clima del país a escalas de tiempo interanual son el ENSO y sus fases El Niño y La Niña, pero otros como la NAO y la PDO ejercen un efecto no despreciable [Poveda, 2004]. Durante las fases del ENSO en Colombia, existen una serie de mecanismos físicos que modulan las anomalías [Poveda and Mesa, 1997] [Poveda and O.J., 1999] [Poveda et al., 1998]: 1) El debilitamiento de la corriente chorro del Chocó; 2) El debilitamiento de la corriente de chorro ecuatorial de los 700 hPa; 3) la reducción de advección de humedad desde el mar Caribe; 4) la reducción en el número e intensidad de ondas tropicales del este sobre el Atlántico tropical norte; 5) El desplazamiento de ZCIT hacia el sur-oeste de su posición normal; y 6) la retro-alimentación entre procesos tierra-atmósfera. El Niño/Oscilación del Sur (ENSO) es el más conocido de los modos globales de variabilidad interanual del clima [IPCC, 2008] [OMM, 2010] [OMM, 2012] y es el principal para Colombia [Poveda, 2004]. Es un fenómeno global de interacción entre el océano y la atmósfera, que origina fluctuaciones de la temperatura superficial y de la presión del aire en el océano Pacífico, en cuyo transcurso se alternan episodios cálidos y fríos, más conocidos como El Niño y La Niña, respectivamente [PNUD, 2000] [NOAA, 2013]. En condiciones normales las temperaturas de la superficie del mar en la zona central y oriental del Océano Pacífico están algunos grados más bajas que las de la zona occidental; durante El Niño, suben por encima de los niveles normales; en consecuencia, se reducen: el gradiente de temperatura del oeste a este, el gradiente de presión del aire y la intensidad de los vientos alisios orientales sobre el pacífico tropical [NOAA, 2013] [OMM, 2012]. Los episodios El Niño suceden a intervalos entre dos y siete años [Dore, 2004], cuando los vientos alisios del trópico se debilitan o invierten su dirección habitual; en este momento desplazan las aguas superficiales hacia el Océano Pacífico oriental y hacia la costa ecuatorial occidental de Suramérica; con esta corriente se desplazan las lluvias que pueden causar inundaciones en unas regiones y sequías en otras [OMM, 2005] [OMM, 2012]. En Colombia, la fase cálida está asociada con incremento en la temperatura media y reducciones en la humedad del suelo, la precipitación y la descarga de los ríos [IDEAM, 1998]. En el fenómeno opuesto, La Niña, los vientos alisios del sur se intensifican frente a las costas suramericanas y provocan una reducción de la temperatura del océano Pacífico [NOAA, 2013]. Influye en el clima nacional y produce reducción en la temperatura ambiente; incremento de las lluvias en las regiones Caribe y Andina; reducción de la radiación solar y efectos en la frecuencia de los ciclones tropicales del Caribe. Los impactos incluyen aumento significativo de los niveles de los ríos y con ellos la probabilidad de inundaciones, crecientes súbitas en las zonas de alta pendiente, e incremento en la probabilidad de deslizamientos de tierra [Velásquez and Rosales, 2002] [IDEAM, 2009]. Los efectos e impactos del ENSO y sus fases El Niño y La Niña, en la precipitación, temperatura, descarga de los ríos y humedad del suelo han sido ampliamente estudiados en Colombia [IDEAM, 1998] [Jaramillo and Ramírez, 2009] [Pabón and Rivera., 1993] [Poveda and O.J., 1993] [Poveda and O.J., 1996] [Poveda, 1994] [Poveda et al., 1998] [Poveda et al., 2001] [Poveda et al., 2002] [Poveda and Jaramillo, 2000] [Salazar, 1990] [Waylen and Poveda., 2002] [Puertas and Carvajal., 2008] [Poveda et al., 2011] [Hoyos et al., 2013]. La Niña 2010-2011 afectó alrededor de cuatro millones de colombianos y causó pérdidas económicas de aproximadamente 7.8 billones de dólares relacionadas con la destrucción de infraestructura, inundaciones de cultivos y pagos de subsidios por el gobierno Nacional [Hoyos et al., 2013] (Hoyos et al., 2013) Como consecuencia, ENSO es el evento natural

de mayor impacto socioeconómico [IPCC, 2008] (IPCC, 2008). Las condiciones meteorológicas extremas asociadas con las fases, cálida -El Niño- y fría -La Niña- producen sequías e inundaciones en las regiones tropicales y subtropicales [OMM, 2004]. El tiempo que trae consigo estos fenómenos es tan extremo, que es una de las grandes prioridades de los servicios meteorológicos e hidrológicos en el mundo [OMM, 2005] [OMM, 2012] quienes vigilan las temperaturas en la superficie del Océano Pacífico [NOAA, 2013] y han desarrollado una serie de índices, para identificar y cuantificar la fuerza del fenómeno.

### 1.2.3. Variabilidad climática estacional

El ciclo anual está dominado por la migración meridional de la zona de convergencia intertropical (ZCIT), la dinámica del transporte de humedad de los vientos alisios del este, la actividad del chorro del Chocó y los sistemas convectivos de mesoescala [Poveda, 2004] (Poveda, 2004); sin embargo, los factores de interacción suelo-atmósfera en esta escala son importantes [Vélez et al., 2002] [Poveda, 2004]. A continuación se describen algunos de estos fenómenos:

**La Zona de Confluencia Intertropical (ZCIT):** es una banda de presión que rodea la Tierra cerca del Ecuador y es la zona donde convergen los vientos alisios provenientes de ambos hemisferios [IDEAM, 2005a]. Esta zona se mueve latitudinalmente durante el año [Mitchell and Wallace., 1992]. A su paso por las distintas regiones genera temporadas de máxima nubosidad [Ruiz et al., 2009] y precipitación de intensidad variable. Por los Andes colombianos pasa dos veces al año, determinando un ciclo anual bimodal de precipitación, con estaciones lluviosas y relativamente secas [Eslava, 1993] [Mejía et al., 1999] [Jaramillo, 2005] [León et al., 2000] [Pabón and Rivera., 1993]

[Pabón et al., 2001] [Poveda, 2004] [Poveda et al., 2007] [Martínez et al., 2013].

**Sistemas convectivos de Meso-escala:** contribuyen al incremento de la precipitación en Colombia y están fuertemente determinados por la migración de la ZCIT. Las zonas más favorables para la formación de estos sistemas son el este del Pacífico, la Costa Pacífica Colombiana, la región Andina, el Valle del río Magdalena, el Norte de Antioquia, la cuenca del Orinoco y la Amazónica [Poveda, 2004].

**La corriente en chorro del Chocó:** corriente de bajo nivel proveniente del Océano Pacífico que penetra en el interior transportando grandes cantidades de humedad; contribuye a la advección de humedad por parte de vientos relativamente fríos que interactúan con los vientos alisios del este más cálidos, causando inestabilidad atmosférica, alta convección profunda y precipitaciones. Esta corriente permite explicar la existencia de complejos convectivos de meso-escala sobre Colombia, los cuales tienen una fuerte incidencia en el ciclo diario. Está fuertemente modulada por el ciclo anual, pero su núcleo se localiza en 5°N a través de todo el año, aunque contribuye a las altas precipitaciones durante septiembre y noviembre [Poveda and O.J., 1999] [Poveda and O.J., 2000] [Poveda, 2004].

**Los sistemas sinópticos del Pacífico:** la circulación atmosférica en el Océano Pacífico frente al litoral colombiano, entre los meses marzo a diciembre y específicamente entre junio y septiembre, se caracteriza por el recurvamiento de los vientos Alisios, tomando direcciones del sur con componente Oeste. Este fenómeno combinado con el flujo de las ondas del Este del Mar Caribe, favorecen el desarrollo de un sistema de baja presión conocido como la baja anclada de Panamá, que da origen a movimientos verticales ascendentes y nubes de gran dimensión vertical [Pabón et al., 1998] [IDEAM, 2005a]. La actividad convectiva es intensificada por las condiciones térmicas de las aguas cálidas del Océano Pacífico, durante el período de verano en el hemisferio norte; la humedad de la zona y las condiciones orográficas de la cordillera de los Andes. Este sistema afecta significativamente el tiempo en la región Andina [Pabón et al., 1998] [Poveda, 2004] [IDEAM, 2005a].

**Los sistemas sinópticos de la amazonía:** se presentan entre mayo y noviembre y específicamente a mitad de año, por el movimiento de zonas frontales frías del polo sur que atraviesan el trópico y originan líneas de inestabilidad y convección en la Amazonía colombiana. En el sureste del país, también afecta la Baja Amazónica, que es un sistema de baja presión que tiene su origen en el calor latente de condensación y la canalización del flujo de los vientos Alisios del noreste por efecto de la cordillera de los Andes, lo cual contribuye a intensificar la convección

y las lluvias en el sur del país y al desplazamiento temporal de la ZCIT hacia el sur, a mitad de año. Estos sistemas pueden llegar a modificar el estado del tiempo a lo largo de la cordillera Oriental y en la Orinoquía Colombiana [Pabón et al., 1998] [IDEAM, 2005a] e influyen en el tiempo en la región Andina [Poveda, 2004].

#### 1.2.4. Variabilidad climática intraestacional

En la escala intraestacional, la variabilidad climática está influenciada por las oscilaciones intra-estacionales de 30-60 días, por las ondas del este, por los ciclones tropicales y los frentes fríos de altas altitudes [Poveda et al., 2002] [Poveda, 2004]. Sin embargo, los fenómenos de interacción suelo-atmósfera juegan un papel importante [Poveda, 2004]. A continuación se describen algunos de estos fenómenos.

**Oscilaciones de 30-60 días:** están asociadas con la oscilación de Madden-Julian [Madden and Julian., 1971] llamada también intra-anual. Parece estar vinculada al comienzo y fin del ENSO [Weickman, 1991] y afecta fuertemente el ciclo de precipitaciones en Colombia [Poveda et al., 2002]

**Ondas Tropicales del Este del Caribe:** son perturbaciones de tipo ondulatorio de verano y otoño, entre junio-noviembre, sobre el Atlántico tropical norte; son de especial interés, porque al intensificarse originan tormentas tropicales y ciclones en el Atlántico Norte y en el mar Caribe [Mesa et al., 1997] [IDEAM, 2005a] [Poveda, 2004]. Afectan el estado del tiempo y son frecuentes en la temporada de lluvias del norte del país, el valle del río Magdalena y el oriente del país [IDEAM, 2005a].

**Las vaguadas de latitudes medias del hemisferio Norte:** un tipo especial de ondas del Este, denominado ondas inducidas en los Estes se producen durante el invierno del hemisferio norte, durante diciembre y marzo, cuando hay vientos del Este en la troposfera inferior y del Oeste en la troposfera superior. La vaguada, eje de baja presión, induce una onda de flujo de los Estes y el sistema acoplado de vaguadas de latitudes medias se desplaza hacia el Este, provocando movimiento ascendente y la intensificación de la nubosidad y las lluvias [Pabón et al., 1998] [IDEAM, 2005a].

**La vaguada tropical de la alta troposfera- TUTT:** se presenta entre junio y noviembre. Es uno de los sistemas más importantes en la convección del Caribe y en la formación de ciclones tropicales en el Caribe occidental; genera núcleos nubosos que dependen de la disponibilidad de calor y de humedad [IDEAM, 2005a].

**Ciclones tropicales:** se originan por la intensificación de las perturbaciones tropicales sobre el océano Atlántico y el mar Caribe entre los 5° y 20° de latitud norte, donde las aguas marítimas son muy cálidas entre los meses de junio a noviembre; específicamente entre agosto y octubre. La costa Caribe y el archipiélago de San Andrés y Providencia se ven afectados por estos fenómenos aunque con baja frecuencia [Pabón et al., 1998] [IDEAM, 2005a].

#### 1.2.5. Ciclo diurno

La amplitud del ciclo diurno de temperatura es la características que determina la variabilidad climática en el trópico [Poveda, 2004] (Poveda, 2004). El ciclo diurno de precipitación es ampliamente variable en tiempo y espacio; esta variabilidad es inherente al ciclo diurno de insolación, los procesos de interacción suelo-atmósfera, la dinámica diurna de sistemas convectivos de meso-escala tanto oceánicos como sobre tierra [Poveda, 2004] [Poveda et al., 2005].

En los Andes colombianos se evidencia la existencia de un ciclo diurno (24 h) y semidiurno (12 horas). La conducta de precipitación se clasifica dentro de dos regímenes diferentes de acuerdo con la localización: se presenta un pico de precipitación alrededor del mediodía en las regiones localizadas al oeste y noreste; mientras que se tienen dos máximos de actividad convectiva, uno a media noche y otro hacia las nueve de la mañana en las regiones del sureste [Poveda et al., 2005] [Martínez et al., 2013].

Los máximos de precipitación son producidos por inestabilidades atmosféricas locales que son

influenciadas por la formación de nubes y niebla en los bosques nublados; advección de la humedad producida durante los procesos diurnos, y procesos de enfriamiento y sequía ocasionados por las dinámicas nocturnas [Ruiz et al., 2009] [Martínez et al., 2013].

### 1.3. Caracterización de la zona de estudio

El departamento de Caldas, Colombia es la zona de estudio, su localización se presenta en la Figura 1.1. De acuerdo con la posición astronómica, se encuentra en el Sistema Tropical, en la zona de baja latitud ecuatorial o intertropical. El departamento de Caldas está ubicado en la región Andina colombiana entre las cordilleras Central y Occidental. Su superficie representa una extensión de  $7450 \text{ km}^2$ .

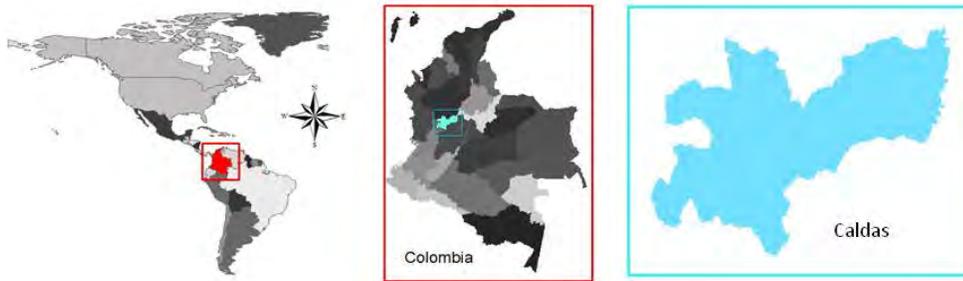


Figura 1.1: Localización del departamento de Caldas. Fuente: propia

El departamento comprende áreas montañosas que corresponden al macizo volcánico de la cordillera Central de los Andes; zonas planas y onduladas al oriente, perteneciente al valle medio del río Magdalena, y al occidente, por los valles de los ríos Risaralda y Cauca. [CORPOCALDAS, 2001] [CORPOCALDAS, 2007], como se observa en el modelo de elevación digital del terreno –MED- que se presenta en la Figura 1.2.

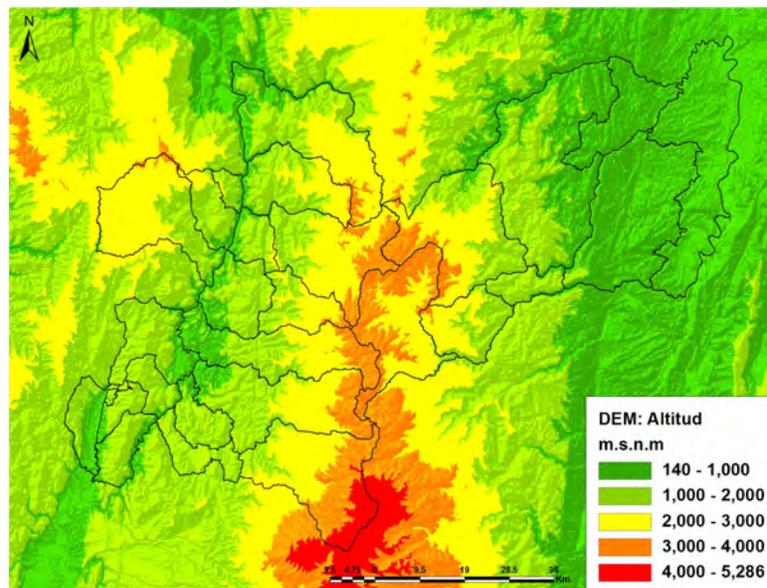


Figura 1.2: Modelo de Elevación Digital del Terreno. Fuente: Adaptado (USGS)

El terreno por tanto, tiene fuertes pendientes en las áreas montañosas y es relativamente plano en los valles de los ríos, como se aprecia en la Figura 1.3. en el mapa de pendientes. Considerando la división político-administrativo que se observa en la Figura 1.4., el departamento está dividido en 27 municipios, distribuidos en seis distritos o provincias, según la regionalización establecida por la gobernación de Caldas: Magdalena Caldense, Alto Oriente, Norte, Alto Occidente, Centro Sur y Bajo Occidente (Figura 1.4).

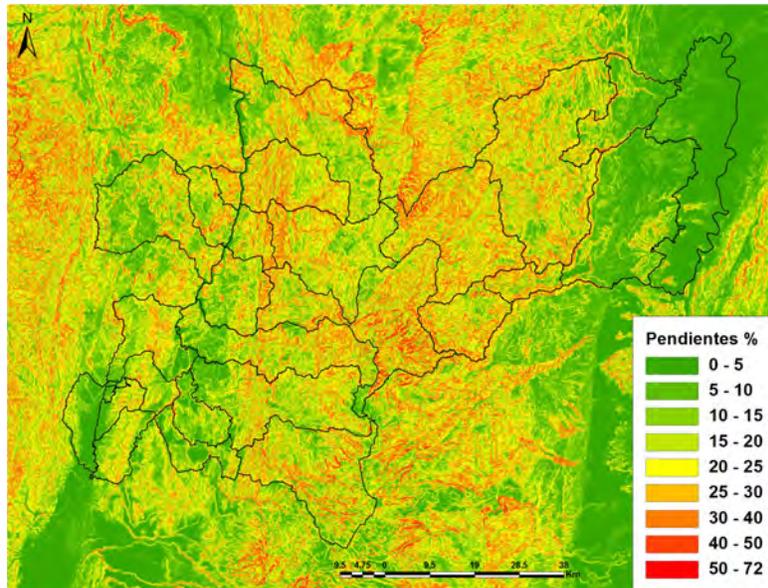


Figura 1.3: Mapa de pendientes. Fuente: Adaptado (USGS)



Figura 1.4: Distritos o Provincias del departamento de Caldas. Fuente: Adaptado (IGAC)

Las cuencas a escala 1:25.000 se presentan en la Figura 1.5.; la delimitación fue realizada a partir del MED empleando las herramientas de los Sistemas de Información Geográfica (SIG). En la vertiente del río Magdalena se tienen las cuencas de los ríos: La Miel, aferentes directos al Magdalena, Guarinó y río Samaná sur; mientras que en la vertiente del río Cauca se encuentran las cuencas de los ríos: Chinchiná, Pozo-Maibá, Risaralda, Arma, Tapias-Tareas, Supía, Pácora, Campoalegre-San Francisco y aferentes directos al Cauca [CORPOCALDAS, 2001] [CORPOCALDAS, 2007].



Figura 1.5: Cuencas hidrográficas del departamento de Caldas. Fuente: propia

[CORPOCALDAS, 2007] reporta 13 zonas de vida que van desde el nival hasta el bosque seco tropical, según la clasificación del Holdridge. La distribución por área se presenta en la tabla 1.1. Se evidencia fragmentación de las zonas de vida en especial los bosques seco, húmedo tropical, muy húmedo y húmedo premontano por las actividades agropecuarias en especial la ganadería y la caficultura que comprenden cerca del 49 % del territorio [CORPOCALDAS, 2007]. El amplio rango altitudinal y la diversidad de climas del departamento favorece la presencia de una gran variedad de especies de flora y fauna y una gran riqueza biótica para el departamento. Sin embargo la biodiversidad está siendo afectada por diversas causas antrópicas [CORPOCALDAS, 2007].

#### 1.4. Análisis de las variables climáticas de caldas

A continuación se describen las principales características de las variables climáticas que caracterizan el clima del departamento de Caldas. Para el análisis se consideraron las estaciones presentes en la zona, información reportada por la Universidad Nacional de Colombia Sede Manizales, la Central Hidroeléctrica de Caldas –CHEC S.A. E.S.P.-, Gestión Energética S.A –GENSA-, CORPOCALDAS, IDEAM y CENICAFÉ. Para el análisis se emplearon los Softwares ArcGis, IBM SPSS Statistics y las herramientas de Microsoft Excel.1.1

Tabla 1.1: Zonas de Vida del departamento de Caldas. Fuente: [CORPOCALDAS, 2007]

Zona de Vida		Área (Ha)	% Territorio
Nival	N	4812.27	0.65
Páramo subpáramo-subalpino	pp-SA	17728.65	2.38
Bosque pluvial –montano	bp-M	40875.79	5.49
Bosque muy húmedo-montano	Bmh-M	58162.83	7.81
Bosque pluvial-montano	bp-MB	2864.41	0.38
Bosque muy húmedo – montano	Bmh-MB	91195.46	12.25
Bosque húmedo – montano bajo	bh-MB	68214.76	9.16
Bosque pluvial – premontano	bp-PM	39561.12	5.31
Bosque muy húmedo premontano	Bmh-PM	125838.33	16.90
Bosque húmedo premontano	bh-PM	83370.34	11.20
Bosque muy húmedo- tropical	Bmh-T	60501.72	8.13
Bosque húmedo- tropical	bh-T	114639.05	15.40
Bosque seco-tropical	bs-T	36828.15	4.95

#### 1.4.1. Brillo solar

El brillo solar, también conocido como insolación, es el número de horas de brillo o de radiación directa. La medición de esta variable se realiza con el Heliógrafo, el cual proporciona las horas de sol efectivo en el día. El valor máximo de brillo solar de una zona cafetera de Colombia, varía entre 11,3-12,6 h/d [Jaramillo, 2005]. Estos valores pueden reducirse por la nubosidad, la orientación y la exposición de las laderas y la amplitud del horizonte. En la zona Andina, el brillo solar anual promedio es de 1550 horas, con extremos superiores de 2050 h/año y mínimos de 1050 h/año [Jaramillo, 2005]. Basados en los mapas de brillo solar para Colombia [IDEAM, 2005b], para el departamento de Caldas fluctúa entre 1 a 7 horas al día, con valores mínimos en las temporadas de lluvias. Para el departamento de Caldas, se tienen registros de brillo solar en las estaciones principales de CENICAFÉ, localizadas en los municipios de Chinchiná, Manizales, Marquetalia, Palestina y en la estación helioplviográfica del municipio de Filadelfia. Para el análisis se incluye también las estaciones localizadas en los departamentos vecinos. La localización geográfica de las estaciones se presenta en la Tabla 1.2. A partir del análisis de los registros históricos se obtiene el comportamiento de la variable para el período 1981-2010 y el análisis de variabilidad climática en diferentes escalas. El comportamiento promedio estacional

Tabla 1.2: Localización geográfica de estaciones climáticas con registros de brillo solar. Fuente: propia

Estación	Municipio	Altitud	Latitud	Longitud	Registros Desde	Registro Hasta	Años Registro
Letras	Manizales	3684.0	5° 3´	75° 20´	01/08/2002	31/12/2011	9
Agromonía	Manizales	2088.0	5° 3´	75° 30´	01/01/1956	31/12/2011	55
Naranjal	Chinchiná	1381.0	4° 58´	75° 39´	01/01/1956	31/12/2011	55
Cenicafé	Chinchiná	1310.0	4° 59´	75° 36´	01/01/1961	31/12/2011	50
Granja Lúker	Palestina	1031.0	5° 4´	75° 41´	01/01/1975	31/12/2011	36
Santágueda	Palestina	1026.0	5° 4´	75° 40´	01/01/1968	31/12/2011	43
El Algarrobo	Manizales	1202	5° 4´	75° 35´	01/01/1993	31/12/2007	14
Rafael Escobar	Supía	1307	5° 27´	75° 38´	01/01/1976	31/12/2004	28
Santa Helena	Marquetalia	1395	5° 19´	75° 0´	01/01/1981	31/12/2011	30
Tricontinental	Fresno	1269	5° 9´	75° 0´	08/01/2002	31/12/2011	9
La Ilusión	Filadelfia	1348	5° 13´	75° 35´	01/04/2003	31/12/2011	8

del brillo solar mensual para el período 1981-2010 se muestra en la Figura 1.6; se evidencia el ciclo bimodal típico, con mayores valores a principios y mediados del año. Con excepción de la estación Santa Helena -ubicada en el oriente de Caldas- donde la estacionalidad es más marcada, las demás estaciones siguen la misma tendencia en el ciclo anual.

Para la espacialización a escala diaria y anual se toma como base el modelo de correlación propuesto por Ocampo et al. [Ocampo et al., 2013] (2013) para la cuenca del río Chinchiná:

$$BrilloSolar < \frac{h}{dia} > = 44.156 * Altitud^{-0.315} \quad (1.1)$$

En las estaciones que no pertenecen a la cuenca del río Chinchiná, los errores en la estimación del brillo solar mensual son: -3 % para Santa Helena y 16 % en Rafael Escobar.

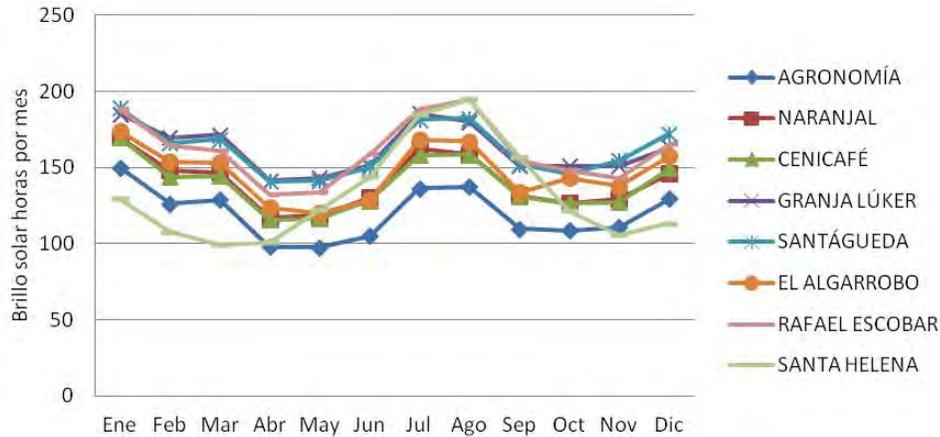


Figura 1.6: Comportamiento estacional del brillo solar mensual 1981-2010. Fuente: propia

La distribución del brillo solar a escala diaria y anual para el período 1981-2010 se presenta en los mapas de las Figuras 1.7 y 1.8, respectivamente. El brillo solar es mayor en el valle del río Magdalena que recibe entre el 68 al 77 % del brillo solar máximo (BSM) posible y menor en las regiones ubicadas en la cordillera Central con solo el 24 al 41 % del BSM. No se dispone de información a menor escala para el análisis del ciclo diario.

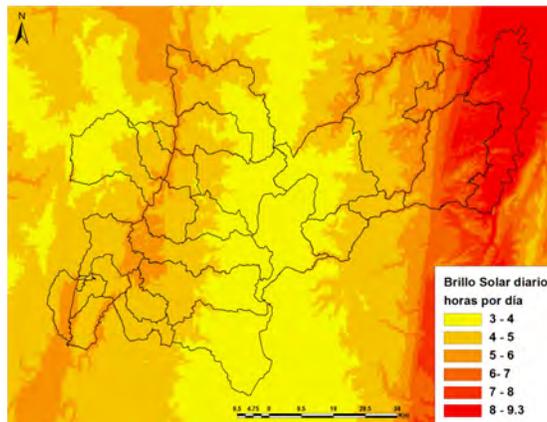


Figura 1.7: Mapa de brillo solar diario 1981-2010. Fuente: propia

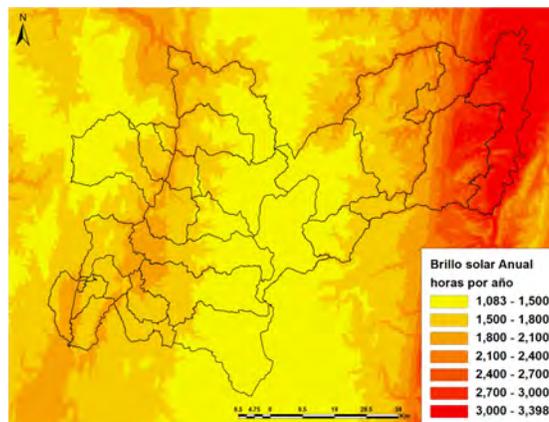


Figura 1.8: Mapa de brillo solar anual 1981-2010. Fuente: propia

En la escala interdecadal, se presentan variaciones en el brillo solar diario para las últimas tres décadas como se observa en la Figura 1.9 para la estación Agronomía. Tomando como base los registros históricos disponibles desde la década de los cincuenta, se tienen tendencias en el brillo solar decrecientes, pero se evidencia un cambio en la tendencia en la última década. El análisis de tendencias en esta escala para la cuenca del río Chinchiná es descrito por [Ocampo, 2012] [Ocampo et al., 2012c] [Ocampo et al., 2012a].

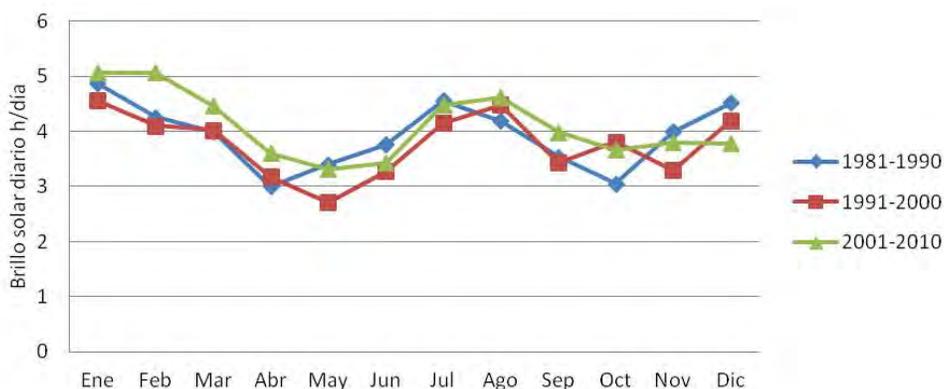


Figura 1.9: Variación Interdecadal del brillo solar diario- Estación Agronomía 1981-2010. Fuente: propia

### 1.4.2. Radiación solar

La radiación electromagnética proveniente del sol se propaga radialmente en el espacio mediante ondas electromagnéticas y partículas. La radiación global es el resultado del componente vertical de la radiación directa más el componente de la radiación difusa. Para la medición se considera la cantidad de energía solar incidente por unidad de área y tiempo o integrada durante un lapso de tiempo que normalmente es un día [IDEAM, 2005b]. Los pirheliómetros y fotómetros se emplean para la medición de la radiación directa; mientras que los piranómetros y pirgeómetros permiten mediciones tanto de la radiación difusa como la global.

La cantidad de radiación solar absorbida por la superficie es afectada por la posición del sol, la altitud, la latitud, la exposición de la ladera, la nubosidad y la neblina [Jaramillo, 2005]. Como Colombia se encuentra en la zona ecuatorial, los valores de radiación solar media son 35.7 - 36.1 MJ/m<sup>2</sup> día [IDEAM, 2005a].

La intensidad de la radiación solar en la región Andina colombiana muestra que las zonas de los valles del Cauca y Magdalena poseen el mayor potencial de esta región y a medida que se asciende hacia las cimas de las cordilleras el potencial se va disminuyendo gradualmente, con excepción de algunos núcleos localizados en los altiplanos, donde se observa un potencial mayor en comparación con las laderas [IDEAM, 2005b]. Basados en los mapas de radiación solar para Colombia [IDEAM, 2005b], la radiación solar para el departamento de Caldas, fluctúa entre 3.5 – 5 kW-h/m<sup>2</sup>. En Caldas, se tienen registros de radiación solar en las estaciones de la red de Monitoreo de Caldas, cuya localización se resume en la Tabla 1.3.

Tabla 1.3: Localización geográfica de estaciones climáticas con registros de radiación solar. Fuente: propia

Estación	Fuente	Municipio	Altitud	Latitud	Longitud	Registros Desde	Registro Hasta	Años Registro
CUMANDAY	CORPOCALDAS	Villamaría	5072	4°53'37.7"	75°20'22.1"	02/09/2003	29/02/2012	9
CISNE-SANTA ISABEL	CORPOCALDAS	Villamaría	4812	4°49'24.6"	75°21'59.9"	13/09/2003	13/08/2012	9
NIZA	Alcaldía de Manizales	Manizales	2256	5°3'15.57"	75°28'47.67"	01/01/2010	31/12/2012	2
INGEOMINAS	Alcaldía de Manizales	Manizales	2226	5°4'17.2"	75°31'27.5"	01/01/2004	31/12/2012	8
YARUMOS	Alcaldía de Manizales	Manizales	2195	5°3'47.41"	75°28'52.63"	01/01/2004	31/12/2012	8
HOSPITAL DE CALDAS	Alcaldía de Manizales	Manizales	2183	5°3'46.6"	75°30'2.1"	14/12/2007	31/12/2012	5
POSGRADOS	Universidad Nacional	Manizales	2179	5°3'22.2"	75°29'32.3"	01/01/2002	31/12/2012	10
BOSQUES DEL NORTE	Alcaldía de Manizales	Manizales	2126	5°5'9.67"	75°29'31.83"	22/09/2006	31/12/2012	6
ENEÁ	Alcaldía de Manizales	Manizales	2092	5°1'43.1"	75°28'20"	01/01/2004	31/12/2012	8
EL CARMEN	Alcaldía de Manizales	Manizales	2112	5°3'4.44"	75°31'11.44"	15/02/2005	31/12/2012	7
ALCAZARES	Alcaldía de Manizales	Manizales	2057	5°4'0.2"	75°31'39.9"	01/06/2010	31/12/2012	2
NEIRA	CORPOCALDAS	Neira	2053	5°9'30.5"	75°31'1.7"	26/04/2012	08/10/2012	0
LA PALMA	Alcaldía de Manizales	Manizales	1967	5°5'45.5"	75°31'45.9"	01/11/2006	31/12/2012	6
HOSPITAL VILLAMARIA	CORPOCALDAS	Villamaría	1790	5°0'02'56.4"	75°31'6.9"	17/09/2010	31/12/2012	2
ARÁNJUEZ	Alcaldía de Manizales	Manizales	1915	5°2'34.2"	75°30'1.6"	26/01/2006	31/12/2012	6
SUPIA	CORPOCALDAS	Supía	1239	5°45'57"	75°65'699"	01/08/2012	31/12/2012	0
MANZANARES	CORPOCALDAS	Manzanares	1941	5°15'40.1"	75°09'14.2"	01/07/2012	31/12/2012	0
MARQUETALIA	CORPOCALDAS	Marquetalia	1591	5°17'56.6"	75°03'26.1"	01/07/2012	31/12/2012	0
MARULANDA	CORPOCALDAS	Marulanda	3110	5°18'03.3"	75°20'58.3"	05/09/2011	31/12/2012	1
PENSILVANIA	CORPOCALDAS	Pensilvania	1793	5°23'06.4"	75°08'05.9"	01/07/2012	31/12/2012	0
SALAMINA	CORPOCALDAS	Salamina	2078	5°23'25.5"	75°29'11.7"	01/04/2012	31/12/2012	0
RIO POZO	CORPOCALDAS	La Merced	840	5°26'43.7"	75°33'35.7"	01/04/2012	31/12/2012	0

Con base en los registros de radiación solar, se efectúa el análisis del comportamiento promedio del ciclo diurno y estacional, considerando que las estaciones de la red cuentan con información cada 15 minutos. El ciclo diurno para la estación Posgrados se presenta en la Figura 1.10 ; se evidencia un comportamiento con máximos valores hacia las 10:00 am.

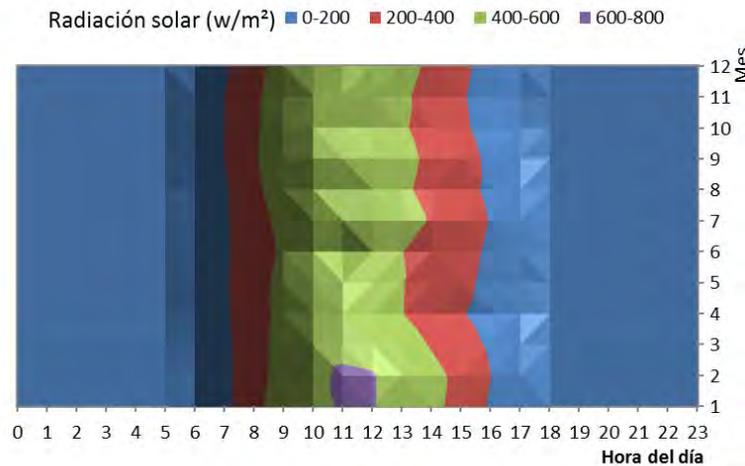


Figura 1.10: Ciclo diurno de la radiación solar en la estación Posgrados 2002-2010. Fuente: propia

La variación estacional de la radiación solar para la estación Posgrados se muestra en la Figura 1.11 si bien esta estación es la de mayor longitud de registro, es relativamente corta para obtener un patrón claro en su comportamiento.

Por tanto, para analizar el comportamiento para el período 1981-2010 se realiza una estimación de la radiación solar diaria a partir de los registros diarios de brillo solar y de la humedad relativa, empleando la Ecuación 1.2, propuesta por [Swartman and Ogulande, 1966]:

$$H = 464 + 265D - 248R \quad (1.2)$$

Donde:

$H$ : Radiación solar  $\langle \text{langleys}/\text{min} \rangle$  ( $\text{cal}/\text{cm}^2 \cdot \text{min}$ );

$D$ : Número de horas de brillo solar/12;

$R$ : Humedad Relativa



Figura 1.11: Comportamiento de la radiación en la estación Posgrados 2002-2010. Fuente: propia

A partir de la estimación de la Radiación Solar (RS) para el período 1981-2010, se construye la Figura 1.12 que muestra el comportamiento típico estacional bimodal, con mayores valores durante los períodos más secos enero-febrero y julio-agosto. Para la espacialización de la variable,

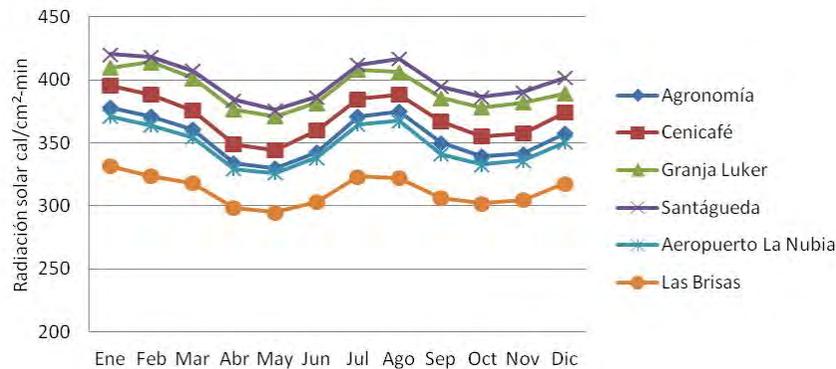


Figura 1.12: Comportamiento estacional de la radiación 1981-2010. Fuente: propia

se toma como base el modelo de correlación propuesto por Ocampo et al. [Ocampo et al., 2013], según la ecuación 1.3, con el cual se construye la Figura 1.13 que muestra la variabilidad claramente influenciada por la orografía, con mayores valores en los valles de los ríos Magdalena y Cauca y menores valores en las regiones montañosas de la cordillera central. En general, se presenta una mayor variación en la radiación que las estimaciones del IDEAM (2005b) para la zona.

$$RS < \frac{(kW - h)}{m^2} \geq 5.0723 - 0.0005H \quad (1.3)$$

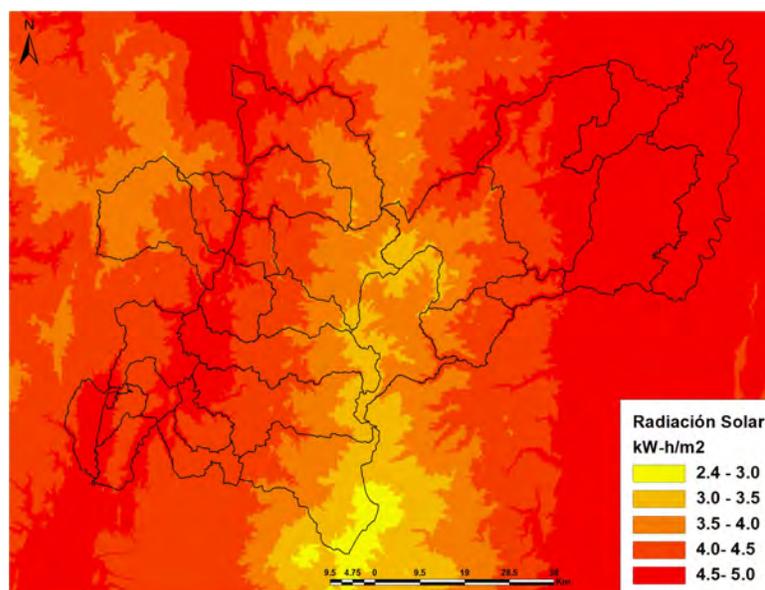


Figura 1.13: Mapa de radiación solar 1981-2010. Fuente: propia

La variabilidad interdecadal se evidencia en la zona de estudio, en el análisis realizado para las últimas tres décadas. La Figura 1.14 muestra el comportamiento para la estación Cenicafé. El análisis de tendencias en esta escala para la cuenca del río Chinchiná es descrito por [Ocampo, 2012] [Ocampo et al., 2012c] [Ocampo et al., 2012a].

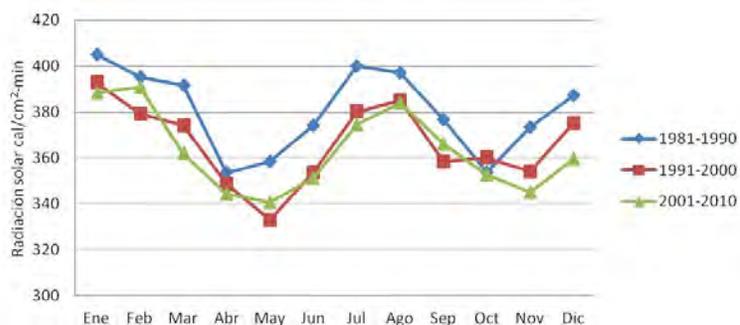


Figura 1.14: Comportamiento Interdecadal de la radiación solar Estación Cenicafé 1981-2010. Fuente: propia

### 1.4.3. Temperatura

La temperatura es la magnitud física que se caracteriza por el movimiento aleatorio medio de las moléculas en un cuerpo físico [OMM, 2012]. Las formas de expresarla son [OMM, 2008]:

- **Temperatura media:** es un indicador térmico que da el resultado del balance energético en la superficie terrestre; por tanto, es un indicador de cambio climático; para propósitos climatológicos es usada la temperatura media diaria de la cual se obtienen los valores mensuales y anuales.

- Temperatura máxima: es la mayor temperatura registrada durante el día. Es un indicador de confort térmico.
- Temperatura mínima: es el valor más bajo de la temperatura del aire registrada durante el día. Es un indicador de cambio climático y determinante de las condiciones de vida y productividad de la región.
- Amplitud o rango de temperatura: Diferencia entre la temperatura máxima y mínima. Tiene valor ecológico y diferente para los distintos cultivos y condiciona los ecosistemas.

La temperatura es proporcional a la radiación total incidente y es afectada por factores como la latitud, la altitud, las corrientes marinas, los vientos, la nubosidad, el relieve, la convección y la turbulencia del aire [Murthy., 2002]. La tasa de variación de la temperatura con la altitud, es el gradiente vertical de la atmósfera, el cual es función del contenido de vapor de agua [Monsalve, 2008]. En la troposfera el decrecimiento de temperatura fluctúa entre 6,0 a 6,5°C por cada km; en Colombia [Cortés, 1989], la variación es 5,53°C/km, en promedio; sin embargo, no es homogénea en el territorio; en la región Andina es 6,13°C/km [Eslava, 1992] [IDEAM, 2010]. Para la región andina, [Eslava, 1992] Eslava (1992) definió los siguientes modelos de regresión entre la temperatura media ( $T$ , °C) y la altitud ( $H$ , m):

$$\text{Magdalena} : T = 29.6 - 0.00624H \quad (1.4)$$

$$\text{Cauca} : T = 29.4 - 0.00617H \quad (1.5)$$

Chaves y Jaramillo [Chaves and Jaramillo., 1998] regionalizaron la relación para Colombia usando la información de 1002 estaciones, para la región Andina se tiene la siguiente correlación de la temperatura con la altitud:

$$\text{RegionAndina} : T = 29.42 - 0.0061H \quad (1.6)$$

Aunque la temperatura media anual es indicador climático importante, puede variar notablemente de un día a otro y a lo largo del año debido a la variabilidad del clima [OMM, 2013b]. Las temperaturas medias anuales en Colombia están por encima de 24°C. En la región Andina, se presentan los pisos térmicos, causados por la disminución de la temperatura con el aumento de la altura sobre el nivel del mar. En los valles de los principales ríos como el Magdalena y el Cauca, se registran temperaturas entre 24 y 28°C, mientras que en los altiplanos se presentan valores bajos, alrededor de 12 a 16°C. En la alta montaña se alcanzan temperaturas inferiores a 8°C [IDEAM, 2005a].

En el valle del río Cauca se presentan dos épocas “cálidas” y dos temporadas “frías”. La primera época “cálida” abarca los meses de febrero y marzo. La segunda temporada “cálida” es menos definida que la primera y abarca los meses de julio y agosto. La amplitud anual de la temperatura media del aire es de 1.5°C en esta región (Poveda, 2001).

Según el IDEAM (2005), se evidencian todos los pisos térmicos en el departamento de Caldas, desde cálido (mayor a 24°C) en los valles del río Magdalena, hasta nival (menor a 1.5°C) en el Parque Nacional Natural Los Nevados. En Caldas, la temperatura del aire es monitoreada con 48 estaciones de diferentes entidades que se presentan en la Tabla 1.4. CORPOCALDAS está implementando la red de monitoreo del departamento y cuenta con 18 estaciones en operación; esta red será una excelente herramienta futura para la estimación de la variabilidad espacio-temporal en la región.

Tabla 1.4: Localización geográfica de estaciones climáticas con registros de temperatura. Fuente: propia

Estación	Fuente	Municipio	Tipo	Altitud	Latitud	Longitud	Registros Desde	Registro Hasta	Años Registro
CUMANDAY	CORPOCALDAS	Villamaría	Climatológica	5072	4,89	-75,34	02/09/2003	29/02/2012	9
CISNE-SANTA ISABEL	CORPOCALDAS	Villamaría	Climatológica	4812	4,82	-75,37	13/09/2003	13/08/2012	9
NEREIDAS	CORPOCALDAS	Villamaría	Aforo	4450	4,88	-75,36	27/01/2004	13/08/2012	8
MOLINOS	CORPOCALDAS	Villamaría	Aforo	4240	4,91	-75,37	27/01/2004	28/02/2012	8
RIO CLARO	CORPOCALDAS	Villamaría	Aforo	4178	4,85	-75,37	24/05/2011	31/12/2012	1
LAS BRISAS	IDEAM	Villamaría	Meteorológica esp.	4150	4,90	-75,35	12/10/1981	30/10/2012	31
LETRAS	CENICAFE	Manizales	Climatológica	3684	5,05	-75,33	01/01/2003	31/12/2011	8
MARULANDA	CORPOCALDAS	Marulanda	Climatológica	3110	5,30	-54,03	05/09/2011	31/12/2012	1
QUEBRADA MANIZALES TESORITO	CORPOCALDAS	Manizales	Aforo	2314	5,04	-75,43	17/12/2010	08/10/2012	2
TESORITO FCA	IDEAM	Manizales	Meteorológica esp.	2290	5,03	-75,43	01/09/1993	30/03/2011	18
NIZA	Alcaldía de Manizales	Manizales	Climatológica	2256	5,05	-75,48	01/01/2010	31/12/2012	2
INGEOMINAS	Alcaldía de Manizales	Manizales	Climatológica	2226	5,07	-75,52	01/01/2004	31/12/2012	8
YARUMOS	Alcaldía de Manizales	Manizales	Climatológica	2195	5,06	-75,48	01/01/2004	31/12/2012	8
HOSPITAL DE CALDAS	Alcaldía de Manizales	Manizales	Climatológica	2183	5,06	-75,50	14/12/2007	31/12/2012	5
POSGRADOS	Universidad Nacional	Manizales	Climatológica	2179	5,06	-75,49	01/01/2002	31/12/2012	10
BOSQUES DEL NORTE	Alcaldía de Manizales	Manizales	Climatológica	2126	5,09	-75,49	22/09/2006	31/12/2012	6
EL CARMEN	Alcaldía de Manizales	Manizales	Climatológica	2112	5,06	-75,52	15/02/2005	31/12/2012	7
ENEA	Alcaldía de Manizales	Manizales	Climatológica	2092	5,03	-75,47	01/01/2004	31/12/2012	8
AGRONOMIA	CENICAFE	Manizales	Climatológica	2088	5,05	-75,50	01/01/1956	31/12/2011	55
EMAS	EMAS	Manizales	Climatológica	2060	5,08	-75,51	01/01/1997	30/12/2012	15
AP TO LA NUBIA	IDEAM	Manizales	Meteorológica esp.	2058	5,03	-75,47	01/05/1981	30/07/2012	31
ALCAZARES	Alcaldía de Manizales	Manizales	Climatológica	2057	5,07	-75,53	01/06/2010	31/12/2012	2
NEIRA	CORPOCALDAS	Neira	Climatológica	2053	5,16	-75,52	26/04/2012	08/10/2012	0
LA PALMA	Alcaldía de Manizales	Manizales	Climatológica	1967	5,10	-75,53	01/11/2006	31/12/2012	6
QUEBRADA OLIVARES	CORPOCALDAS	Manizales	Climatológica	1960	5,06	-75,48	07/09/2010	08/10/2012	2
MANZANARES	CORPOCALDAS	Manzanares	Climatológica	1941	5,26	-75,15	01/07/2012	31/12/2012	0
BELLAVISTA	IDEAM	Anserma	Climatológica	1932	5,27	-75,78	01/12/1978	31/12/2011	33
ARANJUEZ	Alcaldía de Manizales	Manizales	Climatológica	1915	5,04	-75,50	26/01/2006	31/12/2012	6
RIO GUACAICA-CHEC	CORPOCALDAS	Neira	Climatológica	1912	5,14	-75,50	10/08/2011	31/12/2012	1
PENSILVANIA	CORPOCALDAS	Pensilvania	Climatológica	1793	5,39	-75,13	01/07/2012	31/12/2012	0
HOSPITAL VILLAMARIA	CORPOCALDAS	Villamaría	Climatológica	1790	5,05	-75,52	17/09/2010	31/12/2012	2
MARQUETALIA	CORPOCALDAS	Marquetalia	Climatológica	1591	5,30	-75,06	01/07/2012	31/12/2012	0
LLANADAS	CENICAFE	Manzanares	Climatológica	1463	5,20	-75,13	01/01/1979	31/12/1998	19
PLANALTO	CENICAFE	Manizales	Climatológica	1413	4,98	-75,58	01/01/1994	31/12/2011	17
SANTA HELENA	CENICAFE	Marquetalia	Climatológica	1395	5,32	-75,00	01/01/1981	31/12/2011	30
NARANJAL	CENICAFE	Chinchiná	Climatológica	1381	4,97	-75,65	01/01/1956	31/12/2011	55
LA ILUSION	CENICAFE	Filadelfia	Helioplúviográfica	1348	5,22	-75,58	01/04/2003	31/12/2011	8
CENICAFE	CENICAFE	Chinchiná	Climatológica	1310	4,98	-75,60	01/01/1950	31/12/2011	61
RAFAEL ESCOBAR	CENICAFE	Supía	Climatológica	1307	5,45	-75,63	01/01/1976	31/12/2004	28
SUPIA	CORPOCALDAS	Supía	Climatológica	1239	5,46	-75,66	01/08/2012	31/12/2012	0
EL ALGARROBO	CENICAFE	Manizales	Climatológica	1202	5,07	-75,58	01/01/1993	31/12/2007	14
GRANJA LÜKER	CENICAFE	Palestina	Climatológica	1031	5,07	-75,68	01/01/1972	31/12/2011	39
SANTAGUEDA	CENICAFE	Palestina	Climatológica	1026	5,07	-75,67	01/05/1965	31/12/2011	46
AP TO SANTA-GUEDA	IDEAM	Palestina	Meteorológica esp.	933	5,08	-75,68	01/02/1979	31/12/1989	10
RIO GUACAICA-EL JORDAN	CORPOCALDAS	Neira	Aforo	929	5,11	-75,66	05/08/2011	31/12/2012	1
SALAMINA	CORPOCALDAS	Salamina	Climatológica	2078	5,39	-75,49	01/04/2012	31/12/2012	0
RIO POZO	CORPOCALDAS	La Merced	Climatológica	840	5,45	-74,44	01/04/2012	31/12/2012	0

El ciclo diurno de la temperatura media se presenta en la Figura 1.15, para la estación Posgrados, cuyo comportamiento es el esperado, siendo mayor durante el día que en la noche. La temperatura media oscila entre 13 a 23°C en la estación, con un valor promedio de 17°C. La amplitud térmica es de 8.5°C en promedio, con mayores valores en los meses de enero y febrero, 9.9 °C y 9.5°C, respectivamente. A escala horaria, los mayores valores se registran a las 12 del día y los menores a las 5 de la mañana. El coeficiente de variación a escala horaria es del 3% en promedio; la mayor variabilidad se presenta entre las 4 y 5 de la tarde.

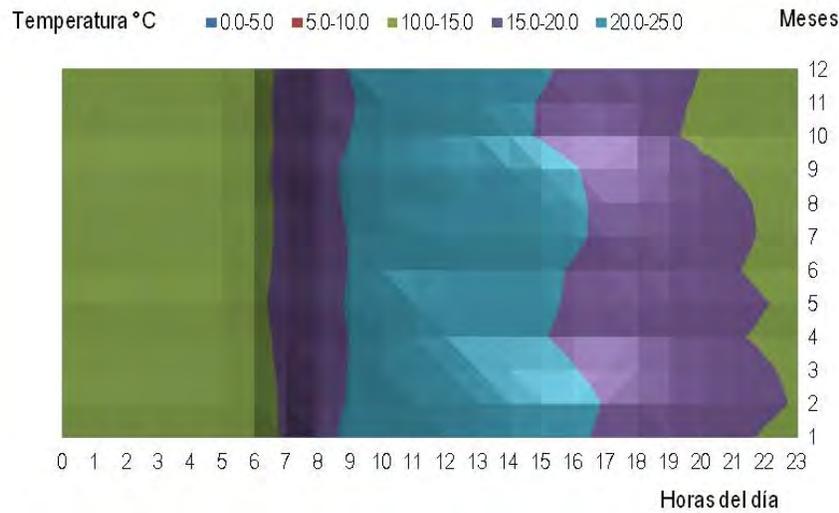


Figura 1.15: Ciclo diurno de la Temperatura Media Estación Posgrados 2002-2010. Fuente: propia

A partir de la información de las estaciones con registros superiores a 25 años, se realiza el análisis del comportamiento estacional de la temperatura media el cual se resumen en la Figura 1.16, que exhibe el comportamiento bimodal típico, con mayores valores en el mes de agosto y a principios de año.

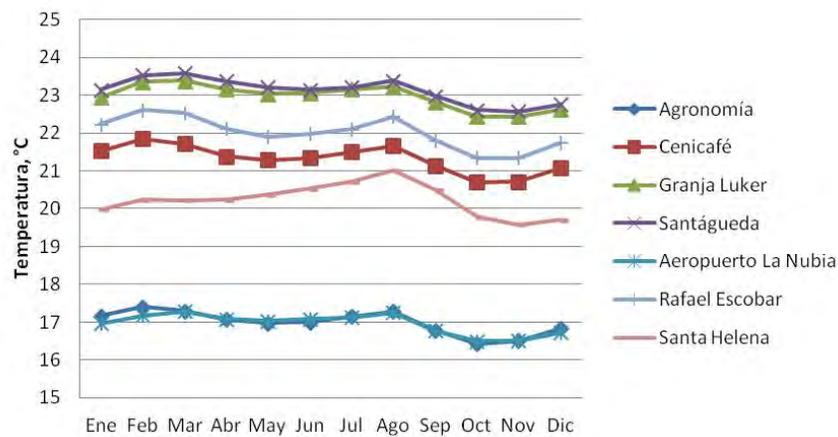


Figura 1.16: Comportamiento estacional Temperatura Media 1981-2010. Fuente: propia

El análisis del comportamiento promedio de la temperatura media para el período de análisis 1981-2010 se realiza a partir de la espacialización de la información; para tal efecto, se parte de los modelos recientes de correlación de la temperatura con la altitud para la zona, propuestos por Ocampo et al. [Ocampo et al., 2013] (2013) que se presentan en la Ecuación 1.7. Este modelo indica una tasa de variación de la temperatura con la altitud de 6.4°C por km.

$$T_{media}(C) = 30.319 - 0.0064H \quad (1.7)$$

El mapa de distribución de la temperatura media anual se visualiza en la Figura 1.17 para el período 1981-2010; se aprecian mayores valores en los valles de los ríos Magdalena y Cauca; las más bajas temperaturas se presentan en la zona del parque Nacional Natural Los Nevados.

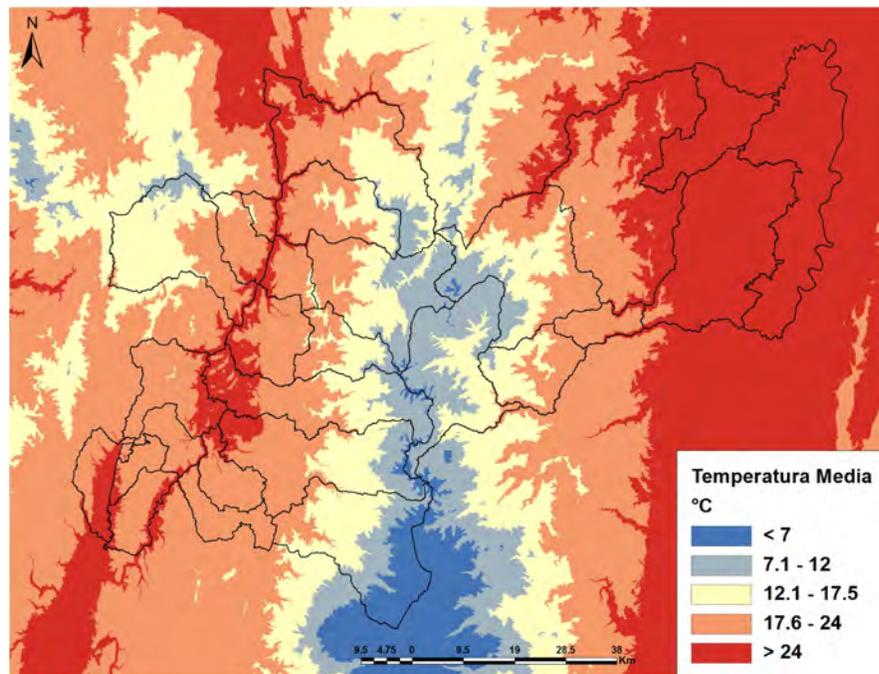


Figura 1.17: Mapa de Temperatura Media 1981-2010. Fuente: propia

Para la estimación de la distribución espacial de las temperaturas máximas promedio, se toma como base el modelo de correlación propuesto en [Ocampo et al., 2013] que se presentan en la Ecuación 1.8. A partir de esta correlación y empleando el modelo de elevación digital del terreno, se elaboró el mapa de distribución de la temperatura máxima el cual se presenta en la Figura 1.18, promedio multianual para el período 1981-2010.

$$T_{maxima}(C) = 34.498 - 0.0056H \quad (1.8)$$

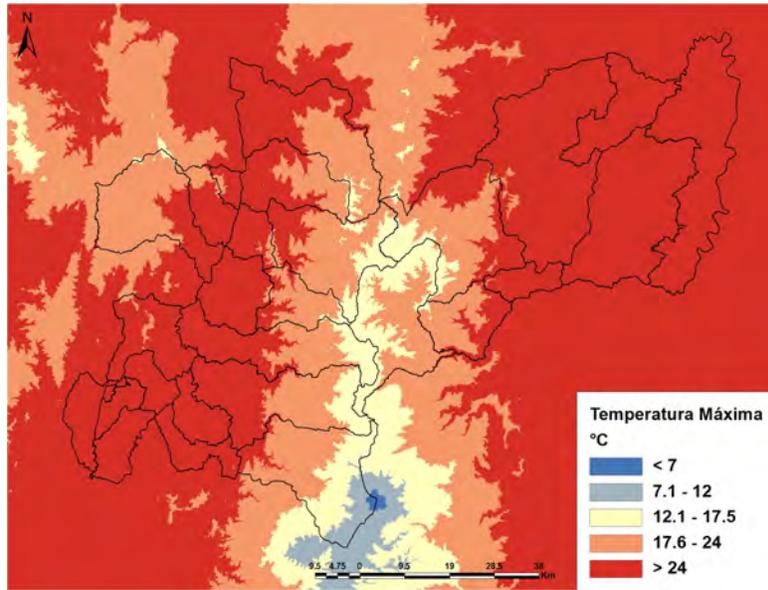


Figura 1.18: Mapa de Temperatura Máxima 1981-2010. Fuente: propia

Las temperaturas mínimas tienen un comportamiento lineal inversamente proporcional con la altitud. Para la espacialización de la variable, se toma como base la correlación propuesta por Ocampo et al. [Ocampo et al., 2013], según la ecuación 1.9. La Figura 1.19 presenta el mapa de distribución espacial de las temperaturas mínimas promedio multianual para el período en evaluación

$$T_{\text{mínima}}(C) = 24.452 - 0.0058H \quad (1.9)$$

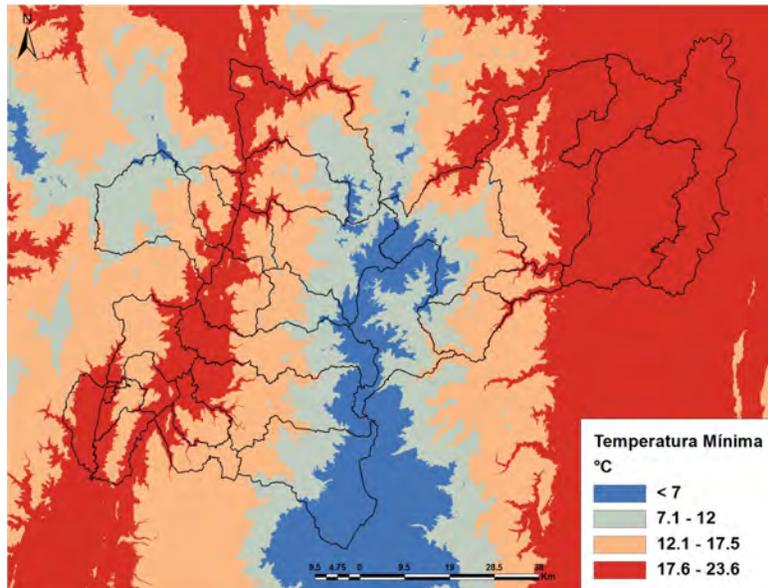


Figura 1.19: Mapa de Temperatura Mínima 1981-2010. Fuente: propia

La Figura 1.20 presenta el mapa de amplitud o rango de temperatura, promedio anual el cual fluctúa entre 10 y 11°C. La distribución por pisos térmicos se presenta al final del capítulo en la clasificación climática. En la escala interanual se evidencian los efectos de los eventos ENSO y sus fases cálida y fría. Para la cuenca del río Chinchiná por ejemplo, los eventos más intensos se presentaron en 1997-1998 con un incremento de 0,92°C en la temperatura y en 2009-2010 con 0,86°C; mientras que para la Niña se presentaron en el 1988-1989 y 2010-2011 con reducciones promedio de la temperatura media de 0,57°C [Ocampo, 2012] [Ocampo et al., 2012c] [Ocampo et al., 2012a].

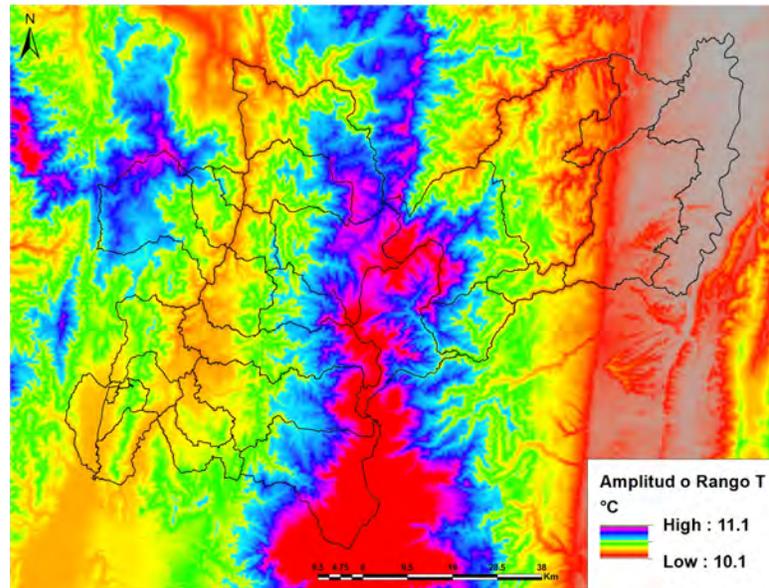


Figura 1.20: Mapa de Temperatura Amplitud o Rango de Temperatura 1981-2010. Fuente: propia

A escala interdecadal, se presentan tendencias crecientes tanto para las temperaturas medias, como para las máximas y mínimas en las estaciones de la zona objeto de estudio, lo cual permite evidenciar las señales de cambio climático en la región. La Figura 1.21 muestra el comportamiento para la estación Agronomía, donde la temperatura es más alta en la última década. El análisis de tendencias de la temperatura en esta escala es descrito por [Ocampo, 2012] [Ocampo et al., 2012c] [Ocampo et al., 2012a].

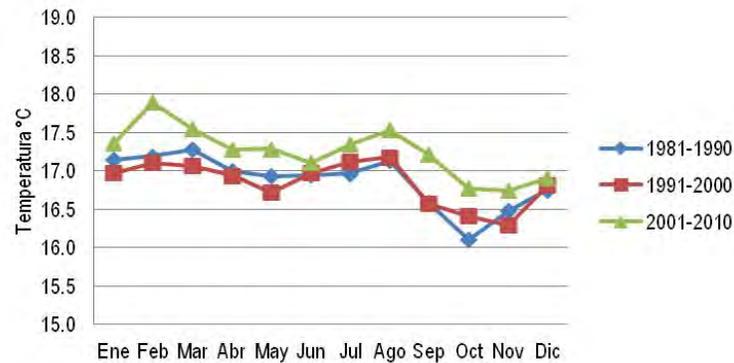


Figura 1.21: Comportamiento Interdecadal de la Temperatura Media Estación Agronomía. Fuente: propia

#### 1.4.4. Humedad relativa

La humedad atmosférica se define como el contenido de vapor de agua en la atmósfera [Silva, 1998]; puede expresarse de diferentes formas: presión de saturación, déficit de saturación, humedad absoluta, humedad específica y humedad relativa [Linacre., 1992]. La humedad relativa es la relación porcentual entre la humedad en un espacio y la cantidad que ese volumen podría contener si estuviera saturado. Se emplean diferentes instrumentos para su medición como psicrómetros, higrómetros, higrógrafos y sensores infrarrojos [IDEAM, 2005a]. En las zonas tropicales, donde las variaciones de la temperatura durante el día son generalmente grandes, la humedad relativa cambia considerablemente en el transcurso del día; es máxima en la noche y mínima en el día; los valores más altos se presentan desde las 20 horas hasta las 6 horas; los intermedios en las primeras horas de la mañana y de la noche y los más bajos al mediodía y al comienzo de la tarde [Eslava et al., 1986e] [Eslava et al., 1986c] [Eslava et al., 1986b] [Eslava et al., 1986d] [Eslava et al., 1986a]. Es afectada por el relieve, el cual se comporta como una barrera que frena el paso de masas húmedas o aumenta su presencia [Jaramillo, 2005]. En la región Andina colombiana se presenta esta influencia. En esta zona se registran humedades relativas anuales entre 66 y 87 % [IDEAM, 2005a]. El comportamiento a través del año es similar en toda la región Andina, de tipo bimodal con dos picos máximos entre abril y mayo y en los meses de noviembre y diciembre y los mínimos entre febrero y marzo y entre julio y septiembre [IDEAM, 2005a]. De acuerdo con lo reportado por el IDEAM [IDEAM, 2005a], el promedio multianual de la humedad relativa del aire para el departamento de Caldas oscila entre 75-90 %, con mayores valores en la zona de páramos y menores en los valles de los ríos Cauca, Risaralda y Magdalena. Para el análisis del comportamiento de la variable humedad relativa en el departamento de Caldas, se parte de los registros de las estaciones presenten en la zona que son básicamente las mismas que registran temperatura, cuya información se muestra en la Tabla 1.4. El análisis del ciclo diurno se presenta para la estación Posgrados que se aprecia en la Figura 1.22. La humedad relativa – HR- fluctúa entre 60 y 95 %, con valores promedio del 82.7 %. Los mayores valores se registran entre las 4 y 5 de la mañana y los menores hacia el mediodía. A escala horaria, el coeficiente de variación es del 3 %; la mayor variación se presenta hacia las 5 de la tarde.

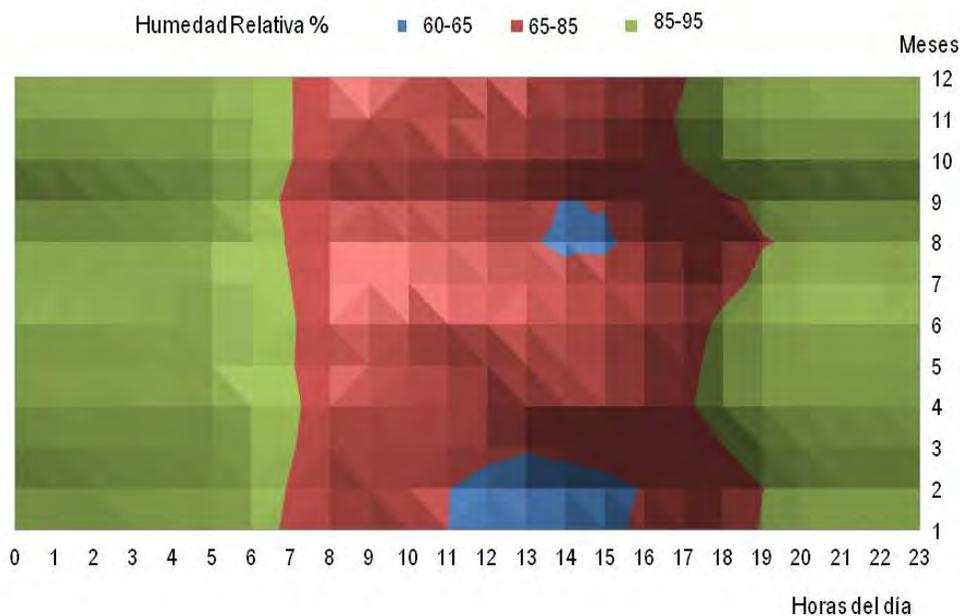


Figura 1.22: Ciclo diario de la humedad relativa estación Posgrados 2002-2010. Fuente: propia

El análisis estacional de la humedad relativa para las diferentes estaciones que se resumen en la Figura 1.23, donde se confirma el comportamiento típico bimodal para la zona, con menores valores en la temporada seca de mitad de año. En la estación Santa Helena, ubicada en el oriente de Caldas, es más marcada la estacionalidad. Para elaborar el mapa de distribución de la humedad relativa (HR) en el departamento de Caldas, se parte del modelo de elevación digital y de la correlación de la Humedad Relativa (HR) con la altitud (H) propuesto por Ocampo et al. [Ocampo et al., 2012a] que se presenta en la ecuación 1.10:

$$HR = 74.08 + 0.0028H \quad (1.10)$$

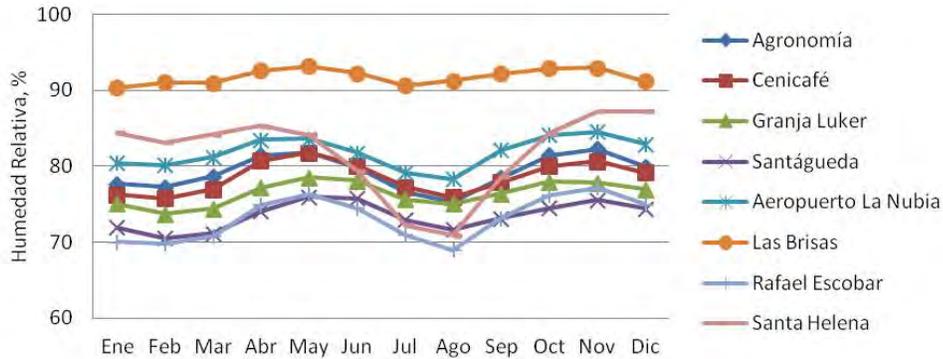


Figura 1.23: Comportamiento estacional de la humedad relativa 1981-2010. Fuente: propia

El mapa de distribución espacial de la humedad relativa se presenta en la Figura 1.24. Es máxima en la zona del parque Nacional Natural Los Nevados y mínima en los valles de los ríos Magdalena, Cauca y Risaralda. Los rangos de variación de la humedad relativa para el período 1981-2010, están acordes a las estimaciones del IDEAM [IDEAM, 2005a]. Sin embargo, se requiere contar con mayores registros en el oriente de Caldas, en especial en la Selva de Florencia, para poder validar en el modelo de regresión para esta zona.

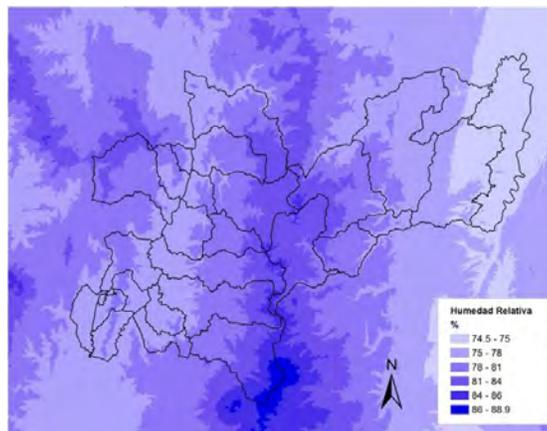


Figura 1.24: Mapa humedad relativa media anual 1981-2010. Fuente: propia

A escala interdecadal se observan tendencias crecientes en las estaciones, en especial para la última década, como se aprecia en la Figura 1.25 para la estación CENICAFÉ. El análisis de tendencias de esta variable se presenta en detalle en Ocampo [Ocampo, 2012] y en Ocampo et al. [Ocampo et al., 2012c] [Ocampo et al., 2012a].

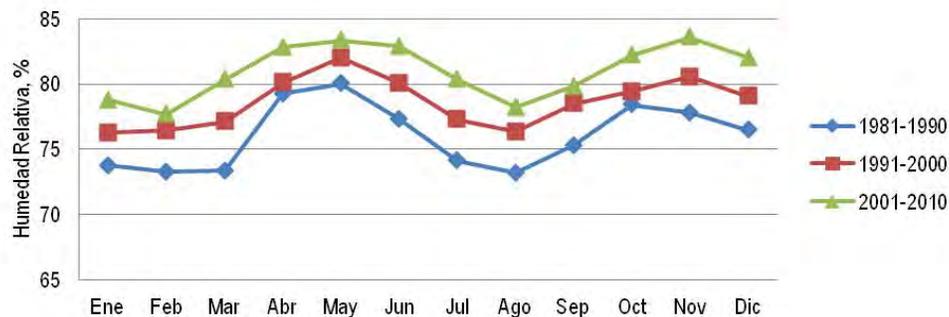


Figura 1.25: Comportamiento interdecadal de la humedad relativa estación CENICAFÉ. Fuente: propia

### 1.4.5. Precipitación

La precipitación corresponde a los productos líquidos o sólidos de la condensación de vapor de agua que caen de las nubes o son depositados desde el aire sobre la tierra, e incluyen lluvia, granizo, nieve, rocío, escarcha y niebla [OMM, 2008]. Para comprender el clima de una región es fundamental comprender el comportamiento y los patrones de precipitación, la cual puede ser descrita en los términos de intensidad, duración, profundidad, frecuencia, distribución temporal y espacial [Ponce, 1989]. La cantidad total de precipitación que llega al suelo en un determinado período de tiempo se expresa en términos de profundidad vertical de agua para la cual cubre un proyección horizontal de la superficie de la tierra [OMM, 2008]; las unidades son mm, que equivalen a  $1 \text{ lt}/\text{m}^2$ . Las características climáticas de una región pueden definirse según la precipitación total anual [OMM, 2008]:

- Poca: < 200 mm
- Escasa: 200 – 500 mm
- Normal: 500 – 1000 mm
- Abundante: 1000 – 2000 mm
- Muy abundante: > 2000 mm

Según la intensidad de lluvia, se han definidos las siguientes clases en las características climáticas de una región [OMM, 2008]:

- Débiles: < 2 mm/h
- Moderadas: >2 y < 15 mm/h
- Fuertes: >15 y < 30 mm/h
- Muy fuertes: >15 y < 30 mm/h
- Torrenciales: >60 mm/h

El conocimiento de la distribución de la precipitación permite conocer la disponibilidad de agua de una región y mejorar la planificación agrícola; es además fundamental para el diseño de obras hidráulicas, para la gestión del riesgo y para la ordenación del territorio. El número de días con lluvia es un indicador de gran importancia para diferentes aplicaciones en especial agricultura y transporte. Mientras que la lluvia acumulada es un indicador de utilidad para conocer la

susceptibilidad al deslizamiento en una zona específica y permite generar niveles de alerta según los valores obtenidos, tanto de los días consecutivos con o sin lluvia como el parámetro A25 que es la lluvia acumulada de los 25 días antecedentes [Vélez et al., 2012]. En los trópicos, la convección es el mecanismo más importante de transporte de calor y el principal proceso de lluvia [Poveda, 2004]. Las nubes convectivas en los trópicos hacen disminuir la amplitud diurna de la temperatura por la ausencia de radiación solar directa [IDEAM, 2005a].

Por otra parte, la orografía tiene una decisiva influencia en la distribución vertical de las lluvias [Oster., 1979] [Mejía et al., 1999] [Pabón et al., 1998] [Pabón et al., 2001] [Poveda et al., 2002] [Poveda, 2004] [Jaramillo, 2005]. En los trópicos se tienen diferentes niveles de condensación según la altitud, que van desde el primer nivel alrededor de los 1800 msnm, el segundo nivel a los 2700 msnm y el tercer nivel entre los 3200-3500 msnm [Mora and Sturmo., 1994].

La alta variabilidad de la lluvia en Colombia según Poveda et al. [Poveda et al., 2002] está asociada con los mecanismos físicos que determinan la lluvia como son el ciclo diurno del brillo solar, los procesos de la interacción suelo-atmósfera, la dinámica diurna de los sistemas convectivos de mesoescala y la circulación de los vientos locales dentro de los valles y vertientes Andinas.

Para Colombia, el promedio de precipitación medial anual es de 3.000 mm. Las máximas lluvias se reportan en la región Pacífica con valores anuales entre 3000-12000 mm. En la Amazonía las lluvias son muy abundantes entre 3000 – 4500 mm. En la región Andina se presentan lluvias abundantes hasta 2000 mm a lo largo de la cordillera Oriental y en los Valles del Alto Magdalena y Alto Cauca y núcleos máximos (3000-5000 mm) en las cuencas del Magdalena Medio y Medio Cauca. En la Orinoquía también se tiene diversidad: hasta 6000 mm en el piedemonte llanero; 2000-3000 mm en la parte central y oriental y menores a 1500 mm en el extremo norte de Arauca. En la región Caribe se reportan lluvias entre 500-2000 mm; las menores lluvias se presentan en la Alta Guajira con valores inferiores a 500 mm [IDEAM, 2005a].

Los núcleos de mayor frecuencia en la precipitación están localizados en el centro y occidente de Antioquia, sur de Santander, oriente de Cundinamarca y Boyacá y eje cafetero con número de días con lluvia entre 200- 250 [IDEAM, 2005a]. Las altas cantidades de lluvia en estas zonas no se deben tanto al valor máximo de las lluvias diarias sino más bien a la alta ocurrencia de los días con lluvia [Oster., 1979] [Jaramillo, 2005].

Los valores más altos de lluvias máximas en 24 horas se registran en la región Norte del litoral Pacífico en las cuencas de los ríos Baudó, San Juan y Atrato con valores cercanos a 500 mm [Eslava, 1994]. El ciclo anual se caracteriza por la ocurrencia de los picos máximos de precipitación según el efecto del movimiento de la ZCIT. Se presenta un régimen bimodal entre los 4° y 7° de latitud Norte, en la mayor parte de la región Andina y el Caribe, con excepción de las cuencas del bajo Nechí y del río Sinú y en sectores de la vertiente oriental de la cordillera central a la altura de Samaná, Caldas y para latitudes inferiores a 3° Norte, donde se evidencia un régimen monomodal. En las regiones con influencia directa de los Alisios del Norte como la Llanura del Caribe, La Guajira y la región del río Catatumbo se presenta también un régimen monomodal; al igual que en la Orinoquía y Amazonía. En la región Pacífica no se presenta una tendencia definida [Pabón et al., 2001] [Jaramillo, 2005] [IDEAM, 2005a].

La región Andina posee una gran diversidad pluviométrica, con lluvias anuales de 2000-5000 mm [IDEAM, 2005b]. La distribución intra-anual de la precipitación es bimodal, con dos temporadas secas en enero - febrero y julio – agosto; mientras que los meses de exceso hídrico son abril-mayo y octubre-noviembre [Jaramillo, 2005]. En la región Andina colombiana, la precipitación está determinada por los sistemas de nubosidad, condicionados por la altitud, la orientación de las montañas y la actividad convectiva del lugar [Jaramillo, 2005]. La ZCIT ejerce un fuerte control en la distribución anual de las lluvias [León et al., 2000] [Pabón et al., 2001] [Poveda et al., 2002] [Poveda, 2004] [Martínez et al., 2013] y su desplazamiento produce un ciclo bimodal de precipitación que permite diferenciar dos períodos de lluvias durante abril a mayo y octubre a noviembre y dos períodos de menor lluvia de diciembre a marzo y junio a agosto [Poveda, 2004] [Jaramillo and Ramírez, 2009].

Según Ramírez y Jaramillo [Jaramillo and Ramírez, 2009], en esta región los meses con baja precipitación, que no tienen influencia de la ZCIT, -diciembre, enero, febrero, junio, julio, agosto- muestran una relación estadísticamente significativa y con mayores coeficientes de correlación entre la lluvia mensual y el índice oceánico El Niño- ONI- .

Los factores locales como las variaciones en la fisiografía influyen en la precipitación [Barry., 1992] [Eslava, 1993]. Las montañas de la zona Andina originan vientos locales de valle-montaña durante el día y de montaña-valle en la noche que afectan la dinámica de la precipitación y fortalecen la interacción de los componentes del sistema atmosférico [Jaramillo, 2005] [IDEAM, 2005a]

Mecanismos físicos complejos como la forma, tamaño y rugosidad de las montañas, la dirección y distancias de las fuentes de humedad, velocidad del viento y los procesos de tipo convectivo de las masas de aire influyen en los patrones de precipitación en la región andina [Jaramillo, 2005].

En el departamento de Caldas, la lluvia exhibe la misma complejidad que para la región Andina; el ciclo anual está dominado por el movimiento de la ZCIT; es influenciada por ENSO y por los factores locales. Según el IDEAM [IDEAM, 2005a], fluctúa entre 1500- 7000 mm anuales, con óptimos pluviométricos hacia el oriente de Caldas y menores valores en la región del Parque Nacional Natural los Nevados.

Para determinar el comportamiento en el ciclo anual en Caldas, se parte de la información reportada por las estaciones principales, heliográficas, pluviográficas y pluviométricas presentes en la zona de la red de Monitoreo de Caldas de CORPOCALDAS, IDEAM, CENICAFÉ, CHEC, GENSA, ISAGEN y la Universidad Nacional de Colombia sede Manizales.

El comportamiento del ciclo diurno de precipitación se muestra en la Figura 1.26 para la estación Posgrados. El coeficiente de variación es del 62 %, que da una idea de la gran variabilidad de la lluvia en esta escala. Los valores más altos de precipitación se registran entre las 2:00 pm y las 4:00 pm; mientras que los más bajos se reportan entre las 9 y 10 am. Entre los meses de octubre a diciembre se tienen los valores más altos de precipitación, con mayor frecuencia en horas de la tarde.

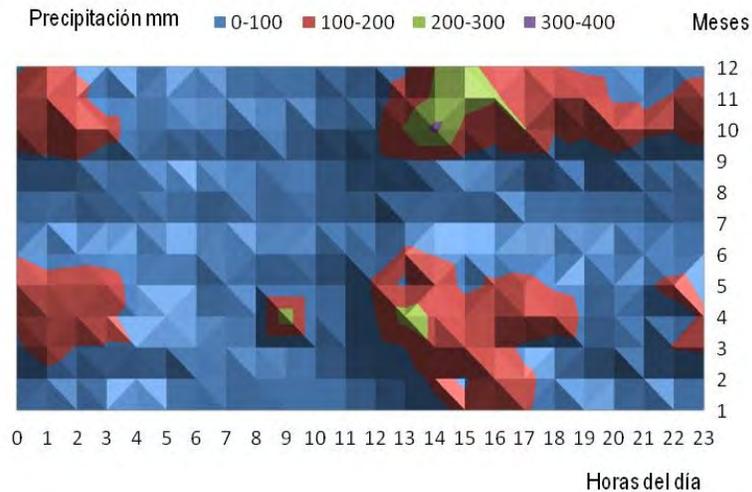


Figura 1.26: Comportamiento del ciclo diurno de la precipitación estación Posgrados 2002-2010. Fuente: propia

La variación estacional de la precipitación en diferentes estaciones de la cuenca se presenta en la Figura 1.27 para la región Centro Sur y el Oriente del departamento de Caldas; se evidencia el comportamiento típico bimodal con mayores valores en los meses de abril-mayo y octubre-noviembre.

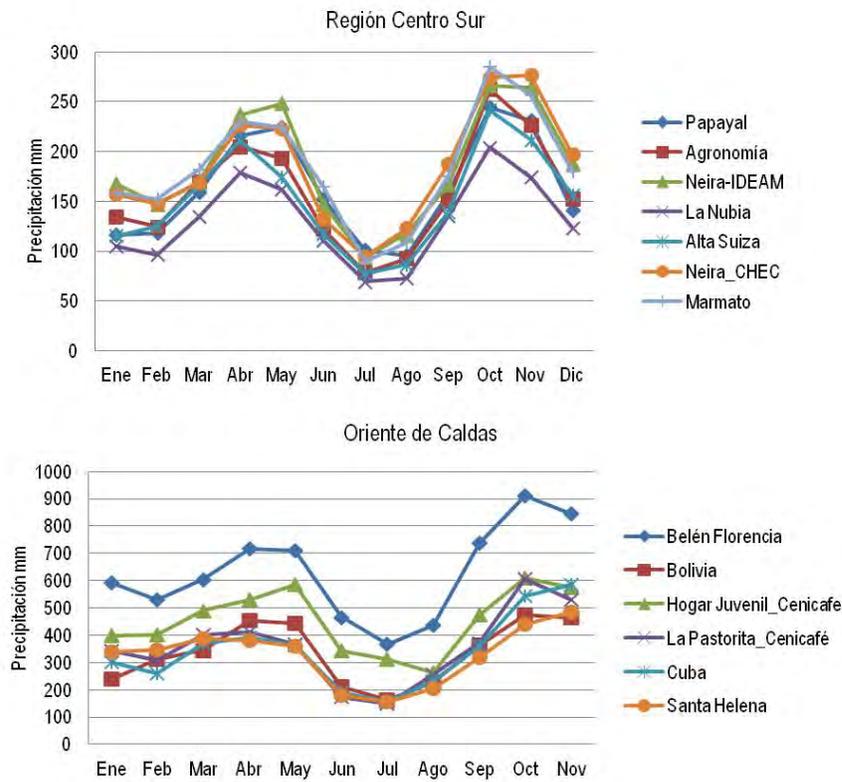


Figura 1.27: Comportamiento estacional de la precipitación media 1981-2010. Fuente: propia

El mapa de precipitación media anual se muestra en la Figura 1.28 que permite concluir que la precipitación es abundante en la zona y que se presenta gran variabilidad espacial, las regiones con menores valores de precipitación son las del Parque Nacional Natural los Nevados; mientras que las más lluviosas se encuentran en la Selva de Florencia, donde se presenta el óptimo pluviométrico para el departamento de Caldas.

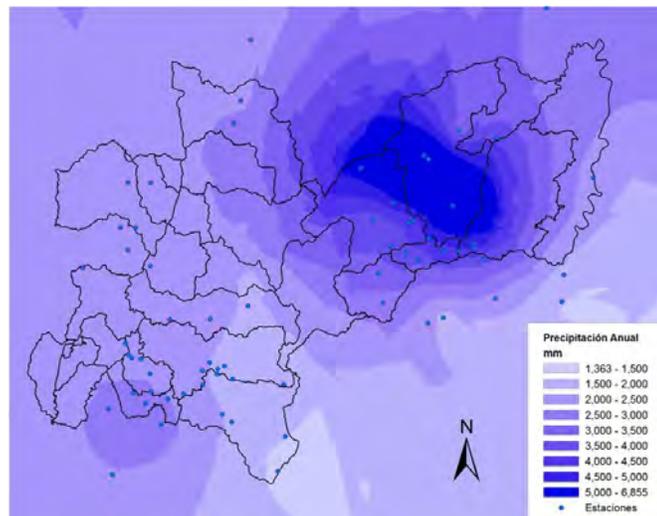


Figura 1.28: Mapa de precipitación media anual 1981-2010. Fuente: propia

En la escala inter-anual se tiene influencia del fenómeno ENSO con incremento de las lluvias durante los episodios fríos –La Niña- y reducción en la fase cálida –El Niño. Durante el período 1981-2011, se evidencia intensificación de La Niña, en especial durante los episodios 2007-2008 y 2010-2011; las anomalías en la precipitación fluctúan entre el 10 al 47 %, con valores extremos de aumento en la precipitación hasta del 140 %. Con respecto al Niño, los eventos más intensos corresponden a los períodos 2009-2010, 1991-1992, con reducciones en la lluvia en promedio del 26 % y 30 %, respectivamente. El análisis de la influencia del ENSO para la cuenca del río Chinchiná se describe por [Ocampo, 2012] [Ocampo et al., 2012c] [Ocampo et al., 2012a] [Ocampo et al., 2012d] [Ocampo et al., 2012b].

En la escala interdecadal se presentan aumento en algunas estaciones, reducciones en otras y también se observan tendencias monótonas sin cambio. El análisis de varianza por pruebas no paramétricas de Kruskal Wallis, para las series diarias y mensuales y por pruebas paramétricas ANOVA para los registros anuales por períodos interdecadales, permite concluir que hay diferencias estadísticas significativas ( $\alpha=0,10$ ) entre los períodos como se observa en la Tabla 1.5 que resume la significancia estadística; se omiten los valores en las estaciones donde no se encontraron diferencias significativas. El cambio en la precipitación se evalúa basado en los medios diarios como aumento en el transcurso del tiempo (+) o disminución (-). En algunas estaciones hay oscilación y por tanto no se evidencia cambio (=).

Tabla 1.5: Significancia estadística análisis de varianza precipitación interdecadal. Fuente: propia

Estación	Diaria		Mensual		Anual	
	Kruskal Wallis significancia	Cambio Lluvia	Kruskal Wallis significancia	Cambio de Lluvia	Anova sig	Cambio de Lluvia
Las Brisas	0,032	-				
La Esperanza	0,000	-	0,065	-		
Papayal	0,000	=				
Agronomía	0,001	+				
Alta Suiza	0,000	+	0,001	+	0,001	+
Neira	0,000	+			0,077	+
Marmato	0,000	+	0,072	-		
Sancancio	0,000	+	0,027	+	0,100	+
La Uribe						
Java	0,000	+	0,026	+	0,058	+
El Recreo	0,000	+	0,003	+	0,057	+
Naranjal	0,002	+	0,025	+	0,008	+
Montevideo	0,000	-	0,004	-	0,005	-
La Selva	0,000	+				
Cenicafé	0,000	+				
Granja Lúker	0,000	+				
Santágueda	0,000	+	0,092	+		
Arauca	0,000	+				

#### 1.4.6. Velocidad y dirección del viento

El viento es la variable de estado del movimiento del aire; se produce por los gradientes de presión atmosférica originados por diferentes calentamientos en las superficies y por efecto de las fuerzas de fricción [Jaramillo, 2005]. A medida que el tiempo transcurre, la componente horizontal del movimiento del aire en un punto dado de la atmósfera o de la superficie terrestre varía en intensidad y en orientación de una manera irregular; se observan variaciones rápidas y amplias, que corresponden al comportamiento habitual del flujo del aire. Las variables físicas de mayor importancia en el monitoreo de este fenómeno natural son: la velocidad del viento y dirección del viento. La velocidad del viento es la relación de la distancia recorrida por el aire con respecto al tiempo. Según la OMM, la velocidad se mide a 2 y 10 metros de altura. Se emplean diversos instrumentos, pero el más común es el anemómetro.

Según la OMM [OMM, 2008], las velocidades del viento se expresan en la escala de Beaufort que se presenta en la Tabla 1.6.

Tabla 1.6: Escala de Beaufort – Velocidad del viento. Fuente: [OMM, 2008]

Número en la Escala de Beaufort	Descripción	Velocidad de viento a una altura de 10 m sobre el suelo (m/s)
0	Calma	0.0 - 0.2
1	Aire ligero	0.3 - 1.5
2	Brisa muy ligera	1.6 - 3.3
3	Brisa suave	3.4 - 5.4
4	Brisa moderada	5.5 - 7.9
5	Brisa fresca	8.0 - 10.7
6	Brisa fuerte	10.8 - 13.8
7	Cercano a vendaval	13.9 - 17.1
8	Vendaval	17.2 - 20.7
9	Vendaval fuerte	20.8 - 24.4
10	Tormenta	24.5 - 28.4
11	Tormenta violenta	28.5 - 32.6
12	Huracán	> 32.7

La dirección del viento se determina según los puntos cardinales, representados gráficamente la rosa de los vientos; también se emplea el azimut, que corresponde al ángulo que forma la dirección del viento, de 0 a 360 grados en la dirección de las manecillas del reloj, con el cero en el norte magnético. Se mide con la veleta (Irigoyen, 2011). Colombia por encontrarse entre el Trópico de Cáncer y de Capricornio está sometida a los vientos alisios que soplan del noreste en el hemisferio norte y del sureste en el hemisferio sur. En el país por las proximidades del Ecuador, la fuerza de Coriolis es muy pequeña y por tanto, los vientos están fuertemente influenciados por las condiciones locales y por el rozamiento proporcionado por la orografía, en especial por la cordillera de los Andes. Por otra parte, el mar Caribe y el Océano Pacífico tienen su papel en el comportamiento del viento [IDEAM, 2005c].

Todos estos factores influyen en la dirección y velocidad del viento. La variabilidad se explica en gran medida por el desplazamiento de la ZCIT; por tanto, la latitud determina la variación a lo largo del año y los patrones de circulación dominantes [IDEAM, 2005c]. En la región Caribe los vientos se intensifican durante los primeros meses del año cuando la ZCIT se encuentra al sur del país. Por el contrario, entre julio y agosto, se intensifican los vientos al oriente de la región Andina [IDEAM, 2005c]. Por otra parte, el tránsito de ciclones tropicales por el norte del país, afectan los vientos en esta zona [IDEAM, 2005a] [Poveda, 2004]. En general, la velocidad de los vientos dominantes tiende a crecer con la altitud, debido a la reducción de la fricción en el suelo. La orografía es un factor que incide considerablemente en la velocidad del viento. En las montañas de gran altura se presenta el efecto Foehn, caracterizado por vientos fuertes, secos y calientes que desarrollan ocasionalmente en las laderas de sotavento de las cordilleras [IDEAM, 2005c]. Sin embargo, muchos factores de orden local pueden alterar la velocidad del viento. Las montañas originan grandes contrastes, algunas áreas pueden quedar protegidas, mientras que en otras soplan fuerte ráfagas con mayor velocidad [IDEAM, 2005a] [OMM, 2005]. Las cadenas montañosas como la cordillera oriental en Colombia por ejemplo se oponen al flujo de los alisios del sureste y puede conducir al debilitamiento o fortalecimiento de los vientos [IDEAM, 2005c]. La variación diurna del viento en superficie también juega un papel importante en los vientos locales, los cuales se forman como consecuencia del distinto calentamiento y enfriamiento que tiene lugar en las superficies. Por tanto, es importante tener en cuenta los factores locales como las áreas costeras, las topografías que contrastan como valles y montañas y las áreas urbanas [IDEAM, 2005a] [OMM, 2005]:

- Brisa valle-montaña: El calentamiento desigual en montañas y valles a lo largo del día genera viento ascendentes durante el día y descendentes a lo largo de las laderas de las montañas. Durante la noche, los cerros y las montañas producen flujos de vientos descendentes porque el aire es más frío en grandes elevaciones.
- Brisas de mar y brisas de tierra: Las interacciones agua-tierra producen las brisas marinas y terrestres, las cuales son originadas por las diferentes propiedades térmicas de la tierra y el agua (mar, lagos, ríos, embalses), que resultan en un calentamiento desigual de las superficies tierra y agua.
- Isla de Calor: Las áreas urbanas tienen propiedades térmicas diferentes porque los materiales de construcción absorben y retienen el calor de manera más eficiente que el suelo y la vegetación, produciendo un aumento de la temperatura del aire desde la periferia de la ciudad hacia el centro urbano durante el día, lo cual provoca vientos locales desde el exterior al interior y puede generar convección y lluvias en torno al centro cálido, si el aire es muy húmedo. El área urbana continúa irradiando calor después de que el sol se pone, generando el efecto de isla de calor.

En Colombia, se presentan variaciones en la velocidad del viento durante el ciclo estacional, con excepción de la península de la Guajira, donde se presentan durante todo el año vientos iguales o superiores a 5 m/s, que pueden alcanzar los 11 m/s [IDEAM, 2005c]. Para el período comprendido entre diciembre y abril se registran vientos que pueden llegar a los 4 m/s en el Alto y Medio Magdalena y en la Orinoquía colombiana. Para el bajo Magdalena puede superar los 5 m/s. En abril hay una reducción significativa en la intensidad del viento hacia el centro del país por la ubicación de la ZCIT que trae consigo vientos más débiles. Entre mayo y septiembre se incrementa la velocidad del viento a valores cercanos a 6 m/s en el bajo Magdalena y en varias regiones en el interior del país. Entre junio y agosto, estos vientos se extienden sobre parte del Tolima, Risaralda, Quindío, suroriente de Caldas y en el Alto Magdalena. En octubre y noviembre los vientos en el centro del país se debilitan; aunque se mantienen en La Guajira, en el litoral Atlántico y el piedemonte Llanero [IDEAM, 2005c]. En relación a la dirección del viento, los flujos sobre los Llanos Orientales y la Amazonía están influenciados por las circulaciones de tipo sinóptico, determinadas por los vientos alisios y la posición de la ZCIT; mientras que en la región Andina se presenta una mayor influencia de la orografía, con flujos encañonados a lo largo de las cordilleras [IDEAM, 2005c]. De acuerdo con los mapas de velocidad media del viento, en Caldas oscila entre 0.5 m/s a 3.5 m/s, con mayores valores en la región del Parque Nacional Natural los Nevados y menores valores en el municipio de La Dorada [IDEAM, 2005c].

Para el departamento de Caldas, se tienen estaciones con registros de velocidad y dirección del viento cuya información se reporta en la Tabla 1.7. Tomando como base esta información se realiza el análisis a las diferentes escalas. Sin embargo, es importante indicar que algunos registros en especial los del Parque Nacional Natural los Nevados son incompletos y que en general, la red tiene información desde años muy recientes, lo cual no permite analizar el comportamiento para todo el período 1981-2010.

Tabla 1.7: Localización geográfica de estaciones con registros de humedad relativa. Fuente: propia

Estación	Municipio	Altitud	Latitud	Longitud	Registros Desde	Registro Hasta	Años Registro
HOSPITAL VILLAMARIA	Villamaría	1790	5.0°02'56.4"	75°31'6.9"	17/09/2010	31/12/2012	2
CUMANDAY	Villamaría	5072	4.0°53'37.7"	75°20'22.1"	02/09/2003	29/02/2012	9
CISNE-SANTA ISABEL	Villamaría	4812	4.0°49'24.6"	75°21'59.9"	13/09/2003	13/08/2012	9
NEIRA	Neira	2053	5°9'30.5"	75°31'1.7"	26/04/2012	08/10/2012	0
LA PALMA	Manizales	1967	5°5'45.5"	75°31'45.9"	01/11/2006	31/12/2012	6
BOSQUES DEL NORTE	Manizales	2126	5°5'9.67"	75°29'31.83"	22/09/2006	31/12/2012	6
EMAS	Manizales	2060	5°4'49.42"	75°30'32.19"	01/01/1997	30/12/2012	15
INGEOMINAS	Manizales	2226	5°4'17.2"	75°31'27.5"	01/01/2004	31/12/2012	8
ALCÁZARES	Manizales	2057	5°4'0.2"	75°31'39.9"	01/06/2010	31/12/2012	2
EL CARMEN	Manizales	2112	5°3'4.44"	75°31'11.44"	15/02/2005	31/12/2012	7
HOSPITAL DE CALDAS	Manizales	2183	5°3'46.6"	75°30'2.1"	14/12/2007	31/12/2012	5
YARUMOS	Manizales	2195	5°3'47.41"	75°28'52.63"	01/01/2004	31/12/2012	8
POSGRADOS	Manizales	2179	5°3'22.2"	75°29'32.3"	01/01/2002	31/12/2012	10
NIZA	Manizales	2256	5°3'15.57"	75°28'47.67"	01/01/2010	31/12/2012	2
ARANJUEZ	Manizales	1915	5°2'34.2"	75°30'1.6"	26/01/2006	31/12/2012	6
ENEA	Manizales	2092	5°1'43.1"	75°28'20"	01/01/2004	31/12/2012	8

A escala diurna, el comportamiento de la velocidad del viento se presenta en la Figura 1.29 para la estación Posgrados. El coeficiente de variación en esta escala es del 20%; sin embargo, a escala horaria es del 9%. La mayor velocidad promedio se presenta hacia las 11:00 am y la menor a las 7:00 am. Las velocidades se encuentran entre las escalas 1 y 2 de la clasificación de Beaufort y corresponden a aire ligero y brisas muy ligeras.

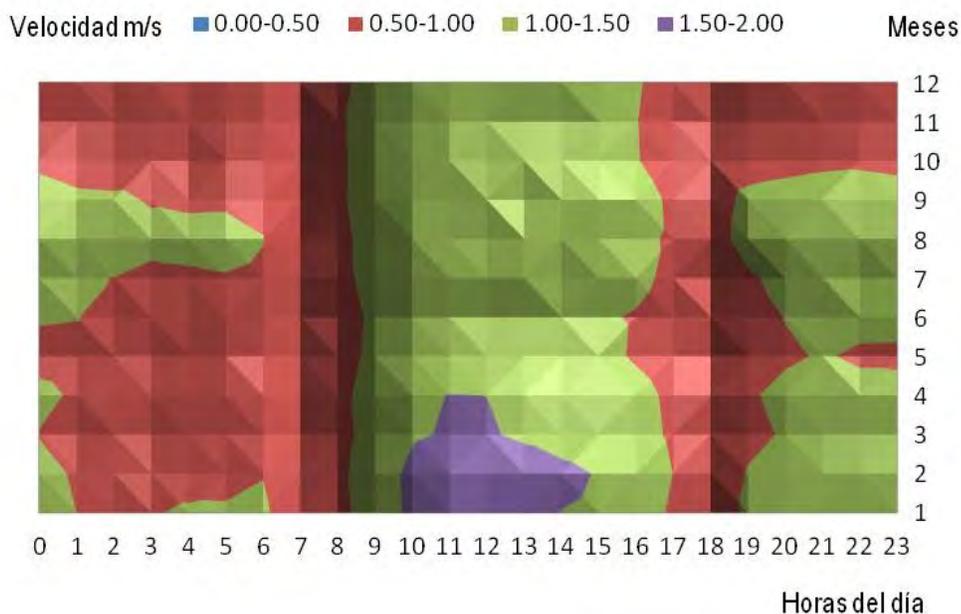


Figura 1.29: Ciclo diario de la velocidad media del viento estación Posgrados 2002-2010. Fuente: propia

El comportamiento estacional de la velocidad del viento promedio multianual para el tiempo de registro se presenta en las Figuras 1.30 y 1.31; en esta última se muestran las estaciones localizadas en el Parque Nacional Natural los Nevados, donde se registran los mayores valores de velocidad del viento, que corresponden a brisas muy ligeras. Con respecto a la dirección del viento, la Figura 1.32 presenta la rosa de los vientos que describe el comportamiento promedio para algunas estaciones localizadas en la zona de estudio.

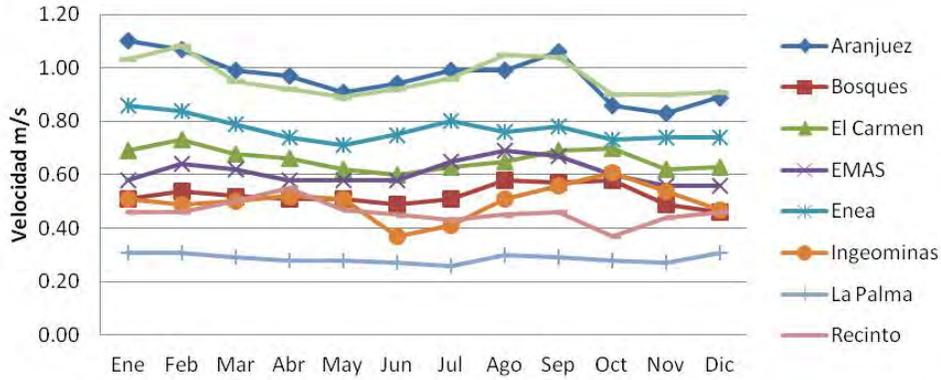


Figura 1.30: Comportamiento de estacional de la velocidad del viento. Fuente: propia

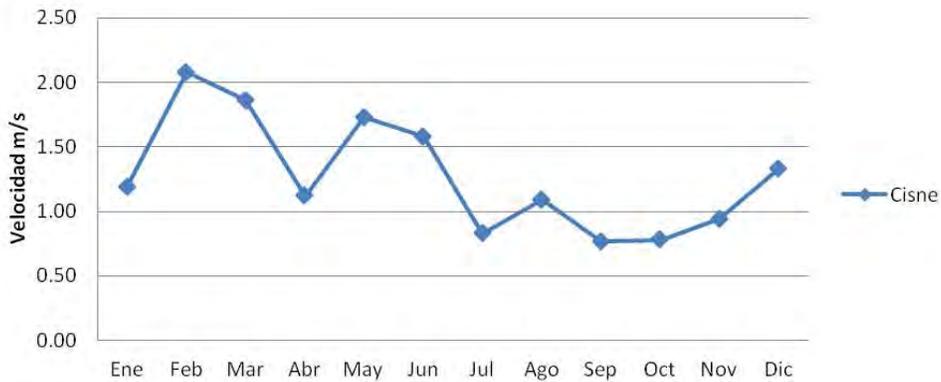


Figura 1.31: Comportamiento de estacional de la velocidad del viento estación El Cisne Parque Nacional Natural Los Nevados. Fuente: propia

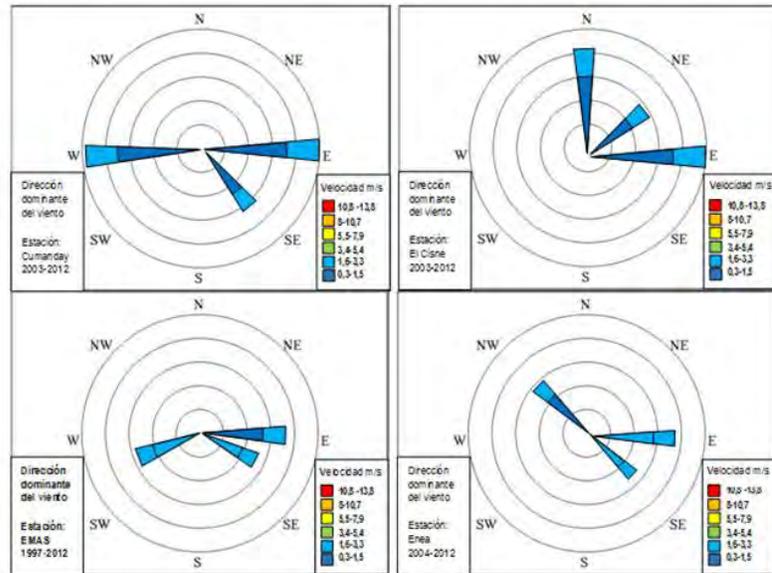


Figura 1.32: Comportamiento de la velocidad y dirección del viento.

## 1.5. Clasificación climática

Existen diferentes tipos de climas: húmedos, secos, cálidos o fríos; ecuatoriales, tropicales, subtropicales, continentales, marítimos, subárticos, mediterráneos, desérticos, de sabana, de estepa, de bosque pluvial, entre otros [OMM, 2011]. Las clasificaciones o zonificaciones climáticas ayudan a identificar este tipo de climas en una región. La clasificación climática está basada en el comportamiento medio de parámetros como la precipitación y la temperatura; entre las más conocidas figuran las clasificaciones de Caldas, Lang y Lang-Caldas. La Clasificación Caldas considera la división en pisos térmicos; esta división se realiza con base en los parámetros que se presentan en la Tabla 1.8. Como sólo considera la variación una variable, la temperatura, se ha tomado como una clasificación incompleta.

Tabla 1.8: Clasificación Climática- Caldas. Adaptada: [IDEAM, 2005a]

Piso Térmico	Símbolo	Rango Altura, m	Temperatura, °C
Cálido	C	<1000	> 24
Templado	T	1001-2000	24 > T > 17,5
Frío	F	2001 a 3000	17,5 > T > 12
Páramo bajo	Pb	3001 a 3700	12 > T > 7
Páramo alto	Pa	37001 a 4200	< 7

La Clasificación Lang por su parte tiene en cuenta el factor de Lang, calculado como la relación entre la precipitación anual en mm y la temperatura en °C. Los tipos de clima identificados por esta clasificación se resumen en la Tabla 1.9. La Clasificación Caldas- Lang (1962) fue propuesta por Schaufleberguer, a partir de las anteriores y por tanto considera la altitud, la temperatura media anual y la precipitación total media anual. La Tabla 1.10 presenta los tipos de climas de esta clasificación.

Tabla 1.9: Clasificación de Lang. Adaptada: [IDEAM, 2005a]

Tipo de Clima	Símbolo	Factor de Lang, P/T
Desértico	D	0 - 20
Árido	A	20,1 - 40
Semiárido	sa	40,1 - 60
Semihúmedo	sh	60,1 - 100
Húmedo	H	100,1 - 160
Superhúmedo	SH	>160

Tabla 1.10: Clasificación de Caldas- Lang. Adaptada: [IDEAM, 2005a]

Tipo climático	Símbolo	P/t	Altitud
Cálido Superhúmedo	CSH	>160	0-1000
Cálido Húmedo	CH	100-160	0-1000
Cálido Semihúmedo	CsH	60-100	0-1000
Cálido Semiárido	Csa	40-60	0-1000
Cálido Árido	CA	20.1-40	0-1000
Cálido Desértico	CD	0-20	0-1000
Templado Superhúmedo	TSH	>160	1000-2000
Templado Húmedo	TH	100-160	1000-2000
Templado Semihúmedo	Tsh	60-100	1000-2000
Templado Semiárido	Tsa	40-60	1000-2000
Templado Árido	TA	20.1-40	1000-2000
Templado Desértico	TD	0-20	1000-2000
Frío Superhúmedo	FSH	>160	2000-3000
Frío Húmedo	FH	100-160	2000-3000
Frío Semihúmedo	Fsh	60-100	2000-3000
Frío Semiárido	Fsa	40-60	2000-3000
Frío Árido	FA	20.1-40	2000-3000
Frío Desértico	FD	0-20	2000-3000
Páramo Bajo Superhúmedo	PBSH	>160	3000-3700
Páramo Bajo Húmedo	PBH	100-160	3000-3700
Páramo Bajo Semihúmedo	PBsh	60-100	3000-3700
Páramo Bajo Semiárido	PBsa	40-60	3000-3700
Páramo Alto Superhúmedo	PASH	20.1-40	3700-4200
Páramo Alto Húmedo	PAH	0-20	3700-4200
Nieves Perpetuas	NP		>4200

Tomando como base el modelo de elevación digital del terreno y el comportamiento promedio de la temperatura y la precipitación, se realizó la clasificación climática de la zona de estudio. La Figura 1.33 muestra que en Caldas se presentan todos los pisos térmicos desde páramo alto, en la zona del Parque Nacional Natural los Nevados, hasta cálido en los valles de los ríos Magdalena y Cauca.

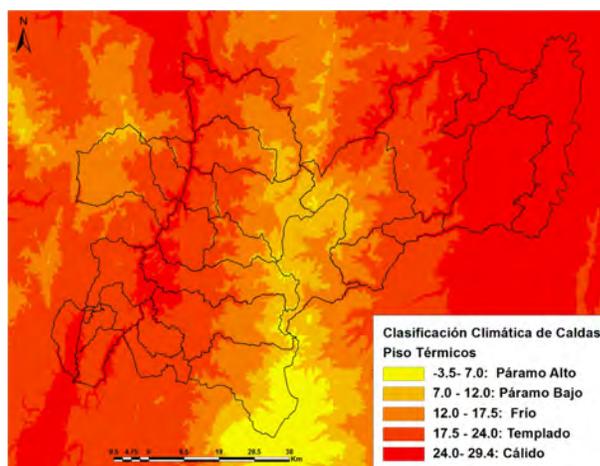


Figura 1.33: Pisos térmicos en el departamento de Caldas 1981-2010. Fuente: propia

La Figura 1.34 presenta la Clasificación climática de Lang; la mayor proporción de la superficie corresponde a clima superhúmedo, el cual está localizado en la Cordillera Central. En los valles de los ríos el clima es semihúmedo.

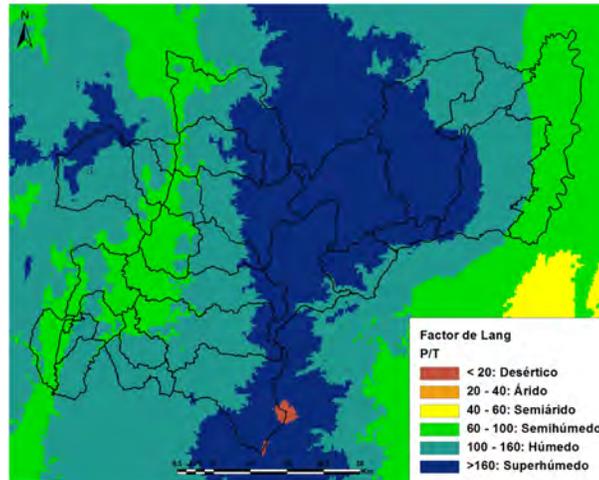


Figura 1.34: Clasificación climática de Lang departamento de Caldas 1981-2010. Fuente: propia

La Figura 1.35 presenta la clasificación climática de Lang-Caldas, la cual permite apreciar la diversidad de climas que se encuentran en el departamento de Caldas.

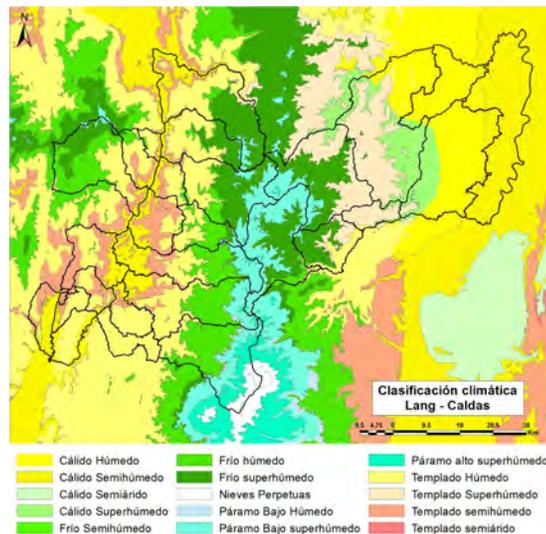


Figura 1.35: Clasificación climática de Lang-Caldas departamento de Caldas 1981-2010. Fuente: propia

## 1.6. Conclusiones

En este capítulo se ha desarrollado una descripción del clima en general, con énfasis en el Departamento de Caldas, una zona compleja y diversa que requiere de un mejor entendimiento de su clima para los procesos de planificación de cuencas que se vienen desarrollando en todo

el país y en el departamento. Esta descripción y análisis de las variables del clima en la región se convierten en la base para la toma de decisiones en muchas áreas del conocimiento y sirven como elemento guía en aspectos relevantes como son la adaptación al cambio climático y a la variabilidad climática.

La temperatura esta influenciada por la altitud, con una variabilidad alta y con un gradiente altitudinal negativo para la región de 6.4°C por km, con un ciclo diurno muy marcado de 12 horas. No hay una variabilidad estacional importante debido a que nos encontramos en el trópico, pero se observan tendencias en los últimos años lo cual permite evidenciar las señales de cambio climático en la región. En la escala interanual se evidencian los efectos de los eventos ENSO y sus fases cálida y fría.

El comportamiento de las lluvias es complejo y diverso en todo el Departamento, ya que hay zonas extremadamente lluviosas como el Bosque de Florencia y zonas secas como las zonas de páramos y glaciares, la lluvia está muy condicionada por aspectos climáticos que se desarrollan a múltiples escalas. La mayor variabilidad interanual se debe al ENSO en sus fases cálida y fría y la variación interanual se correlaciona con el ZCIT, y no se destacan tendencias para los últimos años. En cuanto al ciclo diurno, los valores más altos se registran entre las 2:00 pm y las 4:00 pm; mientras que los más bajos se reportan entre las 9 y 10 am. Estos valores de lluvia sirven de referencia a los constructores para programar vaciados de concretos y pavimentos en horas de la mañana, cuando se observan menores lluvias, mayor radiación y temperaturas altas.

El brillo solar, la radiación solar y la humedad relativa, se comportan de forma razonable para el trópico, con una ligera disminución del brillo en los últimos años. El viento no es relevante para todo el departamento, aunque en pequeñas zonas se pueden reportar velocidades de 6 m/s. Sin embargo, se destaca la poca información disponible en algunas regiones del departamento para estas variables climáticas.

Las evidencias de cambio climático y oscurecimiento han sido reportadas en otros estudios y son confirmados en éste. Para todas las variables el ciclo diurno es muy marcado y debe ser estudiado con mayor detalle. En este estudio sólo se emplean unas pocas estaciones con información disponible a escala horaria, por lo que se hace necesario un mayor cubrimiento para poder detallar mejor el comportamiento en el departamento. La poca cobertura espacial no permite obtener resultados contundentes, pero sí son un buen indicativo de los comportamientos y tendencias de las diferentes variables climáticas. CORPOCALDAS se encuentra en la actualidad realizando un importante esfuerzo para mejorar la densidad de estaciones climatológicas e hidrológicas en el departamento de Caldas, por lo que este tipo de estudios puede ser un referente para el diseño de esta red de monitoreo.

Todos estos elementos del clima son los que tienen mayor influencia sobre los ecosistemas y las actividades desarrolladas por las comunidades y actores asentados en la cuenca hidrográfica. Por lo tanto, es importante que los procesos de Planificación de cuenca y de Ordenación del Territorio que se llevan a cabo en la región recojan estas sugerencias y análisis para poder mejorar la calidad de la información en los estudios futuros, de tal forma que a medida que pase el tiempo sea posible tomar decisiones con más y mejor información en el departamento de Caldas.

## Agradecimientos

Los autores quieren expresar sus agradecimientos CORPOCALDAS, CENICAFÉ, CHEC, GENSA, IDEA y la Alcaldía de Manizales a través de la OMPAD por suministrar la información climática. A la Facultad de Ingeniería y Arquitectura de la Universidad Nacional de Colombia, al programa “Fortalecimiento de capacidades de procesamiento y análisis de información ambiental” del DIMA y al proyecto “Vulnerability and Adaptation to Climate Extremes in the Americas – VACEA” cuyos aportes han sido significativos.

# Referencias Bibliográficas

- [Adger et al., 2007] Adger, W., Agrawala, S., Mirza, M., Conde, C., O'Brien, K., Pulhin, J., Pulwarty, R., Smit, B., and Takahashi, K. (2007). Assessment of adaptation practices, options, constraints and capacity; in climate change 2007: Impacts, adaptation and vulnerability (contribution of working group ii to the fourth assessment report of the intergovernmental panel on climate change). *Cambridge University Press, Cambridge, United Kingdom.*, 17:133–171.
- [Barry., 1992] Barry. (1992). Mountain weather and climate. *Routledge Physical Environmental Series.*, page 402p.
- [Bojariu and Gimeno, 2003] Bojariu, R. and Gimeno, L. (2003). Modelling and predictability of the north atlantic oscillation. *Earth Sci. Rev.*, 63:145–168.
- [CEPAL, 2009] CEPAL (2009). Cambio climático y desarrollo en américa latina y el caribe: una reseña.
- [Chaves and Jaramillo., 1998] Chaves, B. and Jaramillo. (1998). Regionalización de la temperatura del aire en colombia. *Revista Cenicafé*, 24:91–104.
- [CORPOCALDAS, 2001] CORPOCALDAS (2001). Plan de gestión ambiental regional para caldas. pgar 2001-2006. *Manizales: CORPOCALDAS.*
- [CORPOCALDAS, 2007] CORPOCALDAS (2007). Plan de gestión ambiental regional para caldas. 2007-2019. *Manizales: CORPOCALDAS.*
- [Cortés, 1989] Cortés, E. (1989). Estudio del régimen de temperatura en colombia. *HIMAT. Bogotá.*, page 103 pp.
- [Dore, 2004] Dore (2004). Climate change and changes in global precipitation patterns: What do we know? st catharines. *Brock University.*
- [Eslava, 1992] Eslava, J. (1992). Perfil altitudinal de la temperatura del aire en colombia. *Geofis. Colomb.*, 1:37–52.
- [Eslava, 1993] Eslava, J. (1993). Climatología y diversidad climática de colombia. 18(71),):507–538.
- [Eslava, 1994] Eslava, J. (1994). Climatología del pacífico colombiano.
- [Eslava et al., 1986a] Eslava, J., Lopez, V., and Olaya, G. (1986a). Contribución al conocimiento del régimen térmico y pluviométrico de colombia. *Colombia Geográfica*, 12(2):53–118.
- [Eslava et al., 1986b] Eslava, J., Lopez, V., and Olaya, G. (1986b). Los climas de colombia (sistema de caldas - lang). *Atmósfera*, 7:41–77.
- [Eslava et al., 1986c] Eslava, J., Lopez, V., and Olaya, G. (1986c). Los climas de colombia (sistema de c.w.thornthwaite). *Atmósfera*, 6:33–76.
- [Eslava et al., 1986d] Eslava, J., Lopez, V., and Olaya, G. (1986d). Los climas de colombia (sistema de e. de martonne). *Atmósfera*, 8:47–79.

- [Eslava et al., 1986e] Eslava, J., Lopez, V., and Olaya, G. (1986e). Los climas de colombia (sistema de w. koeppen). *Atmósfera*, 5:35–81.
- [FAO, 2003] FAO (2003). World agriculture: Towards 2015/2030.
- [FECYT., 2004] FECYT. (2004). Meteorología y climatología.
- [France, ] France, M.
- [Garreaud et al., 2008] Garreaud, Vuille, Compagnucci, and Marengo. (2008). Present-day south american climate. *PALAEO3 Special Issue*, 281:180–195.
- [Griffiths., 1994] Griffiths. (1994). Atmospheric characteristics. *Oxford University Press.*, pages 3–20.
- [Herzog et al., 2013] Herzog, Martinez, Jorgense, and Tiessen. (2013). Cambio climatico y biodiversidad en los andres tropicales. *Inter-American Institute for Global Change Research and Scientific Committe on Problems of the Environment.*, page 348.
- [Hoyos et al., 2013] Hoyos, Escobar, Restrepo, Arango, and Ortiz. (2013). Impact of the 2010-2011 la niña phenomenon in colombia, south america: The human toll of an extreme weather event. *Applied Geography* 39, pages 16–25.
- [Hurrell and Loon., 1997] Hurrell and Loon., V. (1997). Decadal variations in climate associated with the north atlantic oscillation. *Climatic Change*, 36:301–326.
- [Hurrell et al., 2003] Hurrell, J., Kushnir, Y., Visbeck, M., and Ottersen., G. (2003). An overview of the north atlantic oscillation. *Geophys. Monogr.*, 134.
- [IDEAM, 1998] IDEAM (1998). Posibles efectos naturales y socioeconómicos del fenómeno frío del pacífico (la niña) en colombia en el segundo semestre de 1998 y primer semestre de 1999. page 88.
- [IDEAM, 2005a] IDEAM (2005a). Atlas climatológico de colombia. bogotá: Ministerio de ambiente, vivienda y desarrollo territorial.
- [IDEAM, 2005b] IDEAM (2005b). Atlas de radiación solar de colombia. bogoá: Ministerio de ambiente.
- [IDEAM, 2005c] IDEAM (2005c). Atlas de vientos de colombia. bogotá: Ministerio de ambiente.
- [IDEAM, 2009] IDEAM (2009). Boletines informativos sobre el monitoreo del fenómeno del niño.
- [IDEAM, 2010] IDEAM (2010). 2a comunicación nacional ante la convención marco de las naciones unidas sobre el cambio climático.
- [IPCC., 2007] IPCC. (2007). Cambio climático 2007: Informe de síntesis.
- [IPCC, 2008] IPCC (2008). El cambio climático y el agua.
- [Jaramillo and Ramírez, 2009] Jaramillo and Ramírez (2009). Relación entre el índice oceánico de el niño y la lluvia en la región andina central de colombia. 60(2):161–172.
- [Jaramillo, 2005] Jaramillo, A. (2005). Clima andino y café en colombia.
- [Jones et al., 2003] Jones, P., Osborn, T. J., and Briffa, K. R. (2003). Pressure-based measures of the north atlantic oscillation (nao): A comparison and an assessment of changes in the strength of the nao and in its influence on surface climate parameters, in the north atlantic oscillation: Climate significance and environmental impact. 134:51–62.
- [León et al., 2000] León, Zea, and Eslava. (2000). Circulación general del trópico y la zona de confluencia intertropical en colombia. pages 31–38.
- [Linacre., 1992] Linacre. (1992). Climate data and resources.

- [Lowry, 1973] Lowry. (1973). Compendio de apuntes de climatología para la formación de personal meteorológico de la clase iv. page 167.
- [Madden and Julian., 1971] Madden and Julian. (1971). Detection of a 40-50 day oscillation in the zonal wind in the tropical. pacific. *J. Atmos. Sci.*, 28:702–708.
- [Mantua and Hare., 2002] Mantua and Hare. (2002). Principales sistemas sinópticos en colombia y su influencia en el comportamiento del tiempo. *Atmósfera.*, 16:1–10.
- [Martínez et al., 2013] Martínez, Ruiz, Andrade, Blacutt, Pabón, Jaimes, León, Villacís, Quintana, Montealegre, and Euscátegui. (2013). Sistemas del clima de los andes tropicales. en: Cambio climatico y biodiversidad en los andes tropicales. pages 117–130.
- [Mejía et al., 1999] Mejía, Mesa, Poveda, Vélez, Hoyos, Mantilla, Barco, Cuartas, Montoya, and Botero. (1999). Distribución espacial y ciclos anual y semianual de la precipitación en colombia. 127:7–26.
- [Mesa et al., 1997] Mesa, O., Poveda, G., and Carvajal, L. (1997). Introducción al clima de colombia. universidad nacional de colombia.
- [Mitchell and Wallace., 1992] Mitchell and Wallace. (1992). The anual cycle in the equatorial convection an sea surface temperatura. *Journal of Climate.*, 5:1140–1156.
- [Monsalve, 2008] Monsalve, G. (2008). Hidrología en la ingeniería.
- [Montealegre, 2005] Montealegre, E. (2005). Escalas de variabilidad climática. instituto de hidrología meteorología y medio ambiente.
- [Mora and Sturmo., 1994] Mora and Sturmo. (1994). Estudios ecológicos del páramo y del bosque altoandino cordillera oriental de colombia. page 715.
- [Murthy., 2002] Murthy. (2002). Basic principles of agricultural meteorology. page 261.
- [Newman et al., 2003] Newman, Compo, and Alexander. (2003). Enso-forced variability of the pacific decal oscillation. *Journal of Climate.*, 16:3853–3857.
- [NOAA, 2013] NOAA (2013). National weather service climate prediction center.
- [Ocampo, 2012] Ocampo, O. (2012). Análisis de vulnerabilidad de la cuenca del río chinchiná para condiciones estacionarias y de cambio climático.
- [Ocampo et al., 2013] Ocampo, O., J.J, V., and A., L. (2013). Análisis de vulnerabilidad hídrica de la cuenca del río chinchiná.
- [Ocampo et al., 2012a] Ocampo, O., Vélez, J., and Londoño, A. (2012a). Evidence of climate change on high andean mountain watersheds: Chinchina river basin, colombia.
- [Ocampo et al., 2012b] Ocampo, O., Vélez, J., and Londoño, A. (2012b). Hydrologic impact assessment due to climate change at chinchiná river basin, colombia.
- [Ocampo et al., 2012c] Ocampo, O., Vélez, J., and Londoño, A. (2012c). Señales de cambio climático en cuencas hidrográficas de alta montaña, caso del río chinchiná.
- [Ocampo et al., 2012d] Ocampo, O., Vélez, J., and Londoño, A. (2012d). Señales de cambio climático en cuencas hidrográficas de alta montaña, caso del río chinchiná.
- [OMM, 2002] OMM (2002). Servicio de información y predicción del clima (sipc).
- [OMM, 2003] OMM (2003). Nuestro clima futuro.
- [OMM, 2004] OMM (2004). Tiempo, clima, agua en la era de la información.
- [OMM, 2005] OMM (2005). Tiempo, clima, agua y desarrollo sostenible.
- [OMM, 2006] OMM (2006). La organización meteorológica mundial de un vistazo.

- [OMM, 2008] OMM (2008). Guía de prácticas hidrológicas.
- [OMM, 2010] OMM (2010). Sistema de vigilancia del clima. alerta temprana de anomalías y fenómenos climáticos extremos.
- [OMM, 2011] OMM (2011). El clima y tú.
- [OMM, 2012] OMM (2012). Declaración de la omm sobre el estado del clima mundial en 2011.
- [OMM, 2013a] OMM (2013a). Declaración de la omm sobre el estado del clima mundial en 2013.
- [OMM, 2013b] OMM (2013b). El estado del clima mundial 2001-2010. un decenio de fenómenos climáticos extremos..
- [Oster., 1979] Oster. (1979). Las precipitaciones en colombia. 6(2):5–147.
- [Pabón and Rivera., 1993] Pabón, J. and Rivera. (1993). Análisis de la variabilidad interanual de los caudales de los ríos magdalena y cauca. *Atmósfera*, (20):1–11.
- [Pabón et al., 2001] Pabón, Eslava, and Gómez. (2001). Generalidades de la distribución espacial y temporal de la temperatura del aire y de la precipitación en colombia. *Meteorología Colombiana.*, 4:47–59.
- [Pabón et al., 1998] Pabón, Zea, León, Montealegre, Hurtado, and González. (1998). La atmósfera, el tiempo y el clima. en: El medio ambiente en colombia. pages 38–86.
- [PNUD, 2000] PNUD (2000). Panorama del impacto ambiental de los recientes desastres naturales en américa latina y el caribe.
- [Ponce, 1989] Ponce (1989). Engineering hydrology. *Prentice Hall*.
- [Poveda et al., 2005] Poveda, Mesa, Salazar, Arias, Moreno, Vieira, Agudelo, Toro, and Alvarez. (2005). The diurnal cycle of precipitation in the tropical andes of colombia. *Monthly Weather review*, 113:228–240.
- [Poveda et al., 2002] Poveda, Vélez, Mesa, Hoyos, Mejia, Barco, and Correa. (2002). Influencia de fenómenos macroclimáticos sobre el ciclo anual de la hidrología colombiana: cuantificación lineal, no lineal y percentiles probabilísticos. *Meteorología Colombiana*, (6):121–130.
- [Poveda, 1994] Poveda, G. (1994). Cuantificación de los efectos de el niño y la niña sobre los caudales mensuales de los ríos colombianos.
- [Poveda, 2004] Poveda, G. (2004). La hidroclimatología de colombia. una síntesis desde la década interdecenal hasta la escala diurna. 28(107):201–222.
- [Poveda, 2011] Poveda, G. (2011). Mixed memory, (non) hurst effect, and maximum entropy of rainfall in the tropical andes. 34(2):243–256.
- [Poveda et al., 2011] Poveda, G., Alvarez, and Rueda. (2011). Hydro-climatic variability over the andes of colombia associated with enso: A review of climatic processes and their impact on one of the earth’s most important biodiversity hotspots. *Climate Dynamics*, 36:11–12.
- [Poveda et al., 1998] Poveda, G., Gil, and Quiceno. (1998). El ciclo anual de la hidrología de colombia en relación con el enso y la nao. 27(3):721–731.
- [Poveda and Jaramillo, 2000] Poveda, G. and Jaramillo, A. (2000). Enso-related variability of river discharges and soil moisture in colombia. biospheric aspects of the hydrologic cycle. 8:3–6.
- [Poveda et al., 2001] Poveda, G., Jaramillo, A., Gil, M., Quiceno, N., and Mantilla, R. (2001). Seasonality in enso related precipitation, river discharges, soil moisture, and vegetation index (ndvi) in colombia. 37(8):2169–2178.

- [Poveda and Mesa, 1997] Poveda, G. and Mesa, O. (1997). Feedbacks between hydrological processes in tropical south america and large scale oceanic-atmospheric phenomena. *j. climate*. (10):2690–2702.
- [Poveda and O.J., 1993] Poveda, G. and O.J., M. (1993). Metodologías de predicción de la hidrología colombiana considerando el evento el niño-oscilación del sur (enos). *Atmósfera*, (17).
- [Poveda and O.J., 1996] Poveda, G. and O.J., M. (1996). Las fases extremas del enso - el niño y la niña - y su influencia sobre la hidrología de colombia. *Revista de Ingeniería Hidráulica*, (11):21–37.
- [Poveda and O.J., 1999] Poveda, G. and O.J., M. (1999). La corriente de chorro superficial del oeste (“del chocÓ”) y otras dos corrientes de chorro atmosféricas sobre colombia: Climatología y variabilidad durante las fases del enso. *Revista de la Academia Colombiana de Ciencias*, 23(89):517–528.
- [Poveda and O.J., 2000] Poveda, G. and O.J., M. (2000). On the existence of lloró (the rainiest locality on earth): Enhanced ocean-atmosphere-land interaction by a low-level jet. *Geophysical Research Letters*, 27(11):1675–1678.
- [Poveda et al., 2007] Poveda, G., Vélez, J., Mesa, O., Cuartas, A., Barco, J., Mantilla, R., Mejía, J., Hoyos, C., Ramírez, J., Ceballos, L., Zuluaga, M., Arias, P., Botero, B., Montoya, M., Giraldo, J., and Quevedo, D. (2007). Linking long-term water balances and statistical scaling to estimate river flows along the drainage network of colombia. *journal of hydrologic engineering*. 12(1):1–4.
- [Puertas and Carvajal., 2008] Puertas and Carvajal. (2008). Incidencia de el niño oscilación del sur en la precipitación y la temperatura del aire en colombia utilizando el climate explorer. 23:104–118.
- [Ruiz et al., 2009] Ruiz, Arroyave, Molina, Barros, Gutiérrez, and Zapata. (2009). Signals of climate variability/change in surface water supply of high-mountain watersheds case study: Claro river high mountain basin, los nevados natural park, andean central mountain range, colombia. page 207.
- [Salazar, 1990] Salazar, A. (1990). Ciclicidad de los caudales en colombia, fases húmedas y secas. *Atmósfera*., 14:42–53.
- [Schneider and Cornuelle., 2005] Schneider and Cornuelle. (2005). The forcing of the pacific decadal oscillation. 18:4355– 4373.
- [Silva, 1998] Silva, G. (1998). Hidrología básica.
- [Stern and Nicholas., 2006] Stern and Nicholas. (2006). The economics of te change. the stern review.
- [Swartman and Ogulande, 1966] Swartman, R. K. and Ogulande, O. (1966). An investigation on packed bed collectors. *solar energy*. 10(3):106–110.
- [Trenberth and Thomas., 2003] Trenberth and Thomas. (2003). Modern global te change. *Science* 5, 302.
- [Velásquez and Rosales, 2002] Velásquez, A. and Rosales, C. (2002). Los pasos del niño en colombia 1980-2001.
- [Vélez et al., 2012] Vélez, J. J., Duque, N., Mejía, F., and Orozco, M. (2012). Red de monitoreo climático para dar apoyo a la prevención y atención de desastres en manizales. pages 14–18.
- [Vélez et al., 2002] Vélez, J., Poveda, Mesa, Hoyos, Salazar, and Mejía. (2002). Modelo de interacción suelo-atmósfera para la estimación de caudales medios mensuales en colombia. *Meteorología Colombiana*, 6:81–89.

- [Wang et al., 2008] Wang, Peterson, Lawrimore, Brunet, Cerveny, Donlon, Driouech, Hassan, W., Hollmann, Schwartz, and Zhang. (2008). *Vigilancia del clima de la tierra*. 57:109–113.
- [Waylen and Poveda., 2002] Waylen, P. and Poveda., G. (2002). El niño-southern oscillation and aspects of western south america hydro-climatology. 16:1247–1260.
- [Weickman, 1991] Weickman, K. (1991). El niño/southern oscillation and madden-julian (30-60 day) oscillations during 1981-1982. *Journal Geoph*, 96:3187–3195.
- [WMO, 2007] WMO (2007). The role of climatological normals in a changing climate.

## Capítulo 2

# Predicción de caudales con redes neuronales artificiales y algoritmos evolutivos mediante la exploración de las variables internas en modelos hidrológicos

*Darwin Brochero*<sup>1,2</sup>

*Karol López*<sup>2</sup>

*Jorge Julián Vélez*<sup>3</sup>

<sup>1</sup>Grupo de investigación en pronóstico y acciones hidrológicas. Departamento de Ingeniería Civil e Ingeniería de Aguas. Universidad Laval, Quebec (Quebec), Canadá.

<sup>2</sup>Laboratorio de visión y sistemas numéricos. Departamento de Ingeniería Eléctrica e Informática. Universidad Laval, Quebec (Quebec), Canadá

<sup>3</sup>Departamento de Ingeniería Civil - Facultad de Ingeniería Universidad Nacional de Colombia - Sede Manizales

*... "Ten cuidado con el presente que construyes, debe parecerse al futuro que sueñas". Mujeres  
Creando ...*

## Resumen

Las ventajas del modelo semi-distribuido de simulación hidrológica TETIS han sido demostradas en distintos escenarios de simulación hidrológica, incluyendo la generación de mapas de riesgo por inundación y deslizamientos que están estrechamente relacionados con la compleja dinámica de la humedad del suelo. En este trabajo se presenta un modelo de predicción de caudales del río Risaralda (Colombia) con base en la combinación de distintas técnicas de minería de datos. De forma general, la idea es aprovechar el alto rendimiento del modelo TETIS en la descripción del proceso hidrológico para plantear un modelo regresivo de sus variables internas. Para esto se muestra una aplicación que combina varias tareas clásicas de la minería de datos: la selección de variables, el compromiso de distintos criterios de rendimiento, la regresión no lineal y el agrupamiento (*clustering*) de la información. En síntesis, en este capítulo se presenta un

acoplamiento de un método de regresión no lineal conocido como redes neuronales artificiales (RNA) y un algoritmo evolutivo multi-objetivo denominado Algoritmo Genético de Clasificación No-dominada II (traducción del inglés *Non-dominated Sorting Genetic Algorithm II*, NSGA-II), a los resultados de un modelo hidrológico con fines de predicción de caudales.

**Palabras Clave:** Predicción de caudales, Algoritmos evolutivos, Redes neuronales, Modelo TETIS

## 2.1. Introducción

Para entender la complejidad de la predicción de caudales, se puede partir del esquema básico de la simulación hidrológica, cuya dinámica puede entenderse como un proceso causa-efecto (lluvia-escorrentía) del cual se debe abstraer cuatro grandes subprocesos dentro de la cuenca de estudio:

- La distribución espacial y temporal de la precipitación líquida directa o indirecta (fusión de la nieve).
- Los mecanismos de interacción entre las variables atmosféricas y las características del terreno, en especial la evapotranspiración, y la heterogeneidad de los tipos de cobertura y textura de los suelos.
- La evaluación de la humedad del suelo, la capacidad de recarga hidráulica del suelo, la infiltración y percolación para finalmente determinar los diferentes tipos de flujos que dan lugar a la escorrentía directa, el flujo sub-superficial y al caudal base.
- La conceptualización del esquema hidráulico de transporte (*routing*) de los diferentes tipos de caudal.

Este sistema natural, inherentemente complejo, es representado con un modelo hidrológico. En la última década se viene proponiendo el acoplamiento de modelos hidrológicos especializados para cada subproceso físico, dando lugar a la unión de modelos atmosféricos y esquemas de transferencia entre suelo - atmósfera - vegetación [Pietroniro et al., 2007, Kunstmann et al., 2008, Zabel and Mauser, 2013]. Según el tipo de modelación, un modelo hidrológico puede ser clasificado como:

- Modelo físico o de caja blanca (*white-box or physical based models*): todos los subprocesos son conceptualizados basados en las leyes de la física, considerando el balance de energía y masa del sistema. Idealmente todas las ecuaciones tienen un sentido físico y los parámetros pueden ser calculados a partir de mediciones del sistema.
- Modelos orientados por datos o de caja negra (*black-box or data-driven models*): son métodos basados en los datos donde la dinámica e información de cada subproceso no se considera explícitamente dentro del modelo. En su lugar, el análisis de las series hidroclimatológicas contienen la información necesaria para modelar el sistema sin considerar procesos físicos puntuales. Una reconocida gama de estos modelos es la Hidroinformática, que utiliza técnicas como las redes neuronales artificiales (RNA), los algoritmos evolutivos (AE) u otras técnicas propias del aprendizaje automático (*machine learning*) para el análisis, gestión y previsión de desastres relacionados con fenómenos naturales donde intervienen recursos hídricos.
- Modelos conceptuales o de caja gris (*gray-box or conceptual models*): los subprocesos son considerados a partir de ecuaciones matemáticas o semi-empíricas simplificadas con sentido físico. Los parámetros usados en estos modelos obedecen principalmente a la calibración de los mismos con técnicas de optimización tales como los AE [Eiben and Smith, 2003] u otras

técnicas tales como el Barajado y Evolución de Complejos desarrollado en la Universidad de Arizona, (conocido por sus siglas en inglés, SCE-UA) o una de sus variantes [Duan et al., 1993, Vrugt et al., 2003b, Vrugt et al., 2003a].

### 2.1.1. Hipótesis básicas en la modelación hidrológica

No se debe perder de vista que la simulación hidrológica, como una representación simplificada de la realidad, conlleva a un error inherente entre las mediciones de caudal y los valores simulados. De tal forma, el objetivo del modelador es minimizar tal error, para lo cual se emplean técnicas de optimización que calibran, en algunos casos, decenas de parámetros para ajustar la respuesta del modelo a las observaciones. En este punto, es importante resaltar que siempre se corre el riesgo de que el modelo se comporte como una marioneta matemática jugando a replicar las observaciones aún si las premisas de base no son realistas [Kirchner, 2006].

Ahora bien, partiendo de la aceptación del error en la modelación, éste puede ser concebido como el resultado de la incertidumbre conjunta en relación con las mediciones y las conceptualizaciones. Sin embargo, existen factores que pueden aumentar la incertidumbre, por ejemplo las observaciones no están disponibles en toda la cuenca ni tienen un grado infinito de precisión o las ecuaciones usadas no capturan completamente los procesos dentro de la cuenca. Otro problema que aumenta la incertidumbre es la escala hidrológica, puesto que algunos procesos que son importantes en una escala no lo son necesariamente en otra [Sivapalan et al., 2003].

En el caso de la **predicción** de caudales con modelos físicos o conceptuales, una nueva fuente de incertidumbre es agregada al proceso: la predicción de variables atmosféricas y/o meteorológicas que sirven de entradas al modelo de simulación hidrológica. Al respecto, diversos autores ya han destacado esta nueva fuente de incertidumbre como la más influyente en el proceso de predicción [Cloke and Pappenberger, 2009]. Por el contrario, en los modelos orientados por datos no es necesario plantear la predicción de caudales en dependencia del rendimiento del modelo de predicción de la precipitación; en este caso la hipótesis es la existencia de una relación entre las observaciones pasadas y el horizonte de predicción. En consecuencia, el análisis consiste básicamente en un modelo de regresión basado en la información rezagada de distintas variables hidroclimatológicas. Sin embargo, la selección de los datos a modelar, la estimación de las variables de entrada, la elección del tipo de modelo matemático y su incertidumbre paramétrica, son fuentes activas de incertidumbre en este tipo de modelación donde las RNA son reconocidas por su precisión y alta eficiencia computacional [Maier and Dandy, 2000]. A pesar de las diferencias filosóficas entre los distintos tipos de modelación, los mayores avances en hidrología pueden esperarse en su confluencia [Sivapalan et al., 2003].

### 2.1.2. Las RNA en hidrología

La eficiencia de las RNA como modelo de regresión en hidrología ha sido demostrada en numerosos estudios, una síntesis de tales avances en las últimas dos últimas décadas se encuentra en [Maier and Dandy, 2000, Maier et al., 2010, Abrahart et al., 2010, Abrahart et al., 2012]. No obstante, uno de los grandes retos de la comunidad Hidroinformática es la incursión sólida de las RNA en el “mundo real”, pues aunque en la comunidad académica se reconoce su alto rendimiento con bases de datos reales, la puesta en marcha operativa de tales modelos requiere de un cambio de paradigma, por ejemplo considerando un contexto de aplicación cooperativo con los modelos convencionales en hidrología.

Al respecto, [Abrahart et al., 2010] han presentado el denominado “ciclo de vida de una adopción tecnológica” (Figura 2.1). En este ciclo, se muestra como una primera fase los innovadores, posteriormente los visionarios, seguidos de los pragmáticos, los conservadores y finalmente los escépticos. El estado actual del arte muestra las RNA en pleno desarrollo dentro del estado de los visionarios pero con un inmenso trabajo para llevarlo al siguiente estado de los pragmáticos, es

decir, aquellos que buscan el mejoramiento de sus modelos minimizando cualquier cambio entre las aplicaciones actuales y la adopción de una innovación particular. Note en la Figura 2.1 que el ancho de las interrupciones entre cada estado es proporcional a la dificultad de la transición.

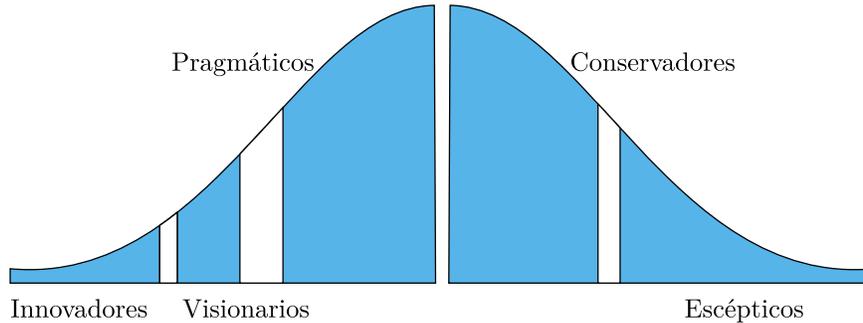


Figura 2.1: Ciclo de vida de una adopción tecnológica.

### 2.1.3. Planteamiento general

El problema planteado en este estudio puede, en su versión semi distribuida, entenderse como la exploración de las variables internas del modelo hidrológico TETIS, el cual incluye las observaciones de precipitación, temperatura y caudal, con un modelo basado en RNA para establecer un sistema de predicción de caudales. Este trabajo está inspirado en el desarrollado por [Anctil et al., 2003] con respecto al análisis de las variables internas de varios modelos hidrológicos para el desarrollo de un modelo de predicción con RNA; sin embargo esta propuesta se complementa con ciertos puntos metodológicos que han sido destacados por [Abrahart et al., 2012] en lo que se concibe como la Agenda de Investigación de las RNA en Hidrología (AIRNAH), la cual considera una investigación internacional coordinada que evite las investigaciones aisladas y desengranadas en la búsqueda de un fin único: la aceptación general de las RNA en la práctica de la hidrología.

De tal forma, a continuación se expone brevemente el planteamiento general de este estudio resaltando la asociación con los siguientes puntos definidos en la AIRNAH:

- Soluciones híbridas.
- Sistemas multi-modelo (*ensemble systems*).
- Manipulación de datos.
- Selección de variables de entrada.
- Acoplamiento neuro-evolutivo.
- Marco multi-objetivo de optimización.

En esencia, la base del planteamiento cooperativo entre las RNA y un modelo hidrológico se destaca como una **solución híbrida** que muestra la importancia de las nuevas tecnologías de la información en la hidrología. Para esto, se considera como modelo de predicción la media de un **sistema multi-modelo** compuesto por 50 RNA del tipo *feed-forward neural network* con igual estructura global pero diferente parametrización. Entonces, surge un primer cuestionamiento sobre la **manipulación de datos** que servirán para su entrenamiento, para lo cual nosotros empleamos la técnica de agrupamiento (*clustering*) conocida como *Self-Organizing Map* tal como lo propuso Anctil y Lauzon (2004). Posteriormente, la **selección de variables de entrada del modelo** es llevada a cabo con un algoritmo evolutivo multi-objetivo denominado Algoritmo

Genético de Clasificación No-dominada II (Non-dominated Sorting Genetic Algorithm II, NSGA-II), lo cual corresponde a un **acoplamiento neuro-evolutivo**. En esta etapa de selección, un **marco multi-objetivo de optimización** explora el compromiso entre tres criterios de error con prioridades independientes en función del promedio, los eventos mínimos y los eventos máximos de los hidrogramas. Los detalles sobre el protocolo de experimentación se muestran en la sección 2.4.

## 2.2. Caso de estudio, cuenca del río Risaralda

El área de estudio corresponde a la cuenca del río Risaralda (Colombia) con sus principales afluentes. Esta zona es de interés porque es un sector de vital importancia para la economía y desarrollo de la región, con grandes extensiones de cultivos de caña, explotación de material del río y pastizales, los cuales generan grandes problemas de erosión y de contaminación. La gestión ambiental de la cuenca es compartida por los departamentos de Caldas y Risaralda con un 40 % y 60 % respectivamente [EAIC, 1996]. En la Figura 2.2 se muestra la localización de la cuenca y su variabilidad altimétrica.

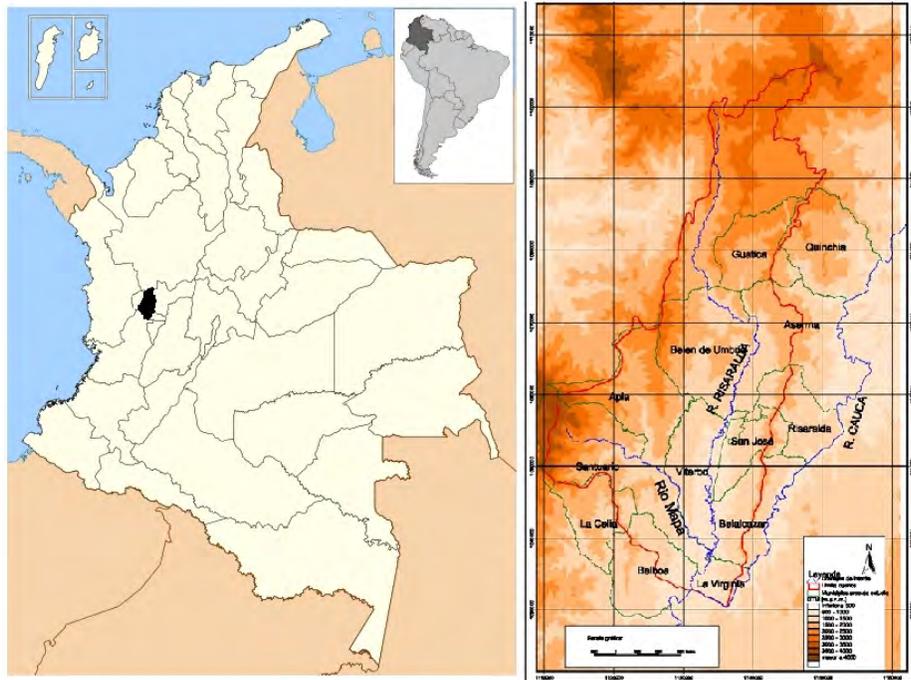


Figura 2.2: Zona de estudio. Fuente: propia.

La cuenca del río Risaralda cubre un área aproximada de 1469 km<sup>2</sup> con una longitud de 110 km desde el Alto de Paramillo donde se desprenden las quebradas Mampay, San Luis, Arrayanal y Dosquebradas, que forman el río y posteriormente desembocan en el río Cauca a la altura del municipio de La Virginia [CARDER, 1996].

El clima en los valles interandinos de los ríos Risaralda y Mapa es cálido y húmedo, conforme a las características climáticas enunciadas en las Agendas para la Gestión Ambiental de los municipios de influencia, tiene una temperatura media máxima de 24°C en la parte baja del valle del río Risaralda, al tiempo que se registra una temperatura mínima de 6°C en el cerro de Tatamá.

La distribución de las lluvias a través del año presenta dos periodos de relativa sequía y dos periodos húmedos. Los periodos secos se presentan de diciembre a marzo y de julio a agosto, siendo el primero el de sequía más pronunciada. Los periodos húmedos se presentan en los meses restantes. Esta distribución corresponde a un régimen de precipitación tipo ecuatorial bimodal influenciado por la zona de confluencia inter-tropical, ZCIT. Es importante mencionar que las precipitaciones experimentan fuerte variación interanual [CARDER, 1996]. En la Figura 2.3 se muestran las cuatro estaciones pluviométricas utilizadas en este estudio, así como la variabilidad espacial modelada de la precipitación media anual [CARDER, 2005].

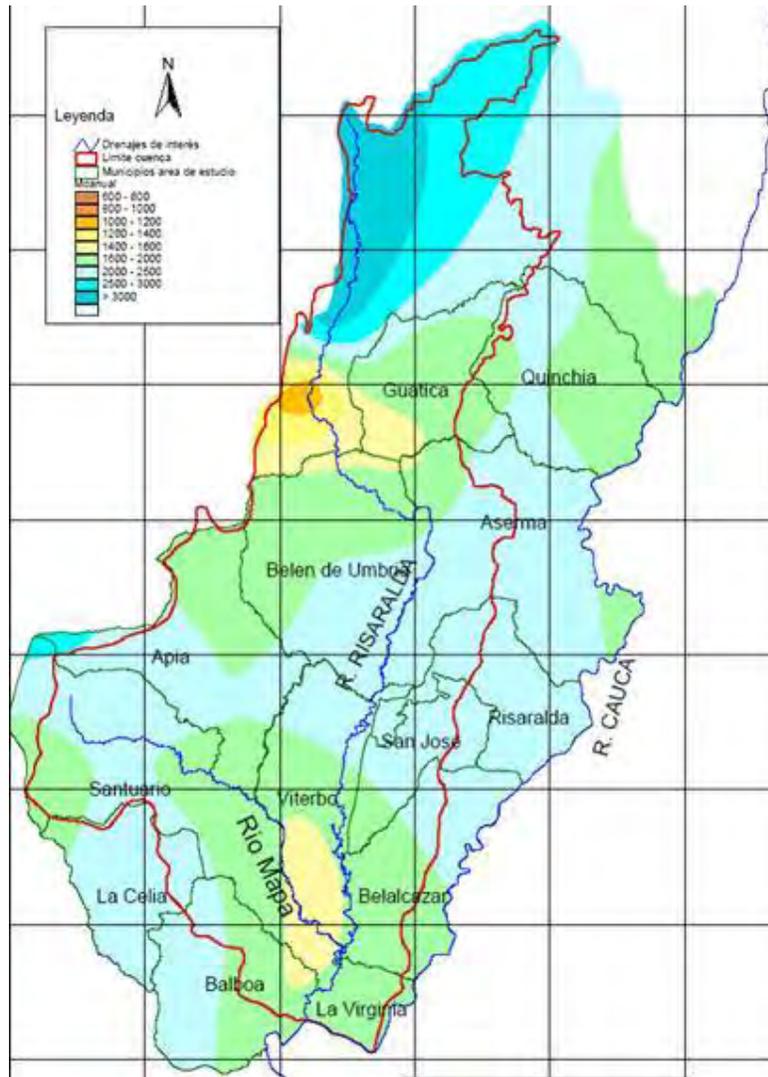


Figura 2.3: Precipitación media anual modelada. Fuente: propia.

El río Risaralda nace en el alto Morro Plancho, en los límites entre Caldas y Antioquia, sobre los 3.750 m.s.n.m. en jurisdicción del municipio de Riosucio (Caldas) y desemboca en el río Cauca frente a la cabecera municipal de La Virginia a los 896 m.s.n.m. Los principales ríos que componen la cuenca son: El Mapa, Totuí, Apía, Guarne y Guática. La cuenca del río Risaralda está separada de la cuenca del San Juan por el cordón montañoso denominado cuchilla del San Juan, de naturaleza volcánica, que se origina en el nudo o alto de los mellizos y se extiende hasta el valle del Risaralda. Se trata de una cuenca muy particular ya que la dirección de la red de drenaje es de norte a sur, similar a la del río Cauca donde desemboca, pero en sentido de flujo contrario, lo cual evidencia la presencia de fallas geológicas que condicionan

la red hidrográfica. El tiempo de concentración es de aproximadamente 10 horas, evaluado como el promedio entre las formulaciones de Ramser-Kirpich, Temez, Giandiotti y Bureau of Reclamation [Suárez, 2008], [Vélez and Botero, 2011]. Para este estudio se dispone de una serie histórica de caudales diarios observados en la estación “Puente Negro” (Figura 2.3). El periodo de análisis se extiende desde el 1 de enero del 2001 hasta el 2 de agosto del 2003.

## 2.3. Descripción de herramientas

Las herramientas empleadas para realizar la predicción de caudales son un modelo hidrológico semi-distribuido previamente calibrado y validado, las RNA que entrenan y evalúan un conjunto de parámetros y los AE como el NSGA-II que permiten optimizar el proceso de búsqueda de parámetros óptimos. Todos ellos se exponen a continuación.

### 2.3.1. Modelo hidrológico TETIS

El modelo hidrológico distribuido de tipo conceptual TETIS, divide la cuenca en celdas rectangulares interconectadas entre sí siguiendo la red de drenaje. En cada celda se modela el proceso lluvia-escorrentía usando seis tanques verticales según las indicaciones de la Figura 2.4, donde cada tanque representa el almacenamiento del agua debido a la nieve, los almacenamientos atmosféricos, el almacenamiento superficial, los almacenamientos gravitacionales, al acuífero y al cauce. La relación entre los distintos tanques se basa en las características del suelo tales como capacidad del campo y las conductividades hidráulicas saturadas del suelo y el subsuelo. Estas características del suelo han sido estimadas previamente en cada celda usando información de tipo ambiental disponible que sirve para caracterizar el suelo como: litología, edafología, usos del suelo y características geomorfológicas.

Las conexiones verticales entre los tanques describen los procesos de: precipitación (lluvia y/o nieve), fusión de la nieve, evapotranspiración, infiltración y percolación. Las conexiones horizontales describen el flujo superficial, el flujo sub-superficial y el flujo base. Finalmente, la salida de las aguas subterráneas se considera en el tanque más bajo para cerrar el balance de agua en lo que se ha denominado las pérdidas subterráneas. El movimiento horizontal del flujo está definido por un acoplamiento tridimensional de los tanques. La traslación del flujo a lo largo de la red de flujo hasta la salida de la cuenca se realiza empleando una combinación de las características geomorfológicas de la cuenca en combinación con el método de la onda cinemática geomorfológico [Vélez, 2001]. Más detalles sobre este modelo hidrológico puede ser consultados en [Vélez et al., 2009].

Los resultados simulados en esta cuenca obedecen a la calibración de los parámetros del modelo minimizando el error medio cuadrático con base en el SCE-UA.

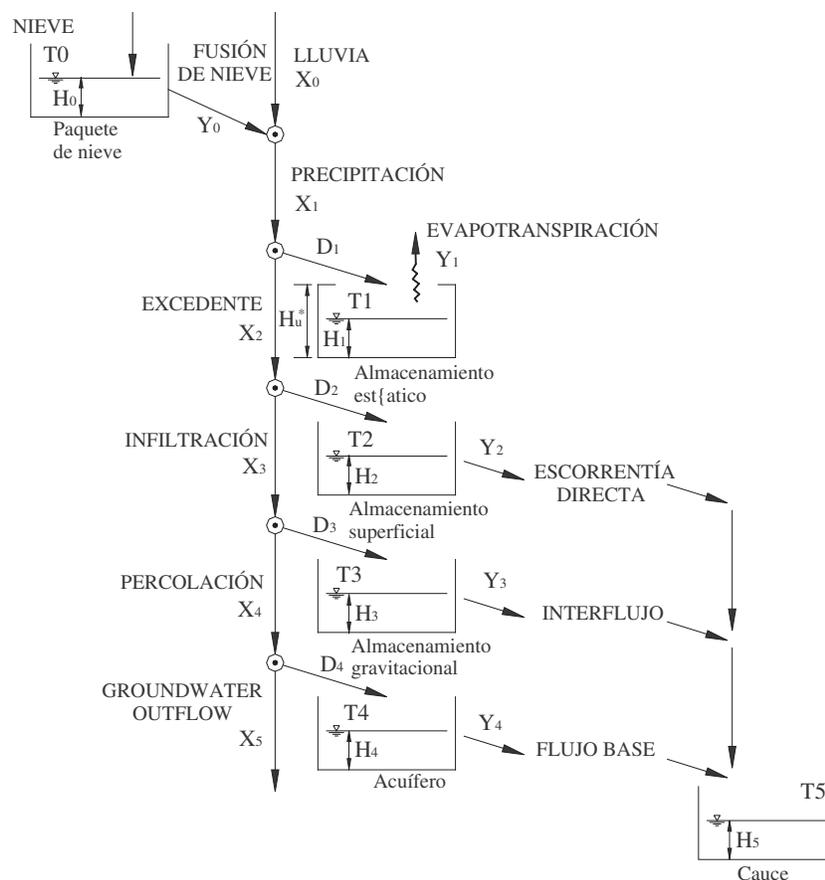


Figura 2.4: Modelo hidrológico TETIS. Fuente: [Vélez et al., 2009]

### 2.3.2. Redes neuronales artificiales (RNA)

En referencia a la estructura de las RNA, recientemente en un grupo experimental de 12 cuencas, [de Vos, 2012] presentó un análisis comparativo del rendimiento de los siguientes tipos de RNA:

- Red clásica de propagación hacia adelante (*feed-forward neural network*, FFNN).
- Redes recurrentes simples o tipo Elman.
- Redes completamente recurrentes.
- Redes con almacenamiento dinámico (*Echo State Network*).

En el anterior estudio se demostró principalmente el alto rendimiento de las redes con almacenamiento dinámico como resultado de la aplicación de técnicas dinámicas no lineales. No obstante, priorizando la sencillez y una fácil aceptación en la comunidad hidrológica, en este trabajo nosotros proponemos la evaluación de la topología comúnmente utilizada en recursos hidráulicos: una FFNN de tres capas formada por una capa de entrada, una capa oculta con una función de transferencia tipo sigmoide y una función de salida con una función de activación lineal [Maier and Dandy, 2000].

En este tipo de redes, cada capa está compuesta de unidades o neuronas y sus respectivos pesos. Cada neurona en la capa de entrada toma la información de las respectivas entradas del modelo multiplicada por los pesos individuales de cada neurona para posteriormente integrarla

por medio de una función de propagación (usualmente la sumatoria). Entonces, en la capa oculta, la función de activación o transferencia convierte el valor propagado (sumatoria) para que éste sea posteriormente procesado en la capa de salida. Finalmente, en la capa de salida el proceso de evaluación de pesos, propagación y función de activación se aplica nuevamente para obtener la salida del sistema.

La evaluación de pesos o parámetros, es decir el entrenamiento de la RNA, es ejecutada mediante un algoritmo de aprendizaje tal como la retro-propagación (*Back-propagation*, BP), el cual permite un aprendizaje controlado por el error entre la salida de la red y las observaciones usando técnicas de gradiente descendente. Muchas variantes de BP han sido propuestas para optimizar el tiempo de cálculo y permitir la búsqueda del mínimo global, por ejemplo el algoritmo Levenberg-Marquardt es reconocido por su convergencia rápida y estable, además de su mayor fiabilidad con respecto a otras variantes de BP [Hagan et al., 1996]. La aplicación de las técnicas BP requiere la minimización de una función de error diferenciable como el error medio cuadrático. Tanto en el campo de las RNA como en la calibración de modelos hidrológicos, y de forma general en métodos basados en la minimización de una función de error, un problema fundamental es la pérdida de generalización debido al sobre-entrenamiento (*overfitting*). En este caso los modelos son llevados a un estado de “perfección ficticia” donde la salida del sistema replica incluso el ruido inherente a la señal evaluada, evitando así una simulación acertada ante observaciones no incluidas en el proceso de entrenamiento.

Dada la importancia de tal fenómeno, considérese la Figura 2.5, donde se presenta un ejemplo sobre la generalización, el sesgo y la varianza. Aquí, los datos de entrenamiento son subdivididos en calibración (círculos blancos) y validación (rectángulos negros). Los datos provienen de un modelo parabólico con un ruido añadido a la señal. Para este análisis se muestran dos ensayos de regresión correspondientes a distintas selecciones de datos para la regresión con cada uno de los polinomios evaluados. Los resultados muestran que el modelo más complejo (orden 10) permite un ajuste perfecto a los datos de calibración en los dos ensayos (sesgo aparentemente bajo, sobre-entrenamiento), pero cambios pequeños en los datos elegidos causan grandes cambios en los polinomios finales (varianza elevada). En contraposición, el modelo de segundo grado no presenta un ajuste perfecto a los datos, los cuales naturalmente contienen ruido, pero es notoriamente más estable (baja varianza). De otra parte, un modelo de un sólo parámetro (polinomio de grado cero) muestra un alto sesgo con baja varianza (sub-entrenamiento, *underfitting*). El modelo óptimo es aquel que tiene el mejor balance entre el sesgo y la varianza, lo cual se traduce en una capacidad adecuada de generalización [Alpaydin, 2010].

En general, dados varios modelos con un error comparable, el modelo más simple debería generalizar mejor. Este principio es conocido como *Navaja de Occam* o principio de parsimonia. A pesar de esto, es importante notar que no hay necesariamente una relación entre el número de parámetros y la tendencia al sobre-entrenamiento, por ejemplo los sistemas multi-modelo muestran cómo se puede lograr un modelo “suavizado” con funciones de combinación de modelos o el caso de las máquinas de vectores de soporte (*support vector machines*), las cuales pueden tener un número infinito de parámetros sin sobre-entrenar el modelo [Domingos, 2012].

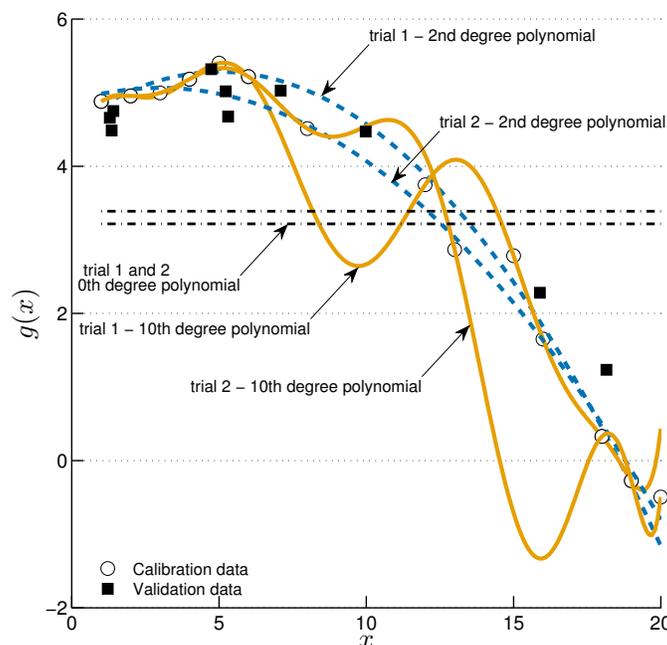


Figura 2.5: Problema de generalización. Fuente: propia.

En el caso de las RNA varios mecanismos pueden reducir el riesgo del sobre-entrenamiento, tales como el clásico *early stopping* o la regularización Bayesiana [Ancitil and Lauzon, 2004]. En el caso del método *early stopping*, es necesario definir dos subconjuntos a partir del conjunto de entrenamiento: uno de estimación y otro de validación. El subconjunto de estimación es usado para buscar iterativamente los parámetros del modelo, mientras el subconjunto de validación es utilizado para supervisar el error de tal forma que si en iteraciones sucesivas comienza a incrementarse, el proceso de entrenamiento es suspendido.

Otro de los interrogantes, tanto en RNA como en los modelos orientados por datos, es la selección de las variables de entrada del modelo matemático que representará el sistema. En el contexto de la selección de variables de entrada, en la comunidad de aprendizaje automático se destacan dos grupos de técnicas: tipo filtro en la cual la selección es independiente del modelo matemático, y tipo *wrapper* donde se usa iterativamente el modelo matemático para analizar la sensibilidad de las variables.

En el caso de las técnicas tipo filtro, varios autores [Bowden et al., 2005, Brown, 2009] han mostrado la eficiencia de un método basado en la información mutua promedio conocida también como correlación no lineal. En contraste, los métodos tipo *wrapper* se reconocen por su alto costo computacional pero también por su alta eficiencia. En este contexto se ha mostrado en diversos escenarios la eficiencia de un método simple e intuitivo conocido como *Forward Greedy Selection* [Ancitil et al., 2009, Brochero et al., 2012, Brochero et al., 2013], en el cual se añade iterativamente la variable que más impacto tiene en la reducción del error, quedando a decisión del modelador la cantidad de variables necesarias para representar adecuadamente el sistema.

Se puede añadir una nueva fuente de complejidad al problema de selección de variables si se tiene en cuenta que la optimización puede centrarse en múltiples objetivos simultáneamente. De tal forma, el planteamiento propuesto reside en la búsqueda de “buenos compromisos” entre los objetivos más que la minimización individual de cada uno de ellos. Así, en lugar de tener una solución única, la optimización multi-objetivo da lugar a un conjunto de soluciones óptimas llamado conjunto o frente de Pareto (*Pareto front*), en el cual cada solución individual representa la imposibilidad de mejorar un objetivo sin degradar el rendimiento de otro criterio (Figura 2.6).

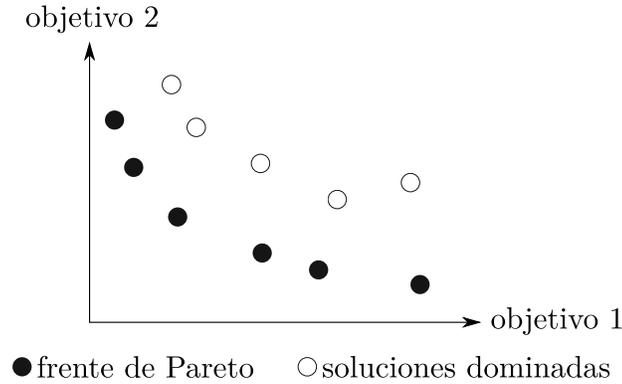


Figura 2.6: Frente de Pareto. Fuente: propia.

En este marco multi-objetivo, en el campo de los AE es reconocida la eficiencia del NSGA-II propuesto por [Deb et al., 2002] y expuesto brevemente a continuación.

### 2.3.3. NSGA-II

Los algoritmos genéticos, o más generalmente AE, surgen como una analogía matemática al proceso de evolución y supervivencia de las especies propuesto por C. Darwin. De tal forma, en un ambiente dado, los mejores “individuos” tienden a vivir lo suficiente para reproducirse mientras los más débiles tienden a desaparecer. En los AE un conjunto de soluciones candidatas, llamado población de individuos, es evaluada con respecto a una función de error o de ajuste (*fitness*), la cual permite medir la calidad de cada individuo. Posteriormente los mejores individuos tienen mayor probabilidad de ser seleccionados para generar una nueva generación de soluciones codificadas debido a los procesos genéticos de cruce y mutación. Este procedimiento continúa hasta que algún criterio de parada se cumpla, por ejemplo el número máximo de generaciones. Más detalles sobre la representación, los operadores de mutación y cruce, los modelos de población y las técnicas de selección pueden ser encontrados en [Eiben and Smith, 2003].

En el caso especial de los AE en un contexto multi-objetivo, el NSGA-II usa un algoritmo genético para la evolución de una población, en combinación con una selección basada en la clasificación de soluciones no dominadas (Figura 2.6) y una distancia de grupos (*crowding distance*) para preservar la diversidad en la población. Los pasos básicos de este algoritmo se resumen a continuación:

- Una población de padres  $M_t$ , de tamaño  $N$ , y una población de hijos  $H_t$ , de tamaño  $N$ , se combinan para formar una población ( $R_t = M_t \cup H_t$ ). Este mecanismo asegura el elitismo en las generaciones.
- La población  $R_t$  es ordenada de acuerdo a un criterio de soluciones no dominadas para identificar diferentes frentes  $F_1, F_2$ , etc.
- Los mejores individuos formarán el primer frente. Una nueva población  $M_{t+1}$  es generada añadiendo los frentes completos si el tamaño no excede a  $N$ .
- Si el número de individuos presentes en  $M_{t+1}$  es menor a  $N$ , un procedimiento basado en la distancia de grupos es aplicado en el primer frente siguiente  $F_i$ , no incluido en  $M_{t+1}$ . El propósito de este operador es insertar los  $N - M_{t+1}$  mejores individuos necesarios en la población  $M_{t+1}$ .
- Una vez los individuos en la población  $M_{t+1}$  son identificados, una nueva población de hijos  $H_{t+1}$  es creada por medio de la selección de padres, el cruce y la mutación.

- La selección de padres usa una versión modificada de torneo que considera primero el rango de dominación y posteriormente la distancia de grupo.
- El proceso continúa de una generación a otra hasta que se cumpla el criterio de parada.

## 2.4. Protocolo de experimentación

En esta sección se definen los detalles de un método de predicción de caudales con RNA basado en la selección de variables de entrada. La novedad del planteamiento presentado proviene del contexto multi-modelo de predicción y del marco multi-objetivo de la selección de variables. De tal forma, en la primera parte se muestra la forma en la que se construye la tabla de entradas candidatas que es la base de la información del sistema. Posteriormente se presenta la definición del sistema de predicción con RNA, desde la selección de datos hasta la definición del entrenamiento y la estructura del sistema, para finalmente indicar el acoplamiento del modelo de predicción con el algoritmo de optimización multi-objetivo NSGA-II.

### 2.4.1. Sistema de información de variables de entrada

La base del modelo la constituye el sistema de información que contiene las variables consideradas importantes en la representación del sistema de predicción. En consecuencia se tienen 16 variables hidroclimatológicas y sus correspondientes rezagos, lo cual nos conduce a una tabla de 71 variables candidatas.

En la Tabla 2.1 se muestra la configuración del sistema de información diferenciando las variables entre tipo medición, extrapolación o simulación hidrológica. Además de una observación que en general denota la zona de la cuenca de la observación, el método de extrapolación y el tipo de respuesta hidrológica en el caso de las variables simuladas con el modelo hidrológico TETIS. Finalmente, en la última columna se presentan los rezagos considerados para la construcción de la tabla de entradas candidatas. En general se evalúa el impacto de los últimos cinco días para la predicción excepto para las variables que representan una respuesta hidrológica muy rápida como la escorrentía, o una respuesta lenta como el flujo de base o muy lenta como el almacenamiento subterráneo.

La configuración del sistema de información muestra nuestro interés en presentar al modelo de predicción la precipitación local en contraste con la precipitación promedio simulada con el método de polígonos de Thiessen, la cual también hace parte del sistema de información. Adicionalmente la temperatura media y el caudal en la salida de la cuenca son también considerados. Asimismo, tanto la evapotranspiración simulada con el método de Thornthwaite como la calculada internamente en el modelo TETIS constituyen la tabla de entradas candidatas. Finalmente, se tienen los caudales y almacenamientos modelados en TETIS para las capas superficial, sub-superficial y subterránea. Vale la pena anotar que aunque en principio la simulación hidrológica del río Risaralda se había subdividido en once subcuencas, en este estudio se trabajaron condiciones medias para toda la cuenca llevando el modelo a un estado global en lugar de una simulación distribuida.

### 2.4.2. Modelo base de predicción

Como se ha descrito anteriormente, el modelo base de predicción consiste en un conjunto de 50 RNA del tipo FFNN con una sola capa oculta. A continuación se presentan los detalles de este modelo.

Tabla 2.1: Tabla de 71 entradas candidatas para la selección de variables.

<b>Variable</b>	<b>Observación</b>	<b>Rezagos</b>
	Mediciones	
Precipitación, $P_1$	Zona media de la cuenca	1, 2, 3, 4, 5
Precipitación, $P_2$	Zona alta de la cuenca	1, 2, 3, 4, 5
Precipitación, $P_3$	Zona media de la cuenca	1, 2, 3, 4, 5
Precipitación, $P_4$	Zona baja de la cuenca	1, 2, 3, 4, 5
Caudal, $Q_o$	Zona baja de la cuenca	1, 2, 3, 4, 5
	Extrapolación de mediciones	
Precipitación, $\bar{P}$	Mét. de Thiessen	1, 2, 3, 4, 5
Temperatura, $\bar{T}$	Promedio de las mediciones	1, 2, 3, 4, 5
Evapotranspiración, E	Mét. Thornthwaite y promedio	1, 2, 3, 4, 5
	Simulación con modelo hidrológico TETIS	
Caudal, $Q_s$	Zona baja de la cuenca	1, 2, 3, 4, 5
Evapotranspiración, $y_1$	Respuesta hidrológica rápida	1, 2, 3, 4, 5
Escorrentía, $y_2$	Respuesta hidrol. muy rápida	1
Caudal subsuperficial, $y_3$	Respuesta hidrológica rápida	1, 2, 3, 4, 5
Caudal de base, $y_4$	Respuesta hidrológica lenta	1
Almacenamiento estático, $H_1$	Respuesta hidrológica rápida	1, 2, 3, 4, 5
Almacenamiento gravitacional, $H_3$	Respuesta hidrológica rápida	1, 2, 3, 4, 5
Almacenamiento subterráneo, $H_4$	Respuesta hidrol. muy lenta	5, 10, 15, 20

#### 2.4.2.1. Definición de los subconjuntos de entrenamiento y prueba

La base de datos de este estudio corresponde a las observaciones y la simulación hidrológica del modelo TETIS entre el 1 de enero del 2001 y el 2 de agosto del 2003. En lo correspondiente a los conjuntos de entrenamiento y prueba, nosotros adoptamos una separación simple de datos con base en una fecha de corte (*cut-off sampling*), en este caso el 3 de agosto de 2002, por lo cual el entrenamiento se considera como la primera parte de la serie mientras la prueba se configura con la parte restante (Figura 2.7).

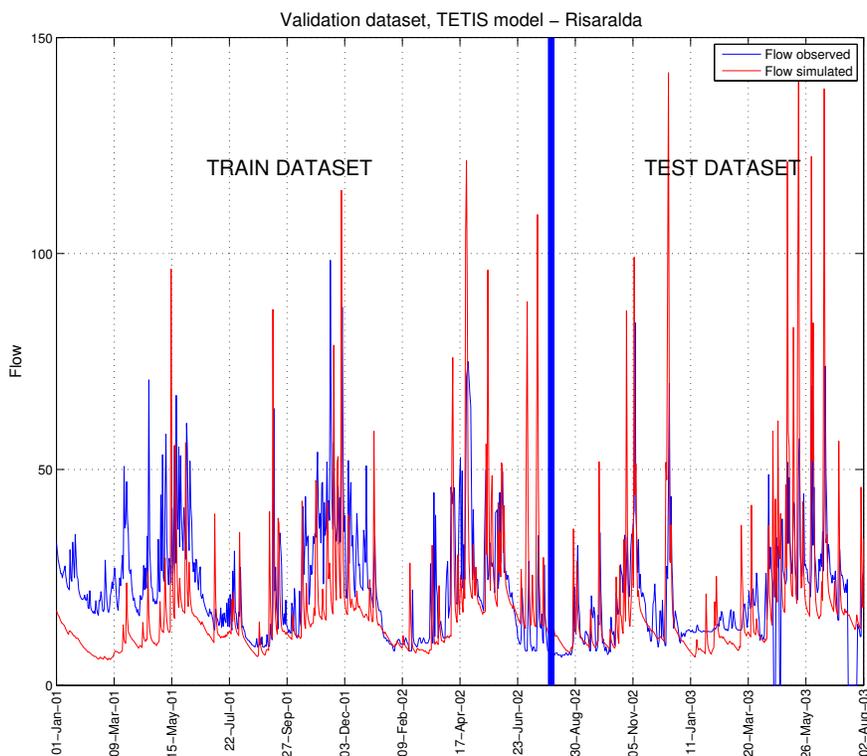


Figura 2.7: Observaciones y simulación hidrológica en el periodo de estudio.

La elección de este método de remuestreo se vio influenciada por su popularidad en la comunidad hidrológica donde frecuentemente los modelos requieren de un periodo de tiempo continuo para su calibración y funcionamiento. Adicionalmente, una simple inspección de la homogeneidad de los estadísticos básicos de las series puede aceptarse como un criterio suficiente para juzgar la representatividad de ambos subconjuntos tal como se muestra en la Tabla 2.2.

Tabla 2.2: Estadísticos subconjuntos de entrenamiento y prueba.

	Variables provenientes de mediciones hidroclimatológicas							
	$P_1$	$P_2$	$P_3$	$P_4$	$\bar{P}$	$\bar{T}$	EVT	$Q_o$
Media (entren.)	4.18	4.30	5.69	4.11	4.59	27.72	3.29	24.30
Media (test)	3.92	4.61	5.51	3.68	4.45	27.85	3.29	19.65
Desv. (entren.)	7.96	7.43	10.80	8.01	5.94	1.64	0.05	14.92
Desv. (test)	7.58	9.22	11.34	7.42	7.23	1.98	0.08	11.31
	Variables provenientes de la simulación hidrológica con TETIS							
	$Q_s$	$y_1$	$y_2$	$y_3$	$y_4$	$H_1$	$H_3$	$H_4$
Media (entren.)	21.04	2.37	0.56	1.50	0.06	229.83	43.40	11.11
Media (test)	19.84	2.39	0.72	1.23	0.05	233.55	35.71	8.94
Desv. (entren.)	16.54	0.24	1.38	0.57	0.01	37.85	16.63	1.04
Desv. (test)	19.64	0.15	1.82	0.36	0.00	27.59	10.53	0.35

Este tipo de remuestreo al menos garantiza la independencia temporal de la información entre los subconjuntos. Hay que recordar que las RNA son reconocidas básicamente por su capacidad de interpolación, es decir, si el conjunto de entrenamiento no contiene eventos máximos, una RNA clásica no será capaz de simular correctamente este tipo de eventos [Hettiarachchi et al., 2005].

De otra parte, en este trabajo se adopta la técnica conocida como *early stopping* para mejorar la generalización de la RNA, por lo cual se hace necesario construir a partir del conjunto de entrenamiento dos nuevos subconjuntos de **estimación** y **validación** de parámetros. Para esto se ha conformado un conjunto de estimación equivalente al 70 % de los datos del entrenamiento y el 30 % restante configura la validación del sistema. La estrategia utilizada en la conformación de estos subconjuntos se basa en el planteamiento de [Anctil and Lauzon, 2004]. En consecuencia los vectores de entrada son clasificados con una red de Kohonen o *Self-Organized Map* empleando una topología de malla de 3x3. Posteriormente, los datos de cada clase son distribuidos aleatoriamente en los conjuntos de estimación y validación preservando los porcentajes previamente definidos.

#### 2.4.2.2. Pre y post-procesamiento

Se lleva a cabo una estandarización de los datos para garantizar que las entradas y salidas del sistema estén en la misma escala. Así, cada variable transformada tiene una media igual a cero y una desviación estándar unitaria. Finalmente, la salida de la RNA debe ser inversamente transformada para obtener la predicción en la escala real del problema.

#### 2.4.2.3. Arquitectura de la RNA

Como se introdujo anteriormente, en nuestro modelo de predicción básico planteamos la utilización de una RNA simple del tipo FFNN con una capa oculta usando el algoritmo de entrenamiento conocido como Levenberg Marquardt *Back-Propagation*.

El número de neuronas fue optimizado por ensayo y error utilizando un esquema inicial de cuatro entradas predefinidas ( $\bar{P}_{t-2}, \bar{P}_{t-1}, Q_{t-2}, Q_{t-1}$ ) para la predicción del caudal en el tiempo "t" ( $Q_t$ ). Los resultados mostraron que cinco neuronas son suficientes, puesto que no se obtiene una ganancia significativa al utilizar más neuronas. En la prueba anterior, al igual que en el resto de las evaluaciones presentadas aquí, la capa oculta tiene una función de activación tipo tangente hiperbólica mientras que la capa de salida se configura con una función de activación tipo lineal.

Adicionalmente, el entrenamiento es ejecutado por lotes *batch*, los pesos son inicializados siguiendo el procedimiento de Nguyen-Widrow [Nguyen and Widrow, 1990], por lo cual las inicializaciones contienen un cierto grado de aleatoriedad. Asimismo, una tasa adaptativa de aprendizaje es inicializada como 0.005. Este valor es multiplicado por 0.1/10 si la función de error disminuye/aumenta en una época.

Con respecto al número máximo de épocas, es decir iteraciones para calibrar los pesos, éste es inicialmente configurado como 50; sin embargo, es el método *early stopping* el que generalmente gobierna el entrenamiento. Las pruebas fueron ejecutadas con el *Neural Networks Toolbox* 7 de Matlab. El lector puede encontrar información detallada sobre la manipulación de los distintos parámetros de una RNA tipo FFNN en [Hagan et al., 1996] o en [Hudson et al., 2012].

### 2.4.3. Definición del espacio de entrada con NSGA-II

La idea central de este estudio se basa en la selección de entradas para el modelo de predicción de RNA con NSGA-II. En este caso, la idea del ensamble neuro-evolutivo con NSGA-II es evaluar simultáneamente la relación entre distintos subconjuntos de variables con diversos objetivos o funciones de error.

Aunque el algoritmo NSGA-II puede utilizarse con múltiples objetivos simultáneos independientemente de la escala y naturaleza de cada uno de ellos, se separa la evaluación de las variables de entrada del modelo de predicción en dos fases: la primera centrada en la determinación del número de variables de entrada; la segunda, en la selección final de entradas con base en la evaluación del compromiso entre tres funciones de error.

Como cualquier AE, su configuración está dada principalmente por la forma de representar las soluciones potenciales, la evaluación de la función objetivo y los mecanismos de evolución. A continuación se presenta la configuración de estas características en cada una de las fases del acoplamiento neuro-evolutivo. Posteriormente, se presentan los demás aspectos comunes en ambas fases del acoplamiento.

#### 2.4.3.1. Fase I: Evaluación del número de entradas

**Representación:** Cada solución potencial está definida por una cadena de 71 alelos de representación binaria, donde el valor unitario representa la inclusión de la respectiva variable de la tabla de entradas candidatas, y cero su exclusión.

**Función de evaluación:** Como se mostró en la sección 2.3.3, se define una distancia de grupo para cada punto como la longitud promedio del cubo definido por los vecinos más próximos en el mismo frente de Pareto. Un valor grande indica una baja densidad de soluciones en el punto evaluado. La distancia de grupo en este caso es definida por el número de variables seleccionadas y el error medio cuadrático (MSE).

**Mecanismos de evolución:** Luego de la selección de los padres se procede a la definición del operador de cruce, en este caso utilizamos un cruce basado en un solo punto de intercambio genético conocido como *one-point crossover*. Este método consiste en la elección aleatoria de un punto de cruce dentro del cromosoma para posteriormente intercambiar el material genético a partir de este punto. La tasa de cruce fue establecida en 0.9 (90%). Otro operador, conocido como mutación, representa la probabilidad de mutar aleatoriamente la información de un alelo. En general esta probabilidad es muy baja para evitar inestabilidad en la evolución. Aquí fue definido un operador de mutación conocido como *bitwise* con una tasa del 2%, es decir que con una probabilidad del 2% la información binaria de un alelo será mutada sea de cero a uno o viceversa.

#### 2.4.3.2. Fase II: Evaluación de la selección final de entradas

**Representación:** El cromosoma está formado por 71 alelos que toman valores enteros entre 1 y 71 evitando que dos alelos tomen valores iguales, lo cual se conoce como una representación por permutaciones. El valor de cada alelo representa una de las 71 variables de la tabla de entradas candidatas. No obstante, la solución se verá únicamente representada por los primeros ocho alelos del cromosoma en lo que resulta como una permutación truncada, decisión que se basa en los resultados de la primera fase de evaluación del número de variables (Sección 2.5.1). Los 63 alelos restantes son reservados para la aplicación de los mecanismos de evolución.

**Función de evaluación:** La distancia de grupo está definida por tres funciones de error negativamente orientadas, es decir que se espera la minimización de cada objetivo. Las siguientes

ecuaciones definen las funciones de error, donde  $Q_{o_t}$  y  $Q_{s_t}$  representan el caudal observado y el simulado en el tiempo “t” respectivamente, mientras  $Q_o$  indica el caudal medio observado. La primera función (Ecuación 2.1) representa el conocido criterio de Nash-Sutcliffe pero orientado en minimización, el cual penaliza los caudales máximos. La Ecuación 2.2 representa el anterior criterio pero calculado sobre la transformación logarítmica de los caudales, lo cual penaliza la predicción de caudales bajos. Finalmente, se tiene el criterio de balance o control de volumen de agua (Ecuación 2.3). La utilización de estas funciones está inspirada en el estudio multi-objetivo del modelo hidrológico GR4J propuesto por [Perrin et al., 2003].

$$C_1 = \frac{\sum_{t=1}^n (Q_{o_t} - Q_{s_t})^2}{\sum_{t=1}^n (Q_{o_t} - Q_o)^2}, \quad (2.1)$$

$$C_2 = \frac{\sum_{t=1}^n (\ln Q_{o_t} - \ln Q_{s_t})^2}{\sum_{t=1}^n (\ln Q_{o_t} - \ln Q_o)^2}, \quad (2.2)$$

$$C_3 = \left| \sqrt{\frac{\sum_{t=1}^n Q_{s_t}}{\sum_{t=1}^n Q_{o_t}}} - \sqrt{\frac{\sum_{t=1}^n Q_{o_t}}{\sum_{t=1}^n Q_{s_t}}} \right|. \quad (2.3)$$

**Mecanismos de evolución:** Dado que la representación es básicamente una permutación, el operador de cruce utilizado se conoce como *Partially Mapped Crossover* propuesto por [Whitley, 2000]. Para esta aplicación uno de los puntos de cruce es elegido entre los alelos 1 y 7. En cuanto al operador de mutación, nosotros usamos el método *swap* [Eiben and Smith, 2003], donde la información de un alelo elegido aleatoriamente es intercambiada con la de un segundo alelo elegido de igual forma. En esta aplicación la primera posición corresponde a uno de los primeros 8 alelos y la segunda a uno de los 63 alelos restantes en el cromosoma. Al igual que en la anterior fase, la tasa de cruce y mutación es del 90% y 2% respectivamente.

#### 2.4.3.3. Elementos comunes en la configuración del AE en ambas fases

**Población y mecanismo de selección de padres:** La población fue establecida en 100 individuos. En cada generación se tienen dos hijos fruto del cruce de dos padres cuando este último tiene lugar. La selección de padres utiliza un mecanismo de torneo (*tournament operator*) en el cual se considera primero el rango de dominio y posteriormente la distancia de grupo. Para el torneo se eligen aleatoriamente cuatro individuos de los cuales los dos de mayor calidad son seleccionados para el cruce.

**Inicialización:** La generación inicial se elige de forma aleatoria.

**Supervivencia:** Las poblaciones de padres e hijos son mezcladas y se determinan los frentes de Pareto. La nueva población de 100 individuos es configurada con la inserción progresiva de los individuos del mejor frente. Si no pueden aceptarse todos los individuos del último frente, ellos serán elegidos con base en la distancia de grupo.

**Condición de parada:** se define 100 generaciones como criterio de parada.

**Análisis del frente de Pareto:** Posterior a la selección de los puntos no dominados del último frente de Pareto, se determinan los centros de los grupos (*clusters*) de puntos que son representativos del frente. Para esto se sigue la metodología propuesta por [Chaudhari et al., 2010], la cual consiste en las siguientes etapas:

- Obtener el conjunto de puntos que representan el frente óptimo de Pareto.
- Evaluar el agrupamiento de datos con *k-means*. Para esto, los objetivos son normalizados con media y varianza igual a cero y uno respectivamente. La disimilitud es medida con base en la distancia euclídeana. El número de grupos se define con base en la maximización del valor medio de la función llamada *silhouette*, la cual mide qué tan cerca están los puntos de un grupo en referencia a los grupos vecinos [Martinez Alvarez et al., 2011].
- Se elige como solución representativa de cada grupo el punto más cercano al centro del mismo.

Aunque el frente de Pareto nos da una información completa con respecto al compromiso de los distintos objetivos, la solución finalmente elegida será aquella que minimice todos los objetivos simultáneamente, correspondiente al grupo en la bifurcación de los objetivos (*knee cluster*). En la Tabla 2.3 se presenta un resumen de la configuración del NSGA-II en las dos fases de la selección de entradas.

Tabla 2.3: Configuración del NSGA-II para la selección de variables. Fuente: propia.

Fase	Número de variables	Selección final
Representación	Binaria	Permutaciones truncadas
Cruce	<i>One-point crossover</i>	<i>Partially mapped crossover</i>
Mutación	<i>Bitwise</i>	<i>Swap</i>
Probabilidad de cruce		90%
Probabilidad de mutación		2%
Selección de padres	Mejores 2 de 4 elegidos aleatoriamente (torneo)	
Supervivencia	Frentes de Pareto y distancia de grupo	
Población	100 individuos	
Hijos por cruce	2	
Inicialización	Aleatoria	
Condición de parada	100 generaciones	
Análisis post-Pareto	<i>k-means</i> con función <i>silhouette</i>	

## 2.5. Resultados y análisis

### 2.5.1. Fase I: Evaluación del número de entradas

En la Figura 2.8 se muestra la evolución del NSGA-II entorno al compromiso entre el número de entradas (implícitamente el tipo de entrada) y el MSE del sistema de predicción con RNA. Éste último es evaluado como el error general en entrenamiento, es decir, mientras se utilizan dos subconjuntos al interior de la RNA para detener la calibración de pesos evitando el sobreentrenamiento, el error se calcula sobre el conjunto total de datos de entrenamiento sin tener en cuenta la anterior subdivisión en conjuntos de estimación y validación. Aquí se puede visualizar el amplio rango de valores de error correspondiente a un mismo número de variables, lo cual nos indica la alta sensibilidad del modelo de predicción a la selección de variables, la cual es la hipótesis de partida de este estudio.

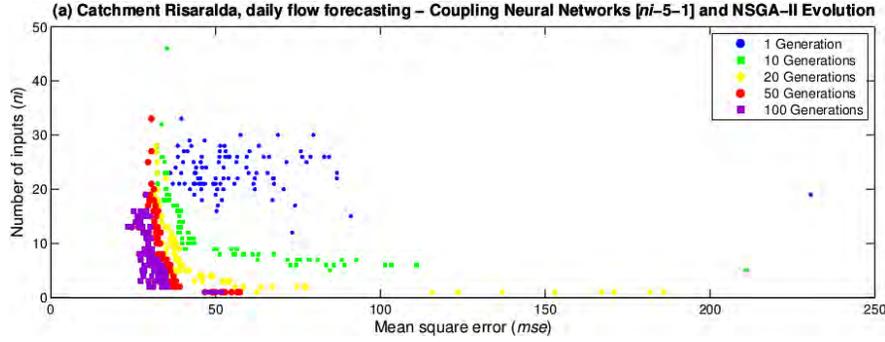


Figura 2.8: Relación entre el número de entradas y el MSE. Fuente: propia.

De otra parte, un análisis más detallado del último frente de Pareto nos muestra la relevancia de los cinco escenarios de selección presentados en la Tabla 2.4. Es importante destacar que el error en prueba no puede ser objeto de la selección del modelo, su carácter es solo informativo, puesto que la selección del modelo se debe centrar exclusivamente en los errores en la estimación y la validación.

El mejor esquema contempla ocho variables de entrada, en el cual se incluye la precipitación local en cada una de las tres zonas de la cuenca en lugar de la precipitación media. Asimismo la presencia del almacenamiento subterráneo ( $H_4$ ), cuya variabilidad es muy baja, puede indicar la posibilidad de optimizar aún más el sistema en referencia al número de entradas. Al respecto un análisis de la sensibilidad de variables en este subconjunto es una tarea fácil de plantear con un algoritmo tal como el *Forward greedy selection* utilizado ampliamente en RNA [Bowden et al., 2005]. De otra parte, se destaca una característica frecuentemente observada en los modelos de predicción hidrológica: la persistencia de las series de caudales, componente que se ve reflejado en la inclusión del caudal observado el día precedente ( $Q_{t-1}$ ) en todos los esquemas de selección. Obsérvese también en el esquema de ocho variables de entrada la inclusión del caudal simulado el día precedente a la predicción ( $Q_{s_{t-1}}$ ), lo cual puede ajustar la predicción en función del error observado en la simulación hidrológica.

Tabla 2.4: Subconjuntos de variables para el modelo de predicción con un conjunto de RNA. La denotación de las variables sigue la identificación dada en la Tabla 2.1. Adicionalmente las variables aparecen tabuladas para visualizar su participación en distintos esquemas de selección.

#	MSE			Variables
	Estim.	Valid.	Test	
1	79.72	190.26	64.23	$Q_{0_{t-1}}$
2	55.13	45.85	43.63	$Q_{0_{t-1}}, \bar{P}_{t-1}$
3	44.14	64.51	37.89	$Q_{0_{t-1}}, \bar{P}_{t-1}, Q_{0_{t-2}}$
8	44.41	35.94	33.58	$Q_{0_{t-1}}, P_{1_{t-1}}, P_{2_{t-1}}, P_{4_{t-1}}, \bar{P}_{t-2}, Q_{s_{t-1}}, H_{4_{t-15}}, H_{4_{t-5}}$
13	41.05	43.24	35.10	$Q_{0_{t-1}}, Q_{0_{t-2}}, P_{4_{t-1}}, Q_{s_{t-1}}, Q_{0_{t-3}}, Q_{s_{t-3}}, T_{t-1}, E_{t-1}, Y_{2_{t-1}}, Y_{3_{t-2}}, H_{1_{t-3}}, H_{3_{t-3}}, H_{3_{t-1}}$

En cuanto a los demás esquemas de selección, se puede resaltar la importancia de combinar la persistencia de las series con la precipitación como impulso del sistema. Así, es notorio el mejoramiento del rendimiento cuando en el esquema de dos variables se incluye la precipitación media, con una ganancia respecto a una sola variable de hasta el 76 % en validación y del 32 % en estimación y prueba. En referencia al esquema de tres variables, se destaca la inclusión del

caudal observado dos días antes, no obstante la ganancia obtenida con respecto a dos variables no es muy significativa en la estimación y la prueba, e incluso es negativa en la validación. De ahí la importancia del esquema de dos variables que llevan el modelo a un error aceptable con una estructura muy parsimoniosa.

Vale la pena anotar que la definición del número de neuronas de la RNA se basa en un modelo ensayo y error que contempla cuatro entradas predefinidas ( $\bar{P}_{t-2}, \bar{P}_{t-1}, Q_{t-2}, Q_{t-1}$ ), lo cual coincide aproximadamente con las variables destacadas en esta fase.

En referencia al modelo con trece variables de entrada, la primera observación se centra directamente en la pérdida de rendimiento en la validación con respecto al modelo de ocho entradas. Asimismo, el error en prueba presenta un ligero deterioro del orden del 5 %. No obstante, se destacan varias características en este modelo:

- Retiene siete variables del modelo hidrológico TETIS sin incluir el almacenamiento subterráneo incorporado en el modelo de ocho entradas.
- Considera únicamente la precipitación precedente de la zona baja de la cuenca.
- Incluye los tres caudales precedentes.
- Abarca la temperatura y la evapotranspiración precedentes ( $T_{t-1}, E_{t-1}$ ).

En síntesis, es importante resaltar que la ganancia media del modelo de ocho variables con respecto al de dos está alrededor del 20 %, por lo cual dependiendo del tipo de aplicación se debe juzgar la pertinencia de tal modelo. En nuestro caso, nosotros lo consideramos pertinente dado que la selección final, presentada en la siguiente fase, busca complementar la información observada de caudal y precipitación con la información del modelo hidrológico TETIS. En la Figura 2.9 se presenta la comparación de los caudales observados y las predicciones del modelo de 50 RNA con el esquema preliminar de ocho variables anteriormente mostrado.

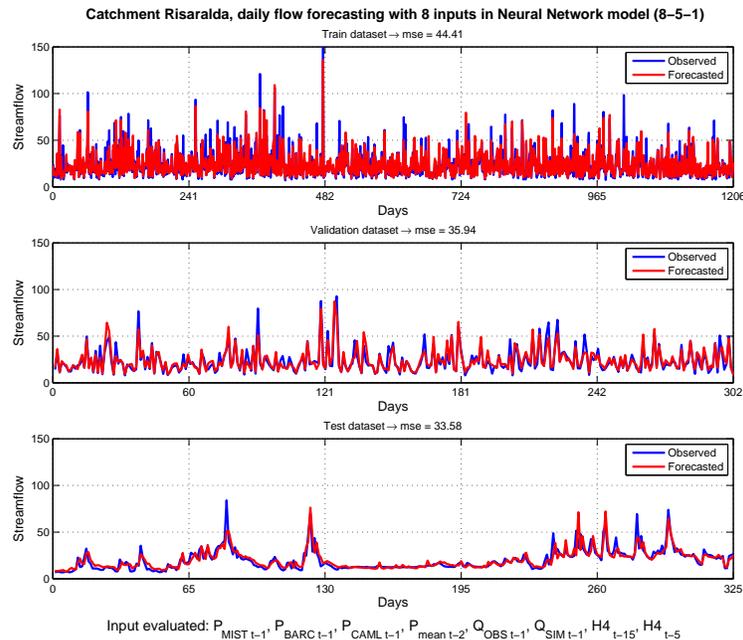


Figura 2.9: Predicción con 50 RNA y esquema de 8 entradas. Fuente: propia.

## 2.5.2. Fase II: Evaluación de la selección final de entradas

De los análisis de la fase anterior, se toman como punto de partida la selección de un esquema de ocho variables, el cual da flexibilidad para incluir las variables de la simulación hidrológica en el modelo de predicción. Sin embargo, en esta fase, la selección se ejecuta nuevamente en un marco multi-objetivo que penaliza en distintas medidas los caudales altos con el criterio de Nash-Sutcliffe, los caudales bajos por medio de la transformación logarítmica de caudales y, finalmente, el volumen de agua con el criterio de balance.

En la Figura 2.10 se presenta la relación entre los tres objetivos evaluados en el entrenamiento, donde cada punto representa diferentes subconjuntos de ocho variables como entradas al modelo de predicción. En el panel superior izquierdo se muestra claramente la evolución del AE en función de la minimización de los tres objetivos. Es importante resaltar que aunque el criterio de Nash-Sutcliffe está orientado frecuentemente hacia la maximización, nosotros lo orientamos negativamente (Ecuación 2.1). En las proyecciones bidimensionales se resaltan cuatro puntos que se destacan como los centros de los grupos (*clusters*) del análisis post-Pareto. En el panel superior derecho se expone una relación directa entre el criterio de Nash y su correspondiente transformación logarítmica, situación que no beneficia la construcción del frente de Pareto, pues en esencia se busca el compromiso antagónico entre los objetivos a minimizar. No obstante, tal característica se visualiza entre estos dos criterios y la función de balance de agua (conservación del volumen). Al respecto, los esquemas de selección representativos son determinados en el contexto tridimensional del problema, razón por la cual en los planos bidimensionales los puntos están aparentemente fuera del último frente de Pareto.

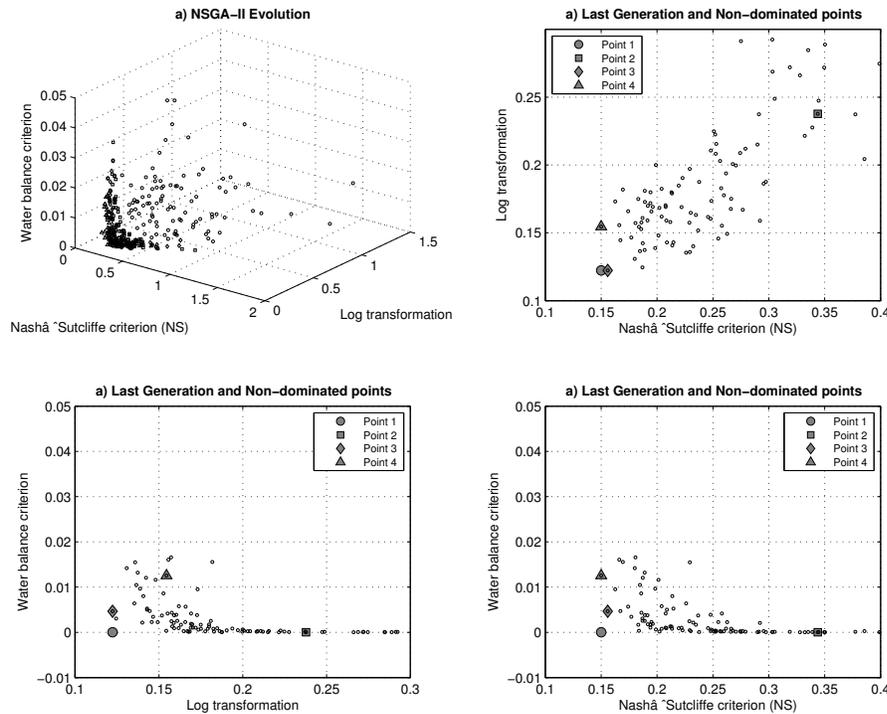


Figura 2.10: Evaluación multi-objetivo de la selección de entradas. Fuente: propia.

Finalmente la solución elegida corresponde al punto 1 de la Figura 2.10, el cual corresponde a una selección que contempla la participación de las distintas variables evaluadas el día precedente a la predicción, excepto el caudal observado que incluye la evaluación de dos días precedentes a la ventana de predicción, tal como se muestra en la Tabla 2.5.

Tabla 2.5: Esquema final de la selección de variables basada en el compromiso de tres funciones objetivo.

Variable	Tipo	Desfases
Caudal	Medición	$Q_{ot-1}, Q_{ot-2}$
Precipitación local	Medición	$P_{1t-1}, P_{2t-1}, P_{4t-1}$
Caudal	Simulación	$Q_{st-1}$
Escorrentía simulada	Simulación	$y_{2t-1}$
Caudal sub-superficial	Simulación	$y_{3t-1}$

Finalmente en la Figura 2.11 se muestra el alto rendimiento del modelo final de predicción, reflejado en un criterio de Nash igual a 0.75 en prueba (0.25 según la formulación de minimización evaluada). Tal valor es prácticamente el mismo encontrado en la primera fase, donde el error medio cuadrático era de 33.58 en prueba, equivalente a un Nash de 0.74. Asimismo se destaca que el criterio de balance es totalmente minimizado y la correlación entre la serie observada y la simulada es igual a 0.88.

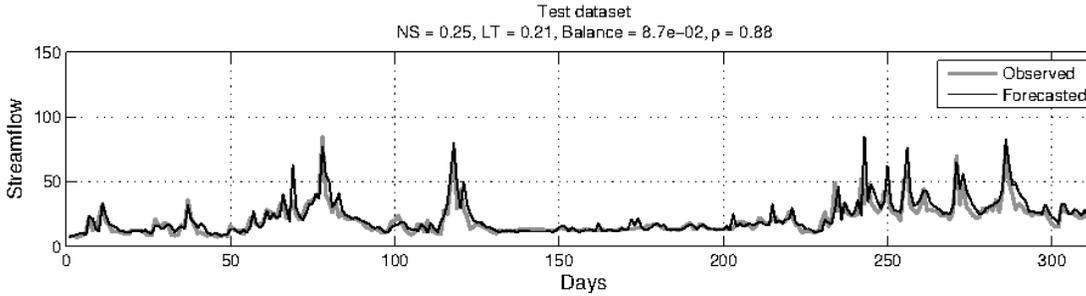


Figura 2.11: Hidrogramas observados y predichos según el esquema de selección final de variables.

## 2.6. Conclusiones

En este estudio se presentan y destacan las ventajas del acoplamiento de un modelo de predicción basado en 50 RNA y un AE multi-objetivo. En este caso el modelo base para la predicción es el valor promedio de la evaluación paralela de 50 RNA del tipo FFNN. Con esto se demostró que aunque se tiene un número elevado de parámetros, el rendimiento de la predicción (generalización) no fue opacado por el sobre-entrenamiento (*overfitting*), lo cual se muestra acorde al principal lineamiento de la predicción multi-respuesta: la combinación, como filtro de múltiples escenarios, conlleva a una mejor capacidad de generalización en contraste con la inestabilidad o el sesgo inducido en un sólo modelo.

Adicionalmente, el principal objetivo de este trabajo se centró en la correcta selección de las variables de entrada para las RNA en un marco multi-objetivo. Los resultados obtenidos confirmaron la gran capacidad de predicción de las RNA con solamente el caudal y la precipitación observada, lo cual es claro para los modelos lluvia-escorrentía. No obstante, se confirmó en la última fase la importancia de incluir la escorrentía superficial y el flujo sub-superficial, variables de respuesta rápida e intermedia simuladas en el modelo hidrológico TETIS. Asimismo, la inclusión reiterativa del caudal simulado en los distintos esquemas de selección deja entrever que el error en la simulación hidrológica sirve como información del sistema de predicción, complementada por las variables de impacto rápido anteriormente citadas que están necesariamente ligadas al orden de magnitud del error. La importancia de estas relaciones radica principalmente en el éxito esperado del trabajo conjunto en referencia tanto al perfeccionamiento del modelo

de simulación hidrológica como del modelo basado en RNA. De otra parte, como se planteó inicialmente, si bien las RNA se destacan por su alto rendimiento en problemas de regresión, es hora de ganar espacios de mayor importancia en la comunidad hidrológica, por lo cual un trabajo combinado de estos modelos matemáticos con modelos hidrológicos convencionales se muestra como estratégico para alcanzar una aceptación generalizada tal como se propone en el “ciclo de vida de una adopción tecnológica” (Figura 2.1).

Con referencia al acoplamiento neuro-evolutivo propuesto, se destaca la gran flexibilidad del AE evaluado. En este orden, aunque en este estudio se presentan dos acoplamientos independientes para determinar el esquema óptimo de entradas para el modelo de predicción, es fácil plantear el mismo acoplamiento para evaluar los hiper-parámetros de la red tales como: el número de neuronas en la capa oculta, la selección de variables de entrada y los métodos de subdivisión de la información para el entrenamiento de la red.

Respecto a los resultados obtenidos en este estudio, se destaca principalmente lo que podría llamarse la equifinalidad en el contexto de la selección de variables, pues la evaluación del NSGA-II mostró que modelos con distintas entradas presentaban el mismo orden de error, de lo cual se desprenden dos lineamientos de trabajo independientes: uno de tipo determinista que se enfocaría en la utilización de una técnica de búsqueda local para optimizar la selección de entradas, y otra de tipo probabilista que buscaría activamente la diversidad al interior del sistema usando distintos subconjuntos de entradas, condición fuertemente ligada a la fiabilidad, consistencia y sesgo en este tipo de predicción. Pero en todo momento teniendo en cuenta la física del fenómeno suministrada por el modelo hidrológico, lo cual da un sentido más realista a los resultados obtenidos.

Estos resultados son importantes para los ríos de montaña en zonas andinas, caso del río Risaralda, en donde se observa una baja regulación hídrica expresada en los cambios abruptos y alta variabilidad en los caudales, haciendo que la predicción sea una tarea compleja con fuertes incertidumbres. Sin embargo, el uso de estas nuevas técnicas híbridas son la base para mostrar resultados satisfactorios y para dar continuidad a este tipo de investigaciones.

## Agradecimientos

Este trabajo fue posible gracias a la información suministrada por CARDER y CORPO-CALDAS, al apoyo de la Universidad Nacional de Colombia y la Universidad LAVAL en lo que se estableció como un programa para fortalecer y consolidar los programas de Doctorado y Maestría, así como al programa *Fortalecimiento de capacidades de procesamiento y análisis de información ambiental*. Adicionalmente, los autores agradecen al Dr. Jasper Vrugt por poner a nuestra disponibilidad el código de NSGA-II versión Matlab. Finalmente, los autores agradecen tanto al Dr. Christian Gagné como al Dr. Francois Anctil por sus comentarios, los cuales mejoraron enormemente la calidad de este manuscrito.



# Referencias Bibliográficas

- [Abrahart et al., 2012] Abrahart, R. J., Anctil, F., Coulibaly, P., Dawson, C. W., Mount, N. J., See, L. M., Shamseldin, A. Y., Solomatine, D. P., Toth, E., and Wilby, R. L. (2012). Two decades of anarchy? emerging themes and outstanding challenges for neural network river forecasting. *Prog. in Phys. Geogr.*, 36(4):480–513.
- [Abrahart et al., 2010] Abrahart, R. J., See, L. M., Dawson, C. W., Shamseldin, A. Y., and Wilby, R. L. (2010). *Nearly two decades of neural network hydrologic modeling*, chapter 6, pages 267–346. World Scientific Publishing.
- [Alpaydin, 2010] Alpaydin, E. (2010). *Introduction to Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, MA, USA, 2 edition.
- [Anctil et al., 2009] Anctil, F., Filion, M., and Tournebize, J. (2009). A neural network experiment on the simulation of daily nitrate-nitrogen and suspended sediment fluxes from a small agricultural catchment. *Ecol. Model.*, 220(6):879–887.
- [Anctil and Lauzon, 2004] Anctil, F. and Lauzon, N. (2004). Generalisation for neural networks through data sampling and training procedures, with applications to streamflow predictions. *Hydrol. Earth Syst. Sci.*, 8(5):940–958.
- [Anctil et al., 2003] Anctil, F., Perrin, C., and Andréassian, V. (2003). Ann output updating of lumped conceptual rainfall/runoff forecasting models. *J. Am. Water Resour. Assoc.*, 39(5):1269–1279.
- [Bowden et al., 2005] Bowden, G. J., Dandy, G. C., and Maier, H. R. (2005). Input determination for neural network models in water resources applications. part 1 – background and methodology. *J. Hydrol.*, 301:75–92.
- [Brochero et al., 2012] Brochero, D., Anctil, F., and Gagné, C. (2012). Forward greedy ann input selection in a stacked framework with adaboost.rt – a streamflow forecasting case study exploiting radar rainfall estimates. In *European Geosciences Union (EGU), Geophysical Research Abstract*, volume 14.
- [Brochero et al., 2013] Brochero, D., Anctil, F., López, K., and Gagné, C. (2013). Finding diversity for building one-day ahead hydrological ensemble prediction system based on artificial neural network stacks. In *European Geosciences Union (EGU), Geophysical Research Abstract*, volume 15.
- [Brown, 2009] Brown, G. (2009). A new perspective for information theoretic feature selection. In Van Dyk, D. and Welling, M., editors, *Twelfth International Conference on Artificial Intelligence and Statistics*, volume 5, pages 49–56, Clearwater Beach, Florida USA.
- [CARDER, 1996] CARDER (1996). Diagnostico, aprovechamiento y conservación del recurso hídrico de la cuenca del río risaralda. departamento de risaralda y caldas colombia. informe final numero 216. Technical report, CARDER.
- [CARDER, 2005] CARDER (2005). Conocimiento de la cantidad del recurso hídrico en el departamento de risaralda, a escala semidetallada. Technical report, Universidad Nacional de Colombia sede Manizales.

- [Chaudhari et al., 2010] Chaudhari, P., Dharaskar, R., and Thakare, V. M. (2010). Computing the most significant solution from pareto front obtained in multi-objective evolutionary. *IJACSA*, 1(4):63–68.
- [Cloke and Pappenberger, 2009] Cloke, H. L. and Pappenberger, F. (2009). Ensemble flood forecasting: A review. *J. Hydrol.*, 375(3-4):613–626.
- [de Vos, 2012] de Vos, N. J. (2012). Reservoir computing as an alternative to traditional artificial neural networks in rainfall-runoff modelling. *Hydrol. Earth Syst. Sci. Discuss.*, 9(5):6101–6134.
- [Deb et al., 2002] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Trans. Evol. Comp.*, 6(2):182–197.
- [Domingos, 2012] Domingos, P. (2012). A few useful things to know about machine learning. *Commun. ACM*, 55(10):78–87.
- [Duan et al., 1993] Duan, Q., Gupta, V., and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *J. Optimiz. Theory App.*, 76(3):501–521.
- [EAIC, 1996] EAIC (1996). Estudio hidrológico y geomorfológico de la parte baja de la cuenca del río risaralda. Technical report, Estudios y Asesorías Ingenieros Consultores Ltda.
- [Eiben and Smith, 2003] Eiben, A. and Smith, J. (2003). *Introduction to evolutionary computing*. Natural computing series. Springer, Berlin, Heidelberg, New York, 1, corr. 2nd printing, edition.
- [Hagan et al., 1996] Hagan, M. T., Demuth, H. B., and Beale, M. (1996). *Neural network design*. Electrical Engineering Series. PWS Publishing Co., Boston, MA, USA, 1 edition.
- [Hettiarachchi et al., 2005] Hettiarachchi, P., Hall, M. J., and Minns, A. W. (2005). The extrapolation of artificial neural networks for the modelling of rainfall-runoff relationships. *J. Hydroinf.*, 7(4):291–296.
- [Hudson et al., 2012] Hudson, M., Hagan, M., and Demuth, H. (2012). *Neural Network Toolbox - User's Guide*. The MathWorks.
- [Kirchner, 2006] Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.*, 42:1–5.
- [Kunstmann et al., 2008] Kunstmann, H., Jung, G., Wagner, S., and Clottey, H. (2008). Integration of atmospheric sciences and hydrology for the development of decision support systems in sustainable water management. *Phys. Chem. Earth*, 33(1-2):165–174.
- [Maier and Dandy, 2000] Maier, H. R. and Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ. Modell. Softw.*, 15(1):101–124.
- [Maier et al., 2010] Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Modell. Softw.*, 25(8):891–909.
- [Martinez Alvarez et al., 2011] Martinez Alvarez, F., Troncoso, A., Riquelme, J., and Aguilar Ruiz, J. (2011). Energy time series forecasting based on pattern sequence similarity. *IEEE Trans. Knowl. Data Eng.*, 23(8):1230–1243.
- [Nguyen and Widrow, 1990] Nguyen, D. and Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, volume 3, pages 21–26. IEEE.

- [Perrin et al., 2003] Perrin, C., Michel, C., and Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.*, 279(1-4):275–289.
- [Pietroniro et al., 2007] Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Versegny, D., Soulis, E. D., Caldwell, R., Evora, N., and Pellerin, P. (2007). Development of the mesh modelling system for hydrological ensemble forecasting of the laurentian great lakes at the regional scale. *Hydrol. Earth Syst. Sci.*, 11(4):1279–1294.
- [Sivapalan et al., 2003] Sivapalan, M., Zhang, L., Vertessy, R., and Blöschl, G. (2003). Downward approach to hydrological prediction. *Hydrol. Processes*, 17(11):2099–2099.
- [Suárez, 2008] Suárez, J. N. (2008). Propuesta metodológica para el estudio del proceso lluvia escorrentía en cuencas urbanas de ciudades de media montaña andina. caso de estudio: cuenca experimental quebrada san luís, manizales caldas. Master’s thesis, Universidad Nacional de Colombia sede Manizales.
- [Vélez, 2001] Vélez, J. I. (2001). *Desarrollo de un modelo hidrológico conceptual y distribuido orientado a la simulación de las crecidas*. PhD thesis, Universidad Politécnica de Valencia.
- [Vélez and Botero, 2011] Vélez, J. J. and Botero, A. (2011). Estimación del tiempo de concentración y tiempo de rezago en la cuenca experimental urbana de la quebrada san luis, manizales. *DYNA*, 78(165):58–71.
- [Vélez et al., 2009] Vélez, J. J., Puricelli, M., López Unzu, F., and Francés, F. (2009). Parameter extrapolation to ungauged basins with a hydrological distributed model in a regional framework. *Hydrol. Earth Syst. Sci.*, 13(2):229–246.
- [Vrugt et al., 2003a] Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S. (2003a). Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res.*, 39(8):1–19.
- [Vrugt et al., 2003b] Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S. (2003b). A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.*, 39(8):1–14.
- [Whitley, 2000] Whitley, D. (2000). *Evolutionary Computation 1: Basic Algorithms and Operators*, chapter 33.3, pages 274–284. Institute of Physics Publishing, Bristol and Philadelphia.
- [Zabel and Mauser, 2013] Zabel, F. and Mauser, W. (2013). 2-way coupling the hydrological land surface model promet with the regional climate model mm5. *Hydrol. Earth Syst. Sci.*, 17(5):1705–1714.



## Capítulo 3

# Análisis de Datos de Calidad del Aire en la Ciudad de Manizales Utilizando Matlab y Openair

*Johana De La Paz Cortés Araujo*<sup>1</sup>

*Carlos Mario González Duque*<sup>1</sup>

*Angie Vanessa Arias Suns*<sup>1</sup>

*Mauricio Orozco Izate*<sup>3</sup>

*Beatriz Helena Aristizábal Zuluaga*<sup>1,2</sup>

<sup>1</sup>Grupo de Trabajo Académico en Ingeniería Hidráulica y Ambiental

<sup>2</sup>Departamento de Ingeniería Química - Facultad de Ingeniería

<sup>3</sup>Departamento de Informática y Computación - Facultad de Administración

Universidad Nacional de Colombia - Sede Manizales

*... “El aire limpio es un requisito para la salud y el bienestar humano”. Organización Mundial de la Salud ...*

### 3.1. Resumen

En este capítulo se presenta el análisis de variables de calidad del aire y la influencia de variables meteorológicas para datos de la ciudad de Manizales, mediante la aplicación de las herramientas computacionales MATLAB y el lenguaje R (librería Openair). Específicamente, se muestra como caso de estudio el análisis de los datos de material particulado menor a 10 micrómetros de diámetro ( $PM_{10}$ ) generados en tres estaciones de monitoreo de la red de calidad del aire de Manizales, durante el periodo comprendido entre enero de 2010 y diciembre de 2012. Las estaciones seleccionadas están asociadas a una estación meteorológica y se encuentran ubicadas en lugares estratégicos de la ciudad, representando la zona del centro histórico (estación Liceo), el centro geográfico (estación Palogrande) y la zona sur oriental (estación Nubia). La aplicación de estas herramientas, permitió comprobar que los datos de material particulado generados durante el periodo de análisis se encuentran dentro de un intervalo estadísticamente satisfactorio según los resultados del análisis de cajas. Los registros medios anuales obtenidos para  $PM_{10}$ , se encuentran por debajo de los límites establecidos en la normativa colombiana; sin embargo, al compararlos con límites internacionales, las estaciones de monitoreo presentaron valores medios anuales por encima de los recomendados por la Organización Mundial de la Salud, OMS. La influencia de emisiones por fuentes móviles, se evidenció como uno de los factores principales en

los niveles de exposición por material particulado en la ciudad. El análisis de variables meteorológicas y su asociación con el  $PM_{10}$ , evidenció el fenómeno de lavado atmosférico (*scavenging*) asociado a la depuración atmosférica de  $PM_{10}$  por acción de la lluvia en Manizales. Finalmente, fue verificada la estabilidad atmosférica de la ciudad (ligeramente inestable) mediante la distribución de los niveles de  $PM_{10}$  respecto a la radiación solar y la velocidad viento. Esta última variable meteorológica, a pesar de presentar valores bajos (promedio) en la ciudad, está sujeta a una dinámica diurna y nocturna diferente debido al cambio de dirección predominante del viento del día a la noche.

**Palabras Clave:** Material particulado ( $PM_{10}$ ), Openair, MATLAB, calidad del aire, variables meteorológicas.

## 3.2. Introducción

La contaminación del aire se define como la alteración causada por la presencia de gases en la atmósfera, o partículas sólidas o líquidas en suspensión, en proporciones distintas a las naturales que pueden poner en peligro la salud del hombre, el bienestar de las plantas y animales, atacar diferentes materiales, reducir la visibilidad o producir olores nocivos. Las emisiones de contaminantes al aire, producto de fuentes antrópicas y naturales, han sido estudiadas debido a su importancia en los cambios climáticos y las afecciones que causan a la calidad de vida de la población, así como el desequilibrio de los ecosistemas [Ramanathan, 2009]. Según el documento CONPES 3550, uno de los contaminantes de mayor peligrosidad y al cual se le ha prestado mayor interés a nivel mundial, es el material particulado. Se estima que la contaminación del aire, debida principalmente al material particulado, afecta la salud de 80 millones de personas en América Latina y el Caribe. En Colombia, alrededor de 6000 muertes anuales y 7400 nuevos casos de bronquitis crónica anuales se atribuyen a la calidad del aire en exteriores; más del 30 % de estos casos se presentan en Bogotá y más del 20 % en ciudades con menos de un millón de habitantes, donde se incluye la ciudad de Manizales [CONPES, 2008].

Para una ciudad intermedia y de alta montaña como Manizales, el estudio de temas relacionados con la contaminación atmosférica y la calidad del aire es importante, debido a factores representativos de la ciudad tales como una alta densidad poblacional (6800 habitantes por  $km^2$ ) y vehicular (254 vehículos por cada 1000 habitantes) (Manizales cómo vamos, <http://manizalescomovamos.org/>), altos contenidos de azufre en el combustible y la cercanía del volcán Nevado del Ruiz, ubicado a 28 km del área urbana: una fuente constante de gases azufrados y emisiones de cenizas en periodos de alta actividad volcánica. Adicionalmente, el estudio de los fenómenos de dispersión de contaminantes atmosféricos, como el material particulado y la relación existente con las variables meteorológicas, toma relevancia para la ciudad, debido a sus características topográficas extremas, su altitud (2150 msnm) y condiciones climáticas, como altos niveles de precipitación (2000 mm/año) y la presencia de microclimas [Cortés, 2010].

Actualmente, la ciudad cuenta con una red de monitoreo de calidad del aire que brinda información sobre material particulado (uno de los contaminantes de mayor importancia en el tema de contaminación atmosférica a nivel urbano). Sumado a esto, la ciudad cuenta con una red de monitoreo de variables meteorológicas, las cuales son un complemento importante a la hora de realizar un análisis completo sobre la dinámica de los diferentes contaminantes del aire. Según la Organización Panamericana de la Salud, OPS, las variables meteorológicas tienen un papel importante en la calidad del aire, ya que el transporte y dispersión de los contaminantes emitidos dependen de las condiciones atmosféricas presentes [Organización Panamericana de la Salud, 2005]. Diversos estudios han puesto de manifiesto la influencia de variables meteorológicas como la precipitación, humedad relativa, temperatura, velocidad del viento y radiación solar, sobre la dinámica y los niveles de exposición al material particulado, a través de fenómenos de remoción por vía húmeda (*scavenging*), procesos de resuspensión y movimientos verticales asociados a las condiciones de estabilidad atmosférica [Seinfeld, 2006]; [Zoras, 2006].

En este sentido, resulta de gran utilidad contar con herramientas computacionales que permitan realizar un análisis y manejo adecuado de la información meteorológica y de calidad del

aire. Esto permite comprender a fondo la problemática de la contaminación atmosférica a nivel local y regional, entender los fenómenos de dispersión y dinámica de los contaminantes atmosféricos, identificar fuentes de contaminación locales y regionales, determinar zonas críticas de contaminación, evaluar el cumplimiento de normativa ambiental, evaluar tendencias y posibles escenarios, así como la generación de conocimiento con posible aplicación en temas más avanzados como, por ejemplo, la aplicación de modelos de dispersión de los contaminantes. La correcta aplicación de herramientas que permitan realizar un análisis confiable de los datos, resulta de gran utilidad y se convierte en un insumo de gran importancia tanto para la autoridad ambiental como para la comunidad científica.

El presente capítulo tiene como objetivo aplicar herramientas computacionales como MATLAB y el lenguaje R (mediante su librería Openair), al análisis de variables de calidad del aire y la influencia de variables meteorológicas para la ciudad de Manizales. Específicamente, se muestra como caso de estudio la aplicación de las principales herramientas que brindan estos programas al análisis de los datos generados en la red de calidad del aire para material particulado menor a 10 micrómetros de diámetro ( $PM_{10}$ ), durante el periodo comprendido entre enero de 2010 y diciembre de 2012. Se analizaron, principalmente, las asociaciones entre las variables meteorológicas y el material particulado tanto para relaciones simples como aquellas en las que interviene más de una variable meteorológica. Sumado a esto, se realiza un análisis general de la distribución temporal del  $PM_{10}$  en las estaciones de monitoreo objeto de estudio.

### 3.3. Generalidades

Material particulado o  $PM$  (por sus siglas en inglés), es el término para las partículas presentes en el aire, tales como polvo, suciedad, hollín, humo y finas gotas de agua; las cuales pueden estar en la atmósfera, ya sea por una fracción pequeña de tiempo o largos periodos (días o semanas). Algunas partículas son grandes o lo suficientemente oscuras para ser vistas en forma de hollín o humo, otras son tan pequeñas que individualmente sólo pueden ser detectadas con un microscopio electrónico [Environmental Protection Agency, 2009]. Una característica importante en el material particulado es su tamaño, del cual depende la vida media en suspensión en la atmósfera y está asociado con el riesgo sobre la salud humana. Las partículas con diámetro menor a 10 micrómetros ( $PM_{10}$ ), son de especial interés ya que tienen la capacidad de ingresar al sistema respiratorio y ocasionar problemas en la salud que van desde asma, tos, dificultad para respirar, hasta infartos no letales y muertes prematuras en personas con problemas cardíacos [Muñoz, 2007]; [González, 2012]. Dentro de la clasificación del  $PM_{10}$ , se diferencia la fracción fina, caracterizada por partículas con diámetro menor a 2.5 micrómetros ( $PM_{2.5}$ ) y la fracción gruesa, en la cual se encuentran las partículas con diámetro superior a 2.5 micrómetros. Debido a su pequeño tamaño, las partículas que hacen parte de la fracción fina revisten una alta peligrosidad, ya que pueden alojarse directamente en los pulmones, generando un mayor riesgo para la salud humana [Environmental Protection Agency, 2009].

Entre las principales fuentes de  $PM_{10}$ , se encuentran los procesos de combustión interna, las operaciones de molienda, triturado y tamizado y el polvo de los caminos pavimentados y no pavimentados. Otros tipos de partículas se pueden formar en el aire a partir de la transformación química de los gases. Estas son formas indirectas cuando los gases procedentes de la quema de combustibles reaccionan con la luz solar y el vapor del agua [Environmental Protection Agency, 2009].

#### 3.3.1. El $PM_{10}$ y sus efectos en la salud

Según la OMS, la población del área urbana, tanto de países desarrollados como en vía de desarrollo, se encuentra expuesta a problemas de contaminación atmosférica debido al material particulado. Esto se convierte a la vez, en un problema de salud pública al quedar demostrado

los efectos adversos del  $PM$  sobre la salud humana. Los principales problemas respiratorios y cardiovasculares que puede generar este tipo de contaminación, van desde la irritación de las vías respiratorias, tos, dificultad para respirar, disminución de la función pulmonar, asma agravada, desarrollo de bronquitis crónica, arritmia cardíaca, infartos no letales y muerte prematura en personas con problemas del corazón y pulmones [WHO, 2006]; [Muñoz, 2007]. El material particulado se ha constituido en un problema crítico, dado el impacto directo que genera en las condiciones de salud de la población, en particular a lo relacionado con las infecciones respiratorias agudas (IRA). Por otro lado, también tiene un impacto negativo en cuanto a la visibilidad, olores e higiene [Cárdenas, 2010].

Las emisiones de partículas de diesel causan importantes episodios de contaminación en ambientes urbanos, contribuyendo a su vez a la formación de otros tipos de contaminantes, por ejemplo, los hidrocarburos aromáticos policíclicos. Muchos de los efectos tóxicos y cancerígenos del aire contaminado en las ciudades, se han vinculado a los hidrocarburos aromáticos policíclicos adsorbidos en el material particulado [Quijano, 2010].

Los efectos nocivos sobre la salud mencionados anteriormente, han promovido diferentes estrategias y directrices encaminadas a la adopción de límites de concentración de material particulado estrictos, que aseguren la calidad del aire y la mínima exposición de la población a este tipo de contaminación. Así mismo, la implementación de monitoreos de la calidad del aire para tener un seguimiento y vigilancia a las emisiones de material particulado [WHO, 2006]. Es por esto que en Colombia se han establecido diferentes normativas, como es el caso de la resolución 610 de 2010, en la cual se fijan los límites máximos permisibles en el aire ambiente para los contaminantes criterio a condiciones de referencia (25 °C y 760 mmHg). En la Tabla 3.1, se muestra la evolución de los límites máximos permisibles para niveles de inmisión de  $PM_{10}$  de acuerdo a la normativa colombiana [Ministerio de Ambiente, 2006b]; [Ministerio de Ambiente, 2010a].

En Colombia, existen Decretos como el 948 de 1995, el cual contiene el Reglamento de Protección y Control de la Calidad de Aire [Ministerio del Medio Ambiente, 1995], y el Decreto 244 de 2006, por el cual se crea y reglamenta la comisión Técnica Nacional Intersectorial para la Prevención y el Control de la Contaminación del Aire – CONAIRE [Ministerio de Ambiente, 2006a]. Resoluciones como la 909 de 2008 del MAVDT, por la cual se establecen normas y estándares de emisión admisibles de contaminantes a la atmósfera por fuentes fijas, la Resolución 910 de 2008 [Ministerio de Ambiente, 2008], por la cual se reglamentan los niveles permisibles de emisión de contaminantes que deberán cumplir las fuentes móviles terrestres, y la Resolución 650 de 2010 [Ministerio de Ambiente, 2010b], por la cual se adopta el Protocolo para el Monitoreo y Seguimiento de la Calidad del Aire, el cual ofrece a las autoridades ambientales y a las personas o entidades interesadas, una guía y lineamientos básicos para llevar a cabo el monitoreo y seguimiento a la calidad del aire, cumpliendo con los más altos estándares de calidad.

Tabla 3.1: Límites máximos permisibles para niveles de inmisión de  $PM_{10}$  en Colombia.

Año	Normativa vigente	Límite máximo permisible	Tiempo de exposición
2006 a 2008	Resolución 601 de 2006 [Ministerio de Ambiente, 2006b]	70	Anual
		150	24 horas
2009 y 2010	Resolución 601 de 2006 y 610 de 2010 [Ministerio de Ambiente, 2006b]; [Ministerio de Ambiente, 2010a]	60	Anual
		150	24 horas
A partir del 2011	Resolución 610 de 2010 [Ministerio de Ambiente, 2010a]	50	Anual
		100	24 horas

Para la ciudad de Manizales, la Corporación Autónoma Regional de Caldas, CORPOCALDAS, con el apoyo de la Universidad Nacional de Colombia Sede Manizales, tiene una red de monitoreo de calidad del aire en la cual se miden Partículas Suspendidas Totales (*PST*),  $PM_{10}$  y  $PM_{2.5}$ . Para el caso del  $PM_{10}$ , las estaciones que maneja la Universidad han proporcionado información desde el año 2009, complementando la red de calidad del aire que maneja la Corporación Ambiental, la cual posee datos de  $PM_{10}$  desde el año 2006.

### 3.3.2. Meteorología y su influencia en el estudio de la calidad del aire

Las variables meteorológicas tienen un papel importante en la calidad del aire, ya que el transporte y dispersión de los contaminantes emitidos dependen de las condiciones atmosféricas presentes. Por lo tanto, la legislación y planificación del país deben tener en cuenta su comportamiento para la toma de decisiones a futuro [Organización Panamericana de la Salud, 2005]. Según la Organización Panamericana de la Salud, "la meteorología se usa para predecir el impacto ambiental de una nueva fuente de contaminación del aire y para determinar el efecto de las modificaciones de las fuentes existentes en la calidad del aire"; es por esto que se hace necesario analizar y correlacionar las principales variables meteorológicas con concentraciones de material particulado [Organización Panamericana de la Salud, 2005]. Las variables meteorológicas de mayor interés son:

- **Dirección y velocidad del viento:** La velocidad del viento actúa como un factor de transporte y control de la dispersión de contaminantes atmosféricos como el *PM*. La dirección del viento establece la orientación en la cual se transportan y difunden los contaminantes en el aire. Estas variables son de importancia en el análisis de posibles fuentes de contaminación atmosférica en una determinada zona, y también pueden ser útiles en la determinación de las condiciones de estabilidad atmosférica y el grado en el cual se podrían estar dispersando contaminantes en el aire. Las estaciones meteorológicas de una red de monitoreo, reportan, generalmente, la velocidad del viento en unidades de medida de metros por segundo (m/s), y la dirección del viento en grados.
- **Temperatura y presión barométrica:** Estas variables son importantes ya que se usan para corregir las concentraciones de *PM* de condiciones locales a estándar, y así poder comparar con la normativa vigente del país. Además, los movimientos del aire se pueden asociar a cambios de presión y temperatura. La temperatura puede también jugar un papel importante en la generación de condiciones para la formación de contaminantes secundarios por reacción de contaminantes primarios en la atmósfera. La temperatura se reporta generalmente en grados Celsius (°C) y la presión barométrica en milímetros de mercurio (mmHg) o hectopascales (hPa).
- **Humedad relativa y precipitación:** Se analizan estas variables debido a que pueden promover la disminución de la concentración de contaminantes en el aire, por acción de procesos de lavado atmosférico. La humedad relativa se encuentra medida en porcentaje (%) y la precipitación en milímetros (mm).
- **Radiación solar:** Esta variable y la velocidad del viento, determinan la estabilidad de la atmósfera y, por lo tanto, la dispersión de contaminantes como el *PM*. En una atmósfera estable, los contaminantes presentarán baja dispersión y, como consecuencia, la concentración de éstos aumentará. La unidad de medida de la radiación solar se reporta generalmente en unidad de potencia - Watt - por unidad de área - metros cuadrados ( $W/m^2$ ).

El municipio de Manizales, posee 13 estaciones meteorológicas y 3 estaciones hidrometeorológicas. En las primeras se miden variables como humedad relativa, dirección y velocidad del viento, presión barométrica, temperatura ambiente y precipitación. Las dos últimas variables mencionadas también son medidas en estaciones hidrometeorológicas, además del nivel y el

caudal [CORPOCALDAS, 2008]. Estas estaciones son complemento de la red de monitoreo de calidad del aire, ya que la información meteorológica es de importancia para determinar el comportamiento de la atmósfera y la concentración del  $PM_{10}$ . Además, se pueden realizar correlaciones entre estas variables, para determinar cuáles están directamente relacionadas con los diferentes fenómenos de contaminación.

### 3.3.3. Herramientas para análisis de datos de calidad del aire

Debido a la creciente preocupación por la contaminación atmosférica y cómo manejar los datos históricos reportados por las diferentes estaciones de monitoreo de calidad del aire, surgen herramientas estadísticas y matemáticas que facilitan su correcta interpretación, con el fin de poder comparar con la legislación vigente del país, estándares de calidad internacionales y para fines investigativos. Las herramientas de análisis de datos se presentan como una alternativa para comprender a fondo la problemática de la contaminación en la atmósfera, los posibles factores de control y la aplicación en modelos de dispersión de los contaminantes. Algunos ejemplos de software para el análisis de datos son R, MATLAB, Excel, Disper entre otros, y sus diferentes librerías.

Todos estos software permiten la visualización de datos, y logran con esta técnica la generación de una poderosa estrategia para aprovechar la orientación visual de los seres humanos, ya que es extraordinariamente buena en el reconocimiento de patrones, tendencias y anomalías. Los gráficos de uso común para la visualización de exploración clásica, corresponden a diagramas de caja, histogramas, gráficos de series de tiempo o diagramas de dispersión de dos o más variables [Gibert, 2008]; [Berthold, 2010]. Además, la visualización también permite identificar agrupamientos de datos, confirmar las tendencias esperadas o la contradicción de las mismas, y toda aquella posible información extraíble a través de las gráficas.

En este estudio, se utilizaron los software R (uso libre) y MATLAB. Su selección se realizó teniendo en cuenta las librerías y herramientas que utilizan (Openair y MATLAB toolbox, respectivamente), las cuales son de gran utilidad en el análisis de datos de calidad del aire; específicamente, en el análisis de concentración de  $PM_{10}$  y la correlación de estos datos con las variables meteorológicas mencionadas anteriormente.

### 3.3.4. R y Openair.

R es un software de programación, de uso libre para el análisis estadístico y gráfico de datos. Este software tiene implementada una librería llamada Openair (también es de uso libre), la cual permite el análisis de datos de calidad del aire. El proyecto Openair es desarrollado por el Grupo de Investigación del Medio Ambiente del King's College London y soportado por la Universidad de Leeds [ERG, 2013].

La librería Openair ha sido utilizada ampliamente en el Reino Unido, ya que este país cuenta con una base de datos sólida en el monitoreo de la calidad del aire y ha sido el propulsor de este paquete. En la Tabla 3.2 se exponen algunas de las aplicaciones que ha tenido este software en el manejo de datos de calidad del aire para dicho país.

Tabla 3.2: Algunos estudios de calidad del aire con manejo de datos mediante la aplicación de Openair.

ARTÍCULO	RESUMEN
Openair - An R package for air quality data analysis [Carslaw, 2011]	Los autores realizan una introducción al paquete Openair, muestran algunas de sus funciones y comandos. Caso de estudio calidad del aire para UK, mediante gráficos polares bivariados, estimaciones de las tendencias cuantitativas y el uso de funciones para fines de evaluación de modelos. Finalmente, hace referencia a cómo los datos de contaminación del aire se pueden analizar de forma rápida y eficaz, disminuyendo tiempo y de forma interactiva.
The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks [Mead, 2013]	Autores como [Mead, 2013], hacen un especial hincapié en el uso de sensores electroquímicos para monitorear la calidad del aire a bajos costos. Se concluye que los sensores electromagnéticos son una alternativa ideal para cuantificar el factor de exposición humana a contaminantes atmosféricos en ambientes urbanos, dado su completo desempeño al medir la alta granulometría estructural de la calidad del aire medida en partes por billón.
Quantifying temporal trends in ground level ozone concentration in the UK [Munir, 2013]	Este estudio analiza las tendencias del ozono en varios percentiles, cuantiles, regresiones, función TheilSen y análisis Changeoint. Según [Munir, 2013], las tendencias de ozono muestran una variabilidad significativa en los diferentes indicadores estadísticos (por ejemplo, cuantiles, media, mediana, máximo y seleccionado). Las tendencias máximas fueron negativas, mientras que las tendencias de la mediana y la media fueron positivas durante el período de estudio (1993-2011), tanto en las zonas rurales como en sitios urbanos. Las tendencias del ozono fueron negativas durante los últimos 8 años (2004-2011), lo que podría haber sido causado por la estabilización de la concentración de $NO_x$ durante este período. El análisis estadístico, se realizó mediante tres herramientas estadísticas: el Quantreg, Openair y Changeoint.
Global comparison of VOC and CO observations in urban areas [von Schneidemesser, 2010]	En este estudio se presenta la espacialización de los compuestos orgánicos volátiles (VOC) y el monóxido de carbono (CO) en la vía Marylebone, en el centro de Londres desde 1998 hasta 2008. Las tendencias que obtuvieron los autores, muestran disminuciones estadísticamente significativas para todos los compuestos orgánicos volátiles considerados, que van desde -3 % a -26 % por año. El monóxido de carbono se redujo en un 12 % por año durante el período de medición. Durante el mismo período, de 1998 hasta 2008, la relación que obtuvieron de VOC a CO en Londres fue constante, lo que indicó que las medidas que tomó el gobierno para reducir la emisiones afectaron de igual manera a los VOC y al CO. Los resultados que obtuvieron de las correlaciones, indicaron que las emisiones más significativas proceden de fuentes relacionadas con el tráfico vehicular, la industria y las fuentes relacionadas con el uso de GLP.

A continuación se citan tres casos de interés, donde se ha usado Openair para el tratamiento de los datos de calidad del aire:

- Los diagramas polares bivariados:** Proporcionan un método gráfico para mostrar en conjunto tres variables, por lo general, dos variables meteorológicas y la concentración de un contaminante. En el estudio publicado por [Carslaw, 2012], se analizaron datos provenientes de dos lugares, cada uno con un contaminante diferente. El primero, fue un sitio cercano al aeropuerto de Heathrow, en el oeste de Londres. Los datos obtenidos fueron las concentraciones horarias de  $NO_x$  desde el año 2000 hasta el 2010. Este lugar es de particular interés debido a las diferentes fuentes que pueden estar emitiendo este contaminante

(aeropuerto, varias vías principales y una gran parte de la zona urbana de Londres). El segundo lugar, fue la principal vía pública del centro de Londres (Marylebone). Se analizaron datos de concentraciones horarias de  $SO_2$  de los años 1998 a 2005. El diagrama polar bivariado obtenido para el sitio cercano al aeropuerto de Londres, correlacionó la velocidad y dirección del viento con las concentraciones de  $NO_x$ ; la mayor concentración del contaminante se observó cuando la velocidad del viento fue la más baja, comportamiento asociado a la cercanía de la zona urbana y la estabilidad atmosférica de la zona. En el caso del sitio de Marylebone, el diagrama polar bivariado obtenido correlacionó la dirección del viento y la temperatura ambiente con las concentraciones de  $SO_2$ ; se mostró que existe una relación de dependencia entre la concentración del contaminante y la temperatura. Cuando los vientos provienen del este, la concentración de  $SO_2$  aumenta con el aumento de la temperatura ambiente; esto se debe a que las plumas emitidas de las chimeneas (aproximadamente a 30 km) se liberan debajo de una capa de inversión y el aire que se encuentra debajo de la mencionada capa se vuelve inestable. Cuando la inestabilidad alcanza el nivel de la pluma que se encuentra bajo la capa de inversión, los contaminantes se transportan rápidamente hacia abajo hasta llegar al suelo [Carslaw, 2012].

- **La función Theil-Sen:** Es un método robusto de regresión lineal, que se basa en escoger la mediana de la pendiente del conjunto de datos. Es usado cuando existen valores extremos que afectan la tendencia de los datos, por esto es más precisa que la regresión lineal simple. Con el fin de evaluar las medidas tomadas por el gobierno en Londres para disminuir las emisiones de los vehículos en la zona más transcurrida de esta ciudad denominada LEZ, [Ellison, 2013] analizaron las concentraciones de  $PM_{10}$  y  $NO_x$  desde el año 2001 hasta el 2011 de esta zona. Se encontró que la zona LEZ tuvo un aumento de la tasa de sustitución de los vehículos más antiguos, en comparación con la tasa obtenida en años anteriores. Este cambio parece haber mejorado ligeramente la calidad del aire de LEZ, con una reducción entre el 2.46 y 3.07 % de las concentraciones de  $PM_{10}$  y un porcentaje de reducción entre el 0.5 a 1.5 % de reducción de emisiones de  $NO_x$ . Mediante el uso de la regresión lineal Theil-Sen, se obtuvieron las tendencias del comportamiento de los contaminantes mencionados, se evaluó que para el 2015 se esperan porcentajes de reducción de emisiones del 16.86 % y 37.58 % para  $PM_{10}$  y  $NO_x$ , respectivamente [Ellison, 2013].
- **Las gráficas de cajas o cajas de Tukey:** Son gráficas de gran utilidad para la representación de series de datos, incluyendo en un solo gráfico información completa sobre la simetría, la dispersión y el sesgo de los datos objeto de análisis. Como se observa en el esquema de la Figura 3.1, este tipo de gráficos presenta un resumen de tres cuartiles representados entre la caja (extremo superior, extremo inferior y la mediana), así como los valores máximos y mínimos (bigote superior e inferior de la gráfica, conocidos como whisker). También da la posibilidad de incluir valores atípicos o fuera del rango de análisis [Follos, 2012]. Este tipo de gráficos son de utilidad para realizar una comparación entre estaciones de monitoreo, en la cual están representados los estadísticos principales básicos que permiten determinar el cumplimiento de la normativa, datos atípicos, límites superior e inferior, entre otros.

Para el año 2009, [Harrison, 2012] realizaron un estudio de las concentraciones horarias de  $PM_{2.5}$  para diferentes sitios del Reino Unido. La zona urbana contó con datos de 37 estaciones de monitoreo, la rural con datos de una estación, la zona industrial con tres estaciones y la zona de avenidas de alto tráfico con datos de siete estaciones.

Mediante el uso de diagramas de cajas, los autores obtuvieron que las concentraciones para el año 2009 de las 37 estaciones urbanas, se encontraban, en su mayoría, en un intervalo de 12 a 14  $\mu\text{g}/\text{m}^3$ . Las mayores concentraciones de  $PM_{2.5}$  que encontraron fueron 17.6, 15.5 y 14.8  $\mu\text{g}/\text{m}^3$ , las cuales correspondieron, respectivamente, a Eltham en el este de Londres, Lisburn al norte de Irlanda y Stoke on Trent, en la región central del Reino Unido. Para Lisburn, la concentración de  $PM_{2.5}$  fue asociada al uso doméstico de combustibles sólidos de esta región; para la región central, no hallaron una influencia directa que explicara la concentración de  $PM_{2.5}$ . Para el caso de la zona rural, se tuvo un intervalo de concentración de 3 a 6  $\mu\text{g}/\text{m}^3$ , valores que, según los autores, pudieron ser no representativos debido a que se tomaron en cuenta los datos de monitoreo de una sola

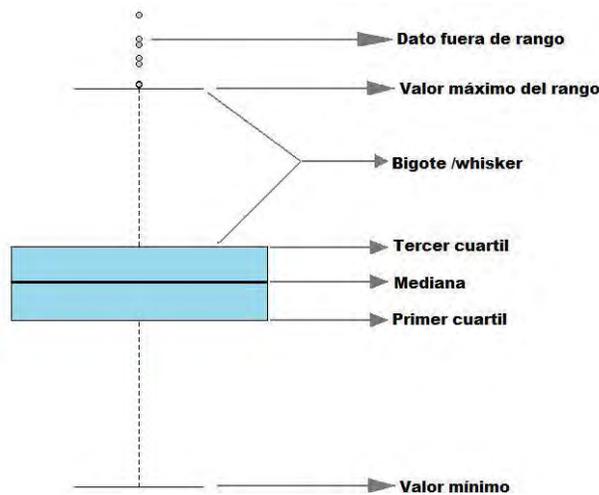


Figura 3.1: Esquema de gráfico de cajas elaborado en Openair (Tomado de [Follos, 2012]).

estación. En el caso de avenidas principales y secundarias, mediante el uso de diagramas de caja, se determinó que las concentraciones de  $PM_{2.5}$  de las carreteras principales fue 0.9 veces mayor que las concentraciones halladas en la vías secundarias; comportamiento asociado a que las avenidas principales presentaron mayor tráfico de vehículos en el centro de la ciudad, en comparación con las carreteras secundarias que poseen menor tráfico, calzadas más amplias y entorno más abierto, llevando a menores concentraciones de este contaminante. Las concentraciones medias anuales de  $PM_{2.5}$  para el sector industrial, presentaron valores de  $8.6 \mu\text{g}/\text{m}^3$  (Grangemouth, cerca de una refinera de petróleo),  $8.1 \mu\text{g}/\text{m}^3$  (Port Talbot, cerca de una planta siderúrgica) y  $14 \mu\text{g}/\text{m}^3$  (SalfordEccles, zona industrial de Manchester). Los autores encontraron concentraciones medias anuales más bajas de la zona industrial, por debajo de los intervalos encontrados para la zona urbana, debido a que estas zonas industriales se encontraban ubicadas en el sector rural; mientras que la zona industrial de Manchester (concentración más alta) estaba dentro del área urbana del Reino Unido, lo que no permite la dispersión del contaminante y, por lo tanto, se observó una mayor concentración de  $PM_{2.5}$  [Harrison, 2012].

- Además de las herramientas mencionadas, Openair permite realizar histogramas de frecuencia, estimar la densidad Kernel (densidad de probabilidades de una variable de la que se desconoce su distribución), rosa de vientos y contaminantes, distribuciones horarias de contaminantes escogidas por el usuario (anuales, mensuales, horarias), gráficos de datos con líneas suavizadas y marcadores, diagramas calendarios de contaminantes, análisis estadístico básico (cálculo de mínimos, máximos, medias, medianas, percentiles), entre otros [Carslaw, 2011]; [Follos, 2012].

### 3.3.5. MATLAB

MATLAB es el nombre abreviado de “MATrixLABoratory”. Es un lenguaje de alto nivel y un entorno interactivo para el cálculo numérico, visualización y programación. Usando MATLAB, pueden analizarse datos, desarrollar algoritmos, crear modelos y aplicaciones. El lenguaje, las herramientas y las funciones integradas de matemáticas, permiten explorar múltiples enfoques y llegar a una solución más rápida que con las hojas de cálculo o lenguajes de programación tradicionales, como C / C++ o Java™ [Mathworks, 2013].

MATLAB proporciona herramientas para adquirir, analizar y visualizar datos, debido a su amplio abanico de programas de apoyo especializados, denominados *Toolboxes*, que extienden

significativamente el número de funciones incorporadas en el programa principal. Estos *Toolboxes* cubren en la actualidad prácticamente casi todas las áreas principales en el mundo de la ingeniería y la simulación, destacando entre ellos los *Toolboxes* de proceso de imágenes, señales, control robusto, estadística, análisis financiero, matemáticas simbólicas, redes neuronales, lógica difusa, identificación de sistemas, simulación de sistemas dinámicos, etc. [Universidad de Santiago de Compostela, 2013]. Para algunas de estas aplicaciones, es utilizado el *PRTools* (Pattern Recognition Tools), el cual permite el reconocimiento de patrones de datos experimentales [van der Heijden, 2004]. El programa cuenta con comandos de análisis estadístico básico y de análisis multivariable (MVA) de los datos.

A continuación se exponen algunas de las aplicaciones que ha tenido este software en el manejo de datos de calidad del aire:

- **Análisis de Componentes Principales aplicado a Compuestos Orgánicos Volátiles:** El PCA (Principal Component Analysis, por sus siglas en inglés) es un método de análisis multivariado usado para reducir la dimensión de un conjunto de datos formado por un gran conjunto de variables que se encuentran relacionadas, permitiendo una interpretación más sencilla de la varianza de los datos. El PCA está basado en la descomposición de los datos originales organizados en una matriz bilineal. [Pérez-Rial, 2010], investigaron la distribución temporal y las posibles fuentes de los compuestos orgánicos volátiles (*VOCs*) en el área suburbana de la ciudad de Coruña (noroeste de España), usando dos métodos de análisis multivariados: el PCA, su variación con la matriz aumentada (MA-PCA) y el análisis factorial paralelo (PARAFAC). Para aplicar los métodos anteriores, realizaron un total de 50 campañas de monitoreo (durante los años 2005 y 2006), que cubrían las diferentes estaciones del año, a cada hora de los días de la semana.

Cuando los autores implementaron el PCA para los datos de un día, identificaron diferentes tipos de fuente para los *VOCs* como aromáticos, alifáticos, halogenados y biogénicos, aunque no fue posible la identificación de tendencias de las emisiones de acuerdo con la hora del día. Pero cuando aplicaron los métodos MA-PCA y PARAFAC al conjunto de datos obtenidos durante una semana, detectaron que las fuentes de *VOCs* (tráfico de carretera, los procesos industriales, las emisiones biogénicas, etc.) son más importantes durante el día que durante la noche, debido a las diferentes reacciones fotoquímicas que pueden ocurrir con otros compuestos presentes en el aire. También identificaron que una de las principales fuentes de *BVOC* (Compuestos orgánicos volátiles biogénicos) durante el día, son las plantas debido a que sintetizan este compuesto estimuladas por la radiación solar y las altas temperaturas. No detectaron patrones cíclicos de emisión para los diferentes días de la semana. Cuando aplicaron la técnica MA-PCA a los datos registrados durante todo el año, identificaron una importante fuente difusa de contaminación relacionada con la liberación de *VOCs* en la atmósfera a partir de fuentes urbanas e industriales durante el día. Además confirmaron el contraste de emisiones entre diferentes fuentes de *VOCs*, las cuales siguen tendencias cíclicas por horas, siendo los compuestos aromáticos y biogénicos predominantes durante el día, mientras que los halogenados durante la noche. No detectaron patrones de emisión claros sobre la época del año para el período en estudio. En los monitoreos que realizaron los autores identificaron 43 *VOCs*, de los cuales los más abundantes eran los compuestos aromáticos; a excepción de algunas muestras tomadas en verano y primavera, en las cuales los compuestos más abundantes eran los *BVOCs*. Entre los aromáticos, uno de los compuestos con la mayor concentración promedio fue el tolueno ( $2.16 \mu\text{g}/\text{m}^3$ ) y el que tuvo una mayor incidencia en las muestras tomadas. Respecto a los compuestos alifáticos, el que mayor concentración reportó fue el n-heptano ( $47.29 \mu\text{g}/\text{m}^3$ ); los compuestos halogenados con concentraciones más altas fueron el tetraclorometano y tetraclorometano con  $17.74$  y  $26.9 \mu\text{g}/\text{m}^3$ , respectivamente. La distribución diaria de los *VOCs* mostró que los compuestos aromáticos y alifáticos incrementaron su concentración durante el día, con valores máximos en la mañana y algunos en la tarde; mientras que los *VOCs* halogenados no mostraron un comportamiento regular respecto al incremento de su concentración.

Respecto a la distribución estacional de los *VOCs*, tanto los compuestos aromáticos como alifáticos mostraron que en las estaciones calurosas (verano y primavera) las concentraciones aumentaban ( $7.74$  y  $2.09 \mu\text{g}/\text{m}^3$ , compuestos aromáticos y alifáticos, respectivamente) respecto a las estaciones más frías invierno y otoño ( $7.34 \mu\text{g}/\text{m}^3$  aromáticos y  $1.90 \mu\text{g}/\text{m}^3$  alifáticos). Para los compuestos halogenados ocurrió lo contrario, las mayores concentraciones que obtuvieron, se encontraban en estaciones frías ( $4.88 \mu\text{g}/\text{m}^3$ ) en comparación con la estación cálida ( $3.16 \mu\text{g}/\text{m}^3$ ). Los (*BVOCs*) mostraron que la concentración en temporada de verano ( $3.87 \mu\text{g}/\text{m}^3$ ) era mayor a la encontrada en temporada fría ( $1.10 \mu\text{g}/\text{m}^3$ ) [Pérez-Rial, 2010].

- **Análisis de Componentes Principales aplicado al ozono troposférico:** [Ozbay, 2011] estudiaron la aplicación de métodos estadísticos multivariados para la predicción de la concentración de ozono ( $O_3$ ) troposférico en función de variables meteorológicas (temperatura, precipitación, humedad, presión, dirección del viento, velocidad del viento y radiación solar) y otros contaminantes ( $PM_{10}$ ,  $SO_2$ ,  $NO$ ,  $NO_2$ ,  $CO$ ,  $CH_4$ ,  $NMHC$ ). En dicho estudio, estas variables fueron medidas cada hora durante un año (septiembre 2008 a agosto 2009) en la región de Dilovasi (noroeste de Kocaeli), debido a que allí se encuentra la zona industrial y existen autopistas cercanas, ferrocarriles y puertos marítimos, que pueden estar contribuyendo a la formación de ozono troposférico. En este estudio se evaluaron tres casos. El modelo 1 evaluó la relación entre las concentraciones predichas y las medidas en campo durante el año de monitoreo. El modelo 2 evaluó la relación entre la concentración predicha para las estaciones cálidas y las medidas en campo; y el modelo 3 evaluó la relación entre las concentraciones predichas para las estaciones frías y las medidas en campo. Los resultados obtenidos mediante el uso de modelos de regresiones lineales multivariadas (MLR) por este autor, muestran coeficientes de correlación ( $R^2$ ) de 0.90, 0.92 y 0.85, respectivamente para los modelos 1, 2, 3. El resultado obtenido muestra el comportamiento global del ozono troposférico, la concentración de este contaminante aumenta en temporadas de verano. Con el fin de reducir el número de variables de entrada, aplicaron PCA para los datos horarios medidos en un año. Es así, como redujeron quince variables iniciales a cuatro grupos de PC. La unión de los PCs explican la formación de  $O_3$  con  $R^2$  de 0.63 y el menor error estándar obtenido en la estimación [Ozbay, 2011].
- La aplicación de las herramientas de MATLAB para el análisis de datos, también incluyen la implementación de aplicaciones más especializadas y detalladas para el tratamiento e interpretación de la información como las Redes Neuronales, las cuales aún son poco empleadas.

### 3.4. Metodología

Con el fin de aplicar las herramientas MATLAB y Openair para el análisis de datos de calidad del aire y la influencia de variables meteorológicas, se seleccionaron tres estaciones de monitoreo que hacen parte de la red de calidad del aire de Manizales. Cada estación se encuentra ubicada en lugares estratégicos de la ciudad y representan la zona centro (estación Liceo), el centro geográfico (estación Palogrande) y la zona sur oriental (estación Nubia). La Tabla 3.3, resume las características principales de las estaciones de monitoreo utilizadas en el análisis. Cada punto de monitoreo está asociado a una estación meteorológica; en el caso de Palogrande y Nubia, se utilizó la información proveniente de las estaciones Posgrados y Enea respectivamente, las cuales hacen parte de la red de estaciones meteorológicas IDEA-OMPAD. En el caso de la estación Liceo, se utilizó la información de la estación meteorológica ubicada en el colegio Liceo Isabel La Católica y que es operada por el personal de CORPOCALDAS.

Tabla 3.3: Características principales de las estaciones de monitoreo analizadas.

Estación	Ubicación (Coordenadas geográficas)	Altura (msnm)	Influencia de fuentes de material particulado
Liceo (L) (Sector centro histórico de Manizales – zona comercial)	N 5° 3' 58.9" W 75° 28' 30"	2156	Influencia de actividad vehicular alta con emisiones de transporte público y particular principalmente, los cuales transitan por vías aledañas como la avenida Santander. La influencia de actividad industrial es baja.
Palogrande (P) (Sector centro geográfico de Manizales – zona residencial)	N 5° 3' 32.5" W 75° 29' 43.5"	2166	Influencia de tráfico vehicular media-alta, con tránsito de vehículos de transporte público y particular principalmente por vías aledañas como la avenida Paralela. Influencia de actividad industrial nula.
Nubia (N) (Sector suroriente de Manizales – cercana a zonas residencial e industrial)	N 5° 1' 57.3" W 75° 30' 49"	2104	Influencia de tráfico vehicular baja. La principal fuente de contaminación sobre esta zona proviene de emisiones de fuentes industriales ubicadas al oriente de la estación.

En la Tabla 3.4 se resumen las especificaciones de cada estación meteorológica utilizada en el análisis. Para este estudio, fueron utilizados los valores de material particulado menor a diez micrómetros de diámetro ( $PM_{10}$ ) y los registros de variables meteorológicas. El análisis comprendió el periodo de enero de 2010 a diciembre de 2012. Los valores de  $PM_{10}$  empleados en los análisis, corresponden a las concentraciones medias durante periodos de 24 horas de monitoreo, calculadas a partir de la medición con equipos de alto volumen (HiVol). Estos equipos están diseñados para mantener un flujo volumétrico de 1130 L/min mediante un dispositivo controlador de flujo (tubo Venturi), recolectando el material particulado en filtros de micro cuarzo o fibra de vidrio. La masa de  $PM_{10}$  en cada estación, se obtiene mediante una técnica gravimétrica y la concentración se expresa a condiciones estándar (25 °C y 1 atm.) en unidades de microgramo de material particulado por metro cúbico de aire ( $\mu\text{g}/\text{m}^3$ ).

Tabla 3.4: Características principales de las estaciones de monitoreo analizadas.

Estación meteorológica	Estación de calidad del aire asociada	VARIABLES meteorológicas	Observaciones sobre la estación meteorológica
Liceo: Ubicada en el colegio Liceo Isabel La Católica – sector Fundadores	Liceo	- Temperatura (°C) - Dirección del viento (grados) - Velocidad del viento (m/s) - Presión barométrica (hPa) - Humedad relativa (%) - Precipitación (mm)	- Estación operada por el personal de CORPOCALDAS. - Se tienen registros promedio diarios de cada una de las variables meteorológicas.
Posgrados: Ubicada en el edificio de posgrados – Campus Palogrande UNAL Manizales.	Palogrande	- Temperatura (°C) - Dirección del viento (grados) - Velocidad del viento (m/s) - Presión barométrica (hPa) - Humedad relativa (%) - Precipitación (mm)	- Estaciones operadas por el personal de la Universidad Nacional de Colombia Sede Manizales. - Se tienen registros para cada 5 minutos de cada una de las variables meteorológicas.
Enea: Ubicada en el Campus La Nubia UNAL Manizales.	Nubia	- Radiación solar (W/m <sup>2</sup> ) - Evapotranspiración (mm)	

### 3.4.1. Conformación de las bases de datos

Previamente a la aplicación de las herramientas de análisis, fue necesario conformar las bases de datos para cada una de las estaciones de monitoreo. En el caso de las estaciones Palogrande y Nubia, se realizó un cálculo de valores diarios de las variables meteorológicas, teniendo en cuenta que la información de  $PM_{10}$  se tiene sobre una base de tiempo diaria y no horaria. La obtención de registros diarios tuvo en cuenta:

- Promedios aritméticos durante periodos de 24 horas para las variables temperatura, presión, velocidad del viento, humedad relativa.
- Sumatoria de registros durante periodos de 24 horas para la variable precipitación.
- Para el caso de la obtención de la dirección del viento predominante para periodos de 24 horas, se empleó la metodología de cálculo descrita en la nota técnica del IDEAM elaborada por [Benavides, 2010], basada en el cálculo de los promedios zonales y meridionales del viento a partir de los registros horarios de cada estación.

A la información en escala diaria de variables meteorológicas, se adicionó la información de  $PM_{10}$  para los días en los cuales se tenían registros de acuerdo al monitoreo realizado en la red de calidad del aire de Manizales, tanto por CORPOCALDAS (entidad encargada de operar la estación Liceo), como por la Universidad Nacional de Colombia Sede Manizales (encargada de operar las estaciones Palogrande y Nubia). Finalmente, los registros de variables meteorológicas y  $PM_{10}$  se guardaron en archivos de texto plano, para el manejo con las herramientas de análisis de Openair, y en ficheros de datos para aplicar las herramientas de MATLAB.

### 3.4.2. Metodología para análisis con Openair

Las bases de datos creadas se guardaron como archivos de texto plano en formato csv, el cual es compatible con Openair y permite su importación a la herramienta. Una vez ingresada la base de datos en Openair, se utilizaron diferentes comandos y sentencias en lenguaje R, con el fin de generar los gráficos y estadísticos básicos que permitieran el análisis de información de calidad del aire. Los comandos o funciones de Openair utilizados en el proceso de análisis son descritos en la Tabla 3.5.

Tabla 3.5: Comandos de Openair utilizados en el procedimiento analítico.

<b>Actividad</b>	<b>Comandos</b>	<b>Función</b>
Elaboración de rosas de viento	<i>windRose</i>	Elaboración de rosas de vientos a partir de información meteorológica de velocidad y dirección del viento.
Gráfico de cajas	<i>Boxplot</i>	Comparación gráfica entre estaciones de monitoreo con base en estadísticos principales básicos.
Resumen de estadísticos principales por año	<i>aqStats</i>	Genera estadísticos principales anuales (mediana, media, máximo, mínimo, percentiles, entre otros) para cada grupo de datos.
Gráfico de distribución diaria	<i>timePlot</i>	Graficar valores diarios para una serie de tiempo específica.
Gráfico de variación media mensual y semanal	<i>timeVariation</i>	Graficar promedios horarios, la evolución semanal (para cada día específico) y mensual de una serie de datos.
Elaboración de rosas de contaminantes	<i>pollutionRose</i>	Análisis de información sobre contaminantes atmosféricos a través de una rosa de contaminantes, la cual puede realizarse en función de otra variable, ya sea meteorológica o de calidad del aire.

El análisis de la distribución de los vientos (de acuerdo a su velocidad y dirección) es de gran utilidad en estudios de calidad del aire, ya que permite tener una idea de las posibles fuentes de contaminantes atmosféricos, así como dilucidar de manera general el grado de dispersión de los contaminantes; se planteó, inicialmente, un análisis de la distribución de los vientos en las estaciones de monitoreo, tanto para intervalos de las variables horarias como para los promedios diarios obtenidos de velocidad y dirección del viento. El objetivo fundamental con este análisis, fue determinar la dinámica de los vientos en cada una de las estaciones analizadas mediante el uso de la herramienta Openair.

Posteriormente, se aplicaron técnicas para cálculo de estadísticos y gráficas en Openair para análisis comparativo de los resultados de  $PM_{10}$  en las estaciones de monitoreo, así como análisis de variación temporal durante el periodo 2010 – 2012. La distribución temporal y la comparación entre estaciones, permiten determinar periodos y zonas críticas de contaminación, además de hacer posible la comparación con límites máximos permisibles definidos en la normativa nacional y protocolos internacionales.

Finalmente, se aplicaron comandos para generar gráficos que permitieran el análisis de la influencia de variables meteorológicas sobre el material particulado. Se realizó el análisis de las variables precipitación, humedad relativa y temperatura, definidas previamente como las variables meteorológicas de mayor influencia en los niveles de concentración del  $PM_{10}$  en la ciudad de Manizales [González, 2012]. Cabe resaltar que se realizó una selección de las posibles herra-

mientas y comandos de Openair que permitieran realizar un análisis temporal y comparativo entre los resultados de las estaciones de monitoreo de calidad del aire, así como un análisis que permitiera dilucidar asociaciones entre variables meteorológicas y de calidad del aire.

### 3.4.3. Metodología para análisis con MATLAB

Los datos fueron guardados como un fichero en archivos de extensión .xls, que le permiten al software cargar e importar la información meteorológica y de calidad del aire. Posteriormente, fueron empleados diferentes comandos del *Toolbox* de MATLAB para la elaboración de gráficos y análisis estadísticos. La Tabla 3.6, muestra los comandos utilizados.

Tabla 3.6: Comandos de MATLAB utilizados en el procedimiento analítico.

Actividad	Comandos	Función
Importar datos	<i>xlsread</i>	Carga e importa los datos y el texto de una hoja de cálculo en un libro de Excel.
Visualización de datos	<i>scatter, scatterd</i>	Grafica la distribución de puntos respecto a dos o tres variables.
Regresión Lineal	<i>cftool</i>	Permite abrir la herramienta para el ajuste de curvas.
Análisis de Componentes Principales (PCA)	<i>pca, printcomp, pcacov (MATLAB Pattern Recognition Toolbox- PRTools)</i>	Ejecutan las funciones estadísticas que conciernen al Análisis de Componentes Principales.

Dado que las variables analizadas tienen comportamientos dinámicos diferentes, es importante visualizar la dispersión de los mismos empleando gráficas que los relacionen entre sí. Estas gráficas permiten observar el grado de relación o influencia de las condiciones meteorológicas en el  $PM_{10}$  y así detectar posibles agrupaciones (*clustering*), tendencias y datos atípicos (*outliers*).

Inicialmente, fueron realizadas gráficas que relacionaron los datos de  $PM_{10}$  con los datos de temperatura, velocidad del viento, presión barométrica, precipitación, humedad relativa y radiación solar. Posteriormente, fueron graficados los datos de  $PM_{10}$  con dos variables meteorológicas simultáneamente. Estas últimas tienen sentido, particularmente, en el conocimiento de la dinámica de la estabilidad atmosférica en la ciudad. La velocidad del viento y la radiación solar, son las variables que proporcionan información de este fenómeno.

Seguido a esto, fueron realizadas regresiones lineales entre los valores del  $PM_{10}$  y los de las variables meteorológicas en donde se identificaron los parámetros estadísticos que definen el nivel de ajuste de los datos. Este análisis se realiza con el objetivo de verificar si este sistema ambiental cuenta con un comportamiento que puede ser representado por una línea recta.

Finalmente, fue realizado el Análisis de Componentes Principales (PCA) a los datos meteorológicos, sobre los cuales se etiquetaron los niveles de  $PM_{10}$  en rangos de concentración bajo, medio y alto. Este procedimiento analítico permitió la reducción del número de variables y la proyección de los componentes principales que retienen la mayor parte de la varianza de los mismos. Así, se pudieron conocer las variables meteorológicas que tienen mayor peso en las estaciones de calidad del aire en la ciudad de Manizales.

De este modo, MATLAB ofrece herramientas que, si bien no son especializadas para la implementación explícita de datos de calidad del aire, permiten el análisis numérico y estadístico de cualquier conjunto de datos. Además, posibilita la visualización de datos en múltiples formas gráficas.

En esta sección, son mostradas solo algunas de estas aplicaciones que pueden ser utilizadas para la generación de conocimiento a partir de datos, en este caso, empleando la información de calidad del aire y meteorológica de la ciudad de Manizales. La implementación de estas herramientas, corresponde a un punto de inicio en la aplicación de técnicas de análisis de datos para estudios atmosféricos en Manizales, que puede ser extendida y profundizada en estudios posteriores.

## 3.5. Resultados y Análisis

### 3.5.1. Estudio de variaciones temporales de $PM_{10}$ y la influencia de variables meteorológicas utilizando Openair

Mediante el manejo del lenguaje estadístico R y su librería Openair, se realizó el análisis de los datos de material particulado ( $PM_{10}$ ) reportados en tres estaciones de monitoreo de calidad del aire ubicadas en la ciudad de Manizales durante el periodo 2010 – 2012. Este manejo de la información es de relevante utilidad, ya que la ciudad cuenta con una amplia base de datos de registros de variables meteorológicas y su red de calidad del aire contiene registros con información periódica de  $PM_{10}$ , la cual es un insumo de gran importancia puesto que su análisis permite comprender fenómenos de dispersión y dinámica de los contaminantes atmosféricos, identificar fuentes de contaminación locales y regionales, determinar zonas críticas de contaminación, evaluar el cumplimiento de normativa ambiental y evaluar tendencias y posibles escenarios. Por lo tanto, la correcta aplicación de herramientas que permitan realizar un análisis confiable de los datos resulta de gran utilidad y se convierte en una fuente de información de interés tanto para la autoridad ambiental como para la comunidad científica.

Los resultados y análisis descritos a continuación, comprenden el uso de gráficas para el análisis de: rosas de vientos, variación temporal del  $PM_{10}$ , comparación de estadísticos principales para cada estación de monitoreo y el estudio de la influencia de variables meteorológicas sobre los niveles de concentración de  $PM_{10}$ . Lo anterior, se elaboró mediante el uso de las herramientas estadísticas y gráficas que ofrece la librería Openair, buscando, principalmente, generar un mayor entendimiento de la dinámica del  $PM_{10}$  en Manizales y su relación con la meteorología, en una ciudad con una alta variabilidad de precipitación y la presencia de microclimas, a pesar de su reducida área.

#### 3.5.1.1. Distribución de vientos para periodos diurno, nocturno y predominancia para registros de 24h (día completo). Uso de la herramienta windRose para la elaboración de la rosa de vientos.

Una de las herramientas principales de Openair es la elaboración de rosas de vientos a partir de información meteorológica, horaria o diaria, mediante el uso del comando “windRose”. A través de sentencias simples es posible realizar en Openair una diferenciación por periodos de interés, por ejemplo, rosas de vientos en periodos diurno y nocturno para cada uno de los años que componen la base de datos. Teniendo en cuenta esta utilidad, se realizó un análisis de vientos para las estaciones de monitoreo empleadas en el análisis. Debido a que Palogrande y Nubia poseen datos con información de variables meteorológicas cada cinco minutos, se realizó el análisis de periodos diurno y nocturno en cada uno de los años (2010 a 2012) para dichas estaciones. La Figura 3.2a-b, muestra las rosas de vientos anuales en Palogrande (Figura 3.2a) y Nubia (Figura 3.2b) según los periodos descritos. En ambas estaciones, se observa una diferenciación entre los patrones de vientos del periodo diurno, con direcciones predominantes provenientes del norte (N), nor-noroeste (NNW) y noroeste (NW); en comparación con el periodo nocturno, para el cual prevalecen vientos provenientes del este (E), este-sureste (ESE) y sureste (SE). Estudios previos han reportado este patrón bidireccional del viento en el departamento de Caldas

caracterizado por una dirección predominante noroeste en el día y sureste en la noche, asociado a efectos orográficos y de flujos de masas de aire con movimientos de tipo montaña-valle y valle-montaña [Cortés, 2010].

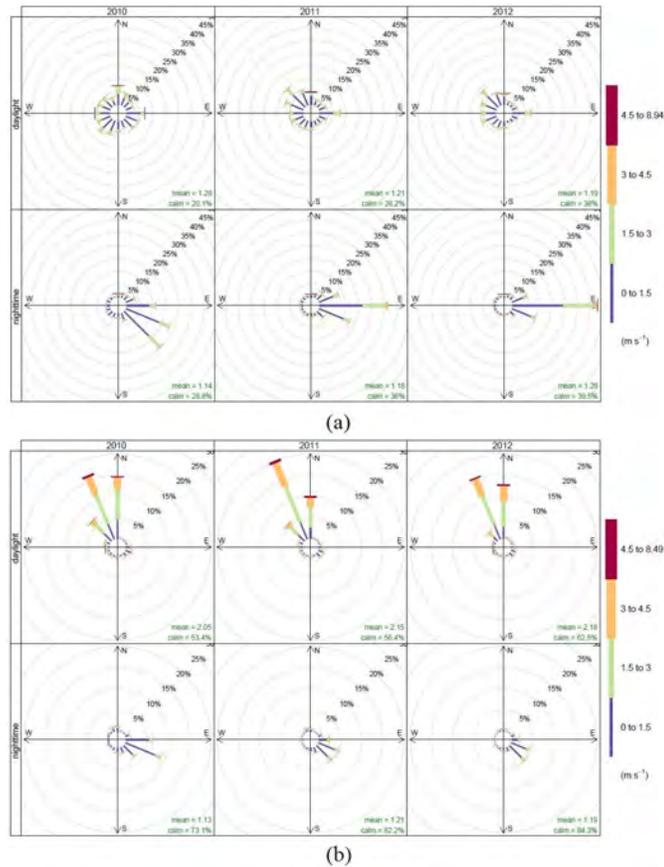


Figura 3.2: Distribución de vientos para periodo diurno (daylight) y nocturno (nighttime) en las estaciones Palogrande y Nubia durante 2010 - 2012. (a) Palogrande y (b) Nubia.

Debido a que la información de calidad del aire existente en la ciudad de Manizales se presenta con base en periodos de 24 horas (diaria), es conveniente analizar las rosas de vientos del día completo, haciéndose necesario el cálculo de las direcciones predominantes para periodos de 24 horas a partir de la información horaria disponible [Benavides, 2010]. La Figura 3.3a-c, muestra los resultados de distribución de vientos para periodos de 24 horas.

Es posible observar cómo hay un cambio en las direcciones predominantes del viento, debido a que éstos son promedios zonales y meridionales basados en registros horarios. Además, se observa un comportamiento diferente al comparar las estaciones de monitoreo analizadas. Por ejemplo, la estación Palogrande registra un patrón predominante promedio diario de vientos proveniente del este, a diferencia de las estaciones Nubia y Liceo, cuyos patrones medios diarios provienen del norte principalmente. Estos cambios zonales pueden estar asociados a efectos orográficos de carácter local y a las características geográficas propias de cada estación, teniendo en cuenta que la ciudad de Manizales presenta una topografía compleja y variable a pesar de su reducida área.

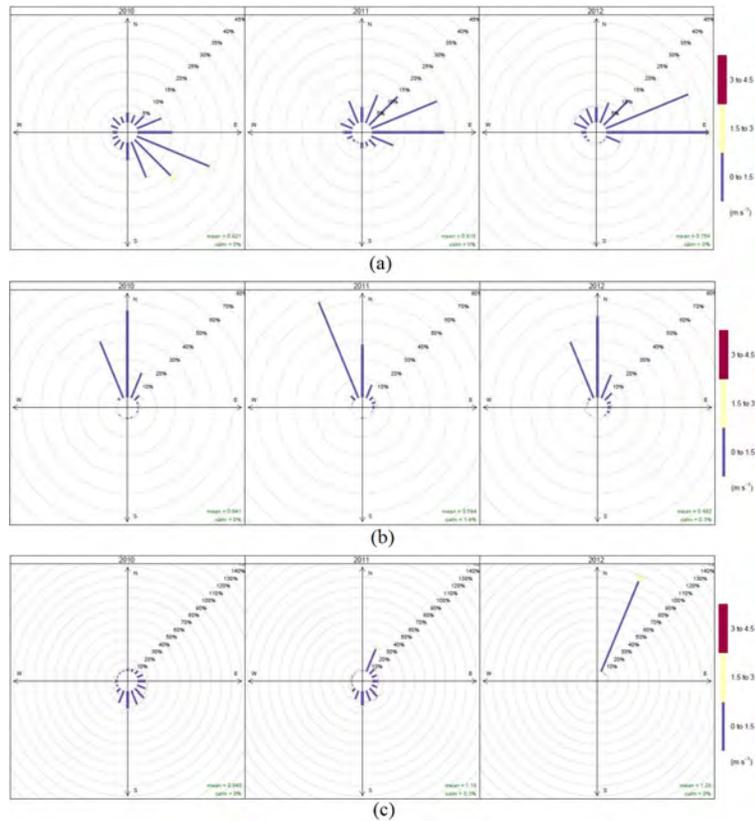


Figura 3.3: Distribución de vientos para periodo diario (24 h) en las estaciones de monitoreo durante 2010 - 2012. (a) Palogrande. (b) Nubia. (c) Liceo.

La forma como se represente la información de vientos en las bases de datos, es muy importante a la hora de la implementación de técnicas de análisis de Openair, las cuales pueden estar basadas en la distribución de concentraciones de contaminantes según la velocidad y dirección de los vientos (por ejemplo la rosas de contaminantes). La información basada en registros horarios, permite una mayor rigurosidad en el análisis, permitiendo diferenciar entre periodos diurnos y nocturnos, así como variaciones horarias en el comportamiento de los contaminantes atmosféricos. Sin embargo, para aquellas redes de calidad del aire en las cuales se tenga información de contaminantes para periodos de 24 horas (como es el caso de Manizales), la representación de distribución de vientos puede brindar un panorama general sobre el comportamiento de cada contaminante, resultando de gran utilidad en los análisis e interpretación de posibles fuentes y dinámica de los mismos.

### 3.5.1.2. Resumen de estadísticos principales en las estaciones de monitoreo para el periodo 2010 - 2012 utilizando el gráfico de cajas y la función aqStats

**Análisis comparativo entre estaciones de monitoreo utilizando el gráfico de cajas:** Para el caso específico de la ciudad de Manizales, se elaboró un gráfico de cajas durante el periodo 2010 - 2012 (Figura 3.4). Se observa claramente un incremento en las concentraciones de  $PM_{10}$  en la estación Liceo con respecto a las estaciones Palogrande y Nubia. La influencia de fuentes de emisión locales como lo son las fuentes móviles, puede influir notablemente en las diferencias presentadas. Estudios previos han sugerido el aporte importante de las fuentes móviles a los niveles de contaminación por  $PM_{10}$  en la ciudad de Manizales [Velasco, 2006]; [González, 2012]. Al analizar los resultados obtenidos en la Figura 3.4, es posible afirmar que los datos de calidad del aire recolectados en las estaciones de monitoreo, presentan uniformidad y se

encuentran dentro de un rango estadísticamente satisfactorio. La mayoría de los datos obtenidos se encuentran entre la distribución de las cajas y sus cuartiles superior e inferior, además son mínimos los valores atípicos en cada una de las estaciones de monitoreo.

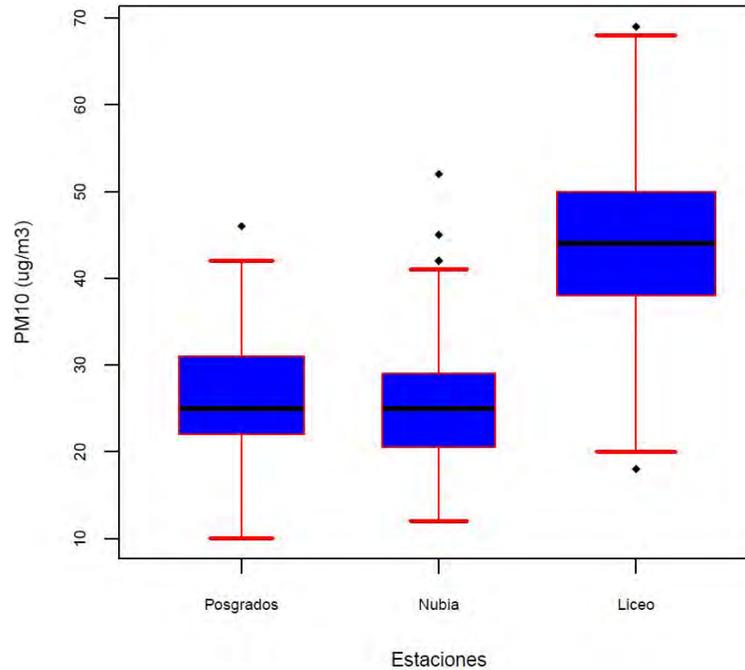


Figura 3.4: Distribución de concentraciones de  $PM_{10}$  para tres estaciones de monitoreo en la ciudad de Manizales durante 2010 – 2012.

**Resumen de estadísticos principales mediante la función aqStats:** Para el caso de la ciudad de Manizales, la Tabla 3.7 muestra los resultados obtenidos al aplicar la función aqStats a los datos de  $PM_{10}$  durante el periodo 2010 – 2012 en las estaciones de monitoreo analizadas. Se puede observar cómo todas las estaciones de monitoreo están por debajo del límite anual establecido en la Resolución 610 de 2010 de  $50 \mu\text{g}/\text{m}^3$  [Ministerio de Ambiente, 2010a] para los valores medios de cada estación. Sin embargo, todas las estaciones están por encima del límite promedio anual recomendado por la OMS de  $20 \mu\text{g}/\text{m}^3$  [WHO, 2006]. Otro aspecto interesante, son los ligeros incrementos en las concentraciones medias anuales que se presentaron en las tres estaciones de monitoreo del año 2010 al año 2012. A pesar de ser incrementos pequeños, si esta tendencia continúa para los años siguientes, se puede hablar de un indicativo de la presión que ejercen las diferentes fuentes de contaminación en la ciudad (móviles y fijas) y un punto de análisis para estrategias futuras con miras a impedir un comportamiento creciente en este aspecto, teniendo en cuenta no solo el cumplimiento de la normativa nacional (que durante los últimos años ha incrementado su rigurosidad), sino también la protección de la población de la ciudad de Manizales. Estos incrementos se evidencian también en los resultados obtenidos para los percentiles 95 y 99, que indican los valores de concentración de  $PM_{10}$  que están por encima de 95 % y 99 % de los datos, respectivamente, en cada estación de monitoreo.

Tabla 3.7: Resumen de estadísticos principales para  $PM_{10}$  en cada estación de monitoreo aplicando la función aqStats de Openair.

Estadísticos estación Liceo ( $\mu\text{g}/\text{m}^3$ )						
Año	Promedio	Mínimo	Máximo	Mediana	Percentil 95	Percentil 99
2010	43	24	62	43	56	61
2011	43	20	68	44	57	67
2012	45	18	69	46	61	68
Estadísticos estación Palogrande ( $\mu\text{g}/\text{m}^3$ )						
Año	Promedio	Mínimo	Máximo	Mediana	Percentil 95	Percentil 99
2010	24	10	37	23	36	37
2011	26	10	41	26	37	40
2012	29	12	46	28	42	44
Estadísticos estación Nubia ( $\mu\text{g}/\text{m}^3$ )						
Año	Promedio	Mínimo	Máximo	Mediana	Percentil 95	Percentil 99
2010	24	12	45	21	40	43
2011	26	15	42	26	37	41
2012	30	20	52	29	46	50

### 3.5.1.3. Variación diaria, media mensual y semanal de $PM_{10}$ en las estaciones de monitoreo durante 2010-2012

Al analizar la distribución diaria del  $PM_{10}$  en las estaciones de monitoreo (Figura 3.5), se observa una clara diferencia entre la distribución temporal de los datos de la estación Liceo en comparación con las estaciones Palogrande y Nubia. Este comportamiento sugiere que las emisiones provenientes de las fuentes móviles en el centro histórico de la ciudad son uno de los factores principales en la definición de los niveles de  $PM_{10}$  alrededor de esta zona de la ciudad. Los resultados obtenidos para la distribución diaria de  $PM_{10}$  son concordantes con el resumen estadístico presentado en la Tabla 3.7, donde se observa claramente que la estación Liceo presenta valores que superan en algunos casos los  $60 \mu\text{g}/\text{m}^3$ , con una línea de tendencia alrededor de los  $45 \mu\text{g}/\text{m}^3$ ; registros muy apartados en comparación con las demás estaciones analizadas, cuyos valores oscilan entre los 20 y  $30 \mu\text{g}/\text{m}^3$ , principalmente.

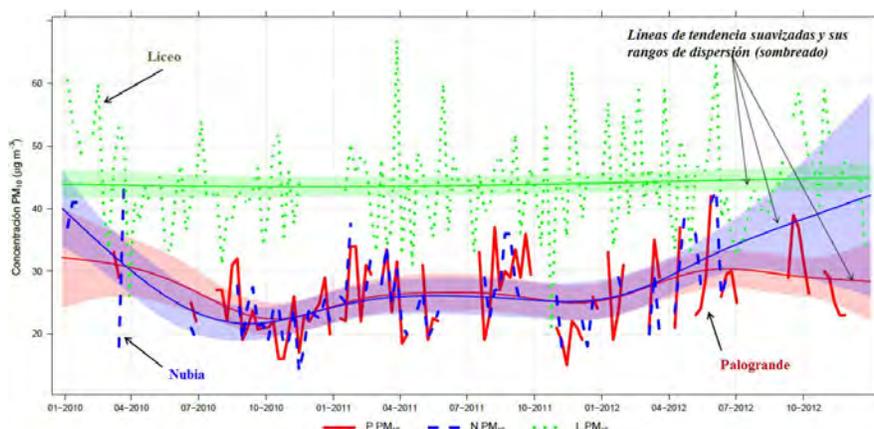


Figura 3.5: Distribución diaria del  $PM_{10}$  en la ciudad de Manizales para las estaciones Palogrande (P), Nubia (N) y Liceo (L) durante 2010 – 2012. Las sombras representan los rangos de dispersión asociados a las líneas de tendencia calculada para cada estación.

La Figura 3.6, muestra la distribución semanal y promedio mensual obtenido en la ciudad de Manizales para las estaciones Liceo, Palogrande y Nubia. En términos de distribución media mensual durante el periodo 2010 - 2012, las tres estaciones presentan los mayores valores en las concentraciones promedio de  $PM_{10}$  durante los primeros dos meses del año (enero y febrero); posteriormente, la variación en  $PM_{10}$  tiende a disminuir y, a partir del mes de mayo, predomina un comportamiento estable. Para la estación Liceo, este comportamiento es constante hasta el mes de diciembre, en comparación con las estaciones Palogrande y Nubia, en las cuales las concentraciones medias mensuales tienden a disminuir nuevamente a partir del mes de septiembre, con posterior incremento durante el mes de diciembre.

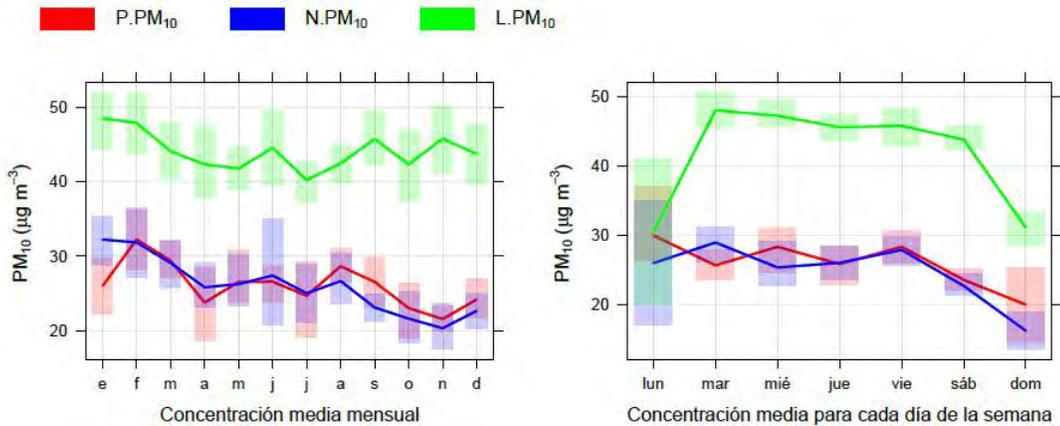


Figura 3.6: Variación media mensual y concentración media por días de la semana del  $PM_{10}$  en la ciudad de Manizales para las estaciones Palogrande (P), Nubia (N) y Liceo (L) durante 2010 - 2012. Las sombras representan los rangos de valores para el cálculo de cada promedio (mensual o por días de la semana).

La distribución mensual obtenida en las estaciones Palogrande y Nubia, puede tener relación con el comportamiento bimodal de la precipitación en la ciudad de Manizales, caracterizado por dos periodos de lluvias intensas (abril - mayo y septiembre - noviembre), teniendo en cuenta el efecto de la precipitación en los procesos de remoción de contaminantes de la atmósfera (*scavenging*). Para el caso de Liceo, este comportamiento no es tan marcado, sin embargo, se debe tener en cuenta que la estación Liceo tiene una influencia directa y constante de emisiones provenientes de actividad vehicular, lo cual puede contribuir a una distribución más uniforme de las concentraciones de  $PM_{10}$ , caso contrario a lo que sucede en Palogrande y Nubia, estaciones en las cuales la influencia de emisiones, producto de actividad vehicular o industrial, no es tan directa.

Los comandos “timePlot” y “timeVariation” de Openair, fueron utilizados para obtener la distribución diaria, así como la evolución semanal (para cada día específico) y promedios mensuales de la serie de datos de  $PM_{10}$  en Manizales.

#### 3.5.1.4. La rosa de contaminantes y su utilidad en el análisis de la influencia de variables meteorológicas individuales

Además de la realización de rosas de vientos con información meteorológica, Openair permite realizar un análisis de información sobre contaminantes atmosféricos a través de una rosa de contaminantes, la cual puede realizarse en función de otra variable, ya sea meteorológica o de calidad del aire. Para el caso de la ciudad de Manizales, es posible realizar rosas de contaminantes para  $PM_{10}$  con información diaria y determinar la influencia de variables meteorológicas como precipitación, humedad relativa, temperatura, entre otras, en los niveles de contaminación por  $PM_{10}$ .

Para cada una de las estaciones de monitoreo analizadas, se realizaron rosas de contaminantes de  $PM_{10}$  y se incluyó en el análisis la variación del  $PM_{10}$  en términos de distintos intervalos de las principales variables meteorológicas (precipitación, humedad relativa y temperatura) que, según reportes previos, influyen en el aumento o disminución de las concentraciones de  $PM_{10}$  en el aire ambiente de Manizales [González, 2012]. El desarrollo de este tipo de gráficos permite distinguir cuáles son los niveles de contaminación por  $PM_{10}$  para diferentes intervalos de variables como precipitación, humedad relativa, temperatura, además de incluir la información de la distribución de los vientos y su relación con las concentraciones del contaminante objeto de análisis. Este análisis de vientos y contaminantes, permite además diferenciar posibles fuentes o focos de contaminación, resultando de gran utilidad cuando se tienen registros horarios, tanto de variables meteorológicas como de contaminantes atmosféricos.

Para el caso de Manizales, donde se tienen registros promedio para cada día de monitoreo, el análisis de la dirección del viento entregará resultados casi idénticos para cualquier rosa de contaminantes generada, sin embargo, permite realizar un análisis preliminar de posibles fuentes de contaminación locales.

En términos de precipitación, la Figura 3.7a-c muestra las variaciones en las concentraciones de  $PM_{10}$  en función de diferentes intervalos de precipitación, además de incluir las concentraciones de  $PM_{10}$  según los registros de la velocidad y dirección del viento. En las estaciones Liceo (Figura 3.7a), Palogrande (Figura 3.7b) y Nubia (Figura 3.7c), se observa cómo a medida que incrementa la precipitación las concentraciones de  $PM_{10}$  son menores, indicando el efecto de esta variable meteorológica en la reducción de la contaminación por  $PM_{10}$  en Manizales. Este efecto está asociado a la capacidad que tiene la lluvia para realizar la depuración de contaminantes en el ambiente, a través de fenómenos de lavado atmosférico (*scavenging*), en el cual se dan procesos de remoción por nucleación (al interior de las nubes) y colisión (por descenso de gotas de lluvia y posterior arrastre de los contaminantes presentes) [Seinfeld, 2006]; [González, 2012].

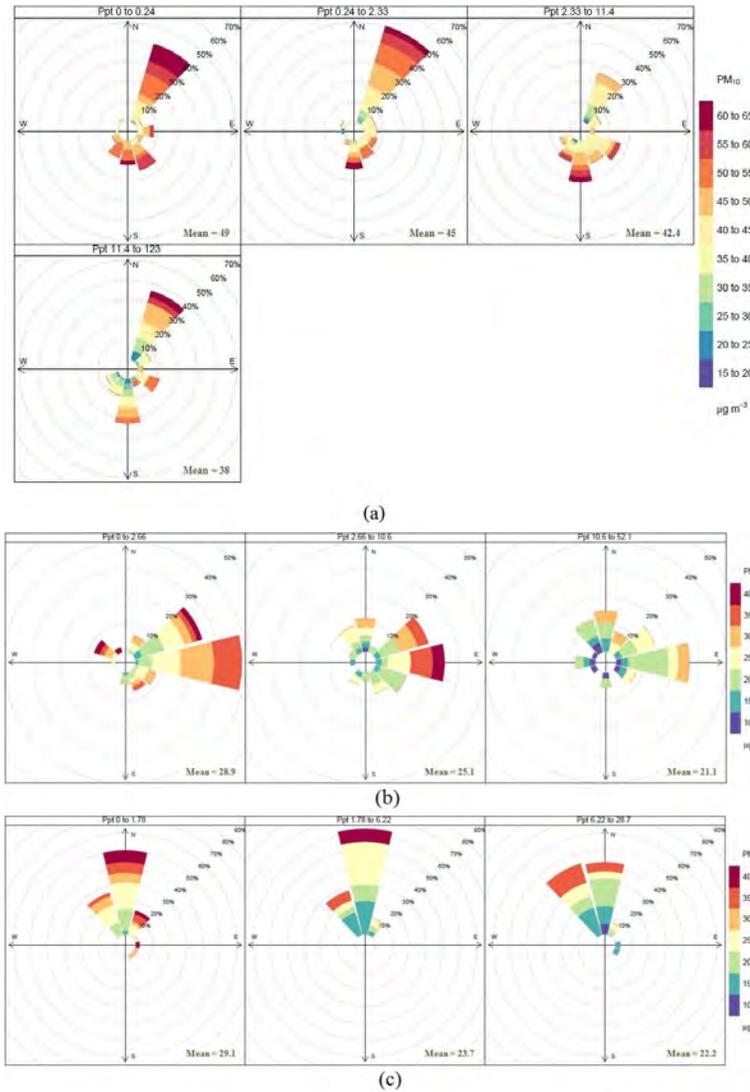


Figura 3.7: Porcentaje de contribución proporcional a la media del contaminante  $PM_{10}$  en función de la velocidad y dirección del viento para diferentes intervalos de precipitación en a) Liceo, b) Palogrande y c) Nubia.

Comportamiento similar se presenta al analizar la influencia de la humedad relativa. Como se observa en la Figura 3.8a-c, el aumento en los valores de la humedad relativa, está asociado a menores concentraciones de  $PM_{10}$  en las tres estaciones de monitoreo.

Existe un descenso en los valores medios de  $PM_{10}$  para cada intervalo de humedad relativa a medida que sus límites son mayores. Este comportamiento es más acentuado a humedades relativas superiores al 90 %, donde se tienen concentraciones de  $PM_{10}$  por debajo de  $55 \mu g/m^3$ , a diferencia de los demás intervalos en los cuales se tuvieron concentraciones en un rango de 60 a  $69 \mu g/m^3$ . El descenso en las concentraciones de  $PM_{10}$  al aumentar la humedad relativa a más del 90 % tiene sentido, ya que humedades relativas cercanas al nivel de saturación (100 %) indican una mayor probabilidad de precipitación y, por tanto, una mayor probabilidad de fenómenos de remoción atmosférica a través de la lluvia [Elminir, 2005].

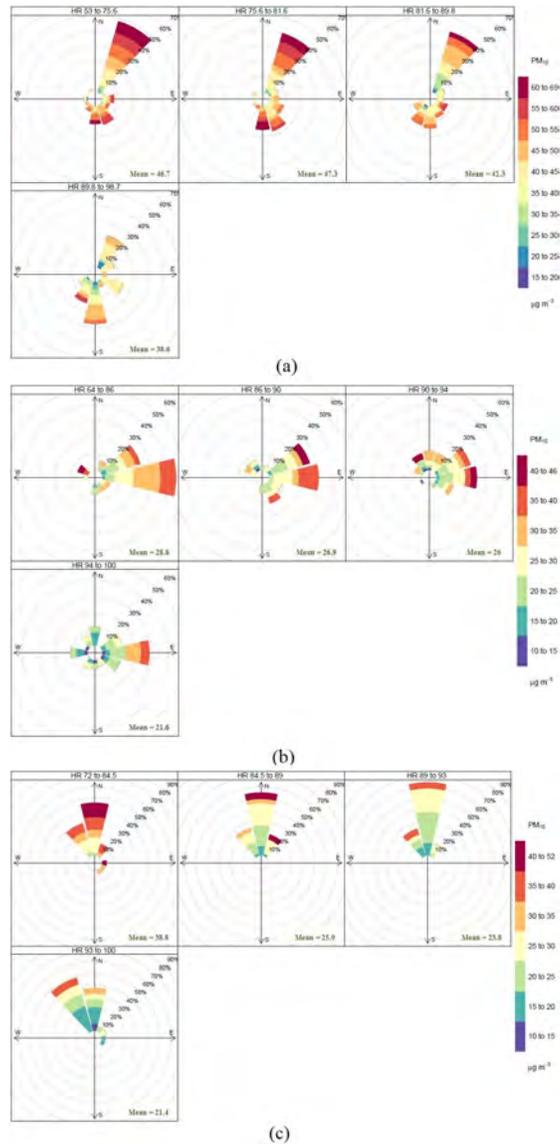


Figura 3.8: Porcentaje de contribución proporcional a la media del contaminante  $PM_{10}$  en función de la velocidad y dirección del viento para diferentes intervalos de humedad relativa en a) Liceo, b) Palogrande y c) Nubia.

En términos del análisis de la influencia de la temperatura, se presenta un efecto contrario en comparación con lo ocurrido con la precipitación y humedad relativa. Para las tres estaciones analizadas, se puede observar cómo el incremento en la temperatura, genera mayores niveles de material particulado (Figura 3.9a-c). Las estaciones de monitoreo, muestran incrementos en las concentraciones medias de  $PM_{10}$  para cada uno de los intervalos de temperatura especificados por Openair. Este comportamiento puede estar asociado a un incremento de la actividad fotoquímica en la troposfera en días con mayor temperatura y, por ende, mayor intensidad de la radiación solar, con la subsecuente formación de material particulado secundario [Chaloulakou, 2003]; [González, 2012]. Sumado a esto, los incrementos en la temperatura ambiente son un indicativo de disminución en la ocurrencia de precipitación (temporadas secas o de lluvias bajas), ocasionando una reducción en los procesos de lavado atmosférico por acción de las lluvias.

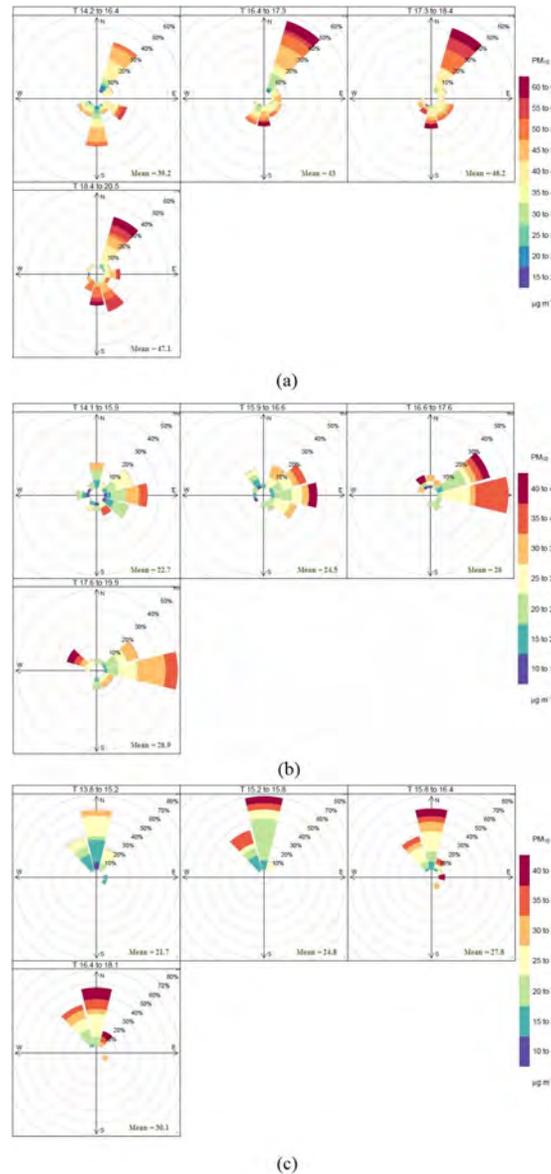


Figura 3.9: Porcentaje de contribución proporcional a la media del contaminante  $PM_{10}$  en función de la velocidad y dirección del viento para diferentes intervalos de temperatura en a) Liceo, b) Palogrande y c) Nubia.

### 3.5.2. Análisis de $PM_{10}$ y su correlación con las variables meteorológicas mediante la implementación de MATLAB

Algunas técnicas para descubrir información a partir de datos incluyen: Procesamiento de datos (depuración, detección de valores atípicos, tratamiento de valores faltantes, transformación y creación de variables), reducción de datos y proyección, visualización, agrupamiento y estimación de densidad, métodos de clasificación y regresión, análisis de asociación, redes neuronales artificiales (ANNs), entre otras [Gibert, 2008].

Con base en lo anterior, utilizando las herramientas del *Toolbox* y del *PRTools* de MATLAB 7.0 (Mathworks, Natick MA, USA), fue realizado el análisis gráfico y estadístico a los mismos datos de  $PM_{10}$  de las estaciones mencionadas.

Los resultados y análisis en esta sección, comprenden el uso de gráficas que relacionan el  $PM_{10}$

con dos y tres variables meteorológicas (técnica de visualización), la aplicación de regresiones lineales (técnica de clasificación y regresión) y el PCA (técnica de reducción de datos y proyección) aplicados al conjunto de datos.

### 3.5.2.1. Técnica de visualización de datos empleando gráficas de dos y tres variables

Los comandos de generación de gráficos de MATLAB, mencionados en la Tabla 3.6, presentan los datos con dos y tres parámetros, los cuales relacionan cada una de las variables meteorológicas con los niveles de  $PM_{10}$  en diferentes combinaciones. Además, la visualización de los datos es mejorada implementando una escala de color para el  $PM_{10}$ , que da color rojo a los mayores valores de concentración y va cambiando de tonalidad a medida que disminuye el nivel de concentración hasta llegar al azul, el cual representa el mínimo valor de concentración.

Inicialmente, fueron graficados los niveles de concentración de  $PM_{10}$  con cada una de las variables meteorológicas de forma independiente. Posteriormente, fueron graficados los niveles de  $PM_{10}$  respecto a dos variables meteorológicas simultáneamente. Los gráficos de estas posibles combinaciones, permiten la observación de la distribución de datos, la verificación de agrupamientos particulares, la detección de valores atípicos (no usuales) y la identificación de posibles tendencias de los mismos.

Algunas de estas combinaciones son mostradas en la Figura 3.10a-f, para las tres estaciones de calidad del aire. Allí se presentan los datos de  $PM_{10}$  respecto a la precipitación y al porcentaje de humedad relativa. Las figuras fueron elaboradas empleando el comando *scatter*. Las Figuras 3.10a, 3.10c y 3.10e, muestran un comportamiento similar del  $PM_{10}$  respecto a la precipitación para todas las estaciones. En estas se observa que el mayor número de datos de  $PM_{10}$  coinciden con niveles de precipitación bajos (inferiores a 20 mm de lluvia) o nulos (0 mm de lluvia). Esta tendencia, confirma la mayor acción contaminante del  $PM_{10}$  en ausencia de la lluvia, comportamiento que se ajusta al descrito por [Elminir, 2005] en el Cairo (Egipto), [Lee, 2011] en Seúl (Korea) y [González, 2012] en Manizales (Colombia).

Más puntualmente, en estas figuras se observan algunos valores aislados (atípicos) del conjunto de datos para la estación Liceo y la estación Palogrande que se acogen a esta tendencia. En ellos, se identifican los máximos niveles de precipitación para niveles bajos de  $PM_{10}$ . En la estación Liceo, se tiene que los puntos corresponden a 83.76 mm de precipitación para una concentración de  $35 \mu\text{g}/\text{m}^3$ , 123.12 mm para  $36 \mu\text{g}/\text{m}^3$ ; y en Palogrande de 51.82 mm para  $14 \mu\text{g}/\text{m}^3$ , y 52.1 mm para  $14 \mu\text{g}/\text{m}^3$ . También, se identifican los niveles máximos de  $PM_{10}$  con los niveles más bajos de precipitación, como valores atípicos del conjunto de datos:  $68 \mu\text{g}/\text{m}^3$  para 0 mm, y  $69 \mu\text{g}/\text{m}^3$  para 0 mm.

Las Figuras 3.10b, 3.10d y 3.10f, revelan una tendencia decreciente entre los niveles de concentración de  $PM_{10}$  y la humedad relativa del aire. En ellas, se observan los menores niveles de  $PM_{10}$  con valores de humedad altos. Lo que es un hecho directamente vinculante con el evento de precipitación, en donde generalmente se registran mayores valores de humedad cuando llueve y, por lo tanto, mayor remoción atmosférica de  $PM_{10}$  en el aire.

A partir de las gráficas mencionadas, se observa que en ninguna de las tres estaciones se excede el límite de nivel máximo permisible por la legislación colombiana actual para el  $PM_{10}$  monitoreado en 24 horas de  $100 \mu\text{g}/\text{m}^3$  [Ministerio de Ambiente, 2010a]. Sin embargo, algunos datos sí superan los límites recomendados por la OMS en 24 horas de  $50 \mu\text{g}/\text{m}^3$  [WHO, 2006] en dos de las estaciones: para la estación Liceo, el 25 % de los datos superan este nivel y para la estación Nubia, el 1 %. De esta forma, se hace evidente la mayor emisión de partículas provenientes del proceso de combustión de diesel en la estación en donde se concentra el mayor tráfico vehicular de las tres estaciones (estación Liceo). Tendencias similares han sido obtenidas por otros autores [Velasco, 2006]; [González, 2012].

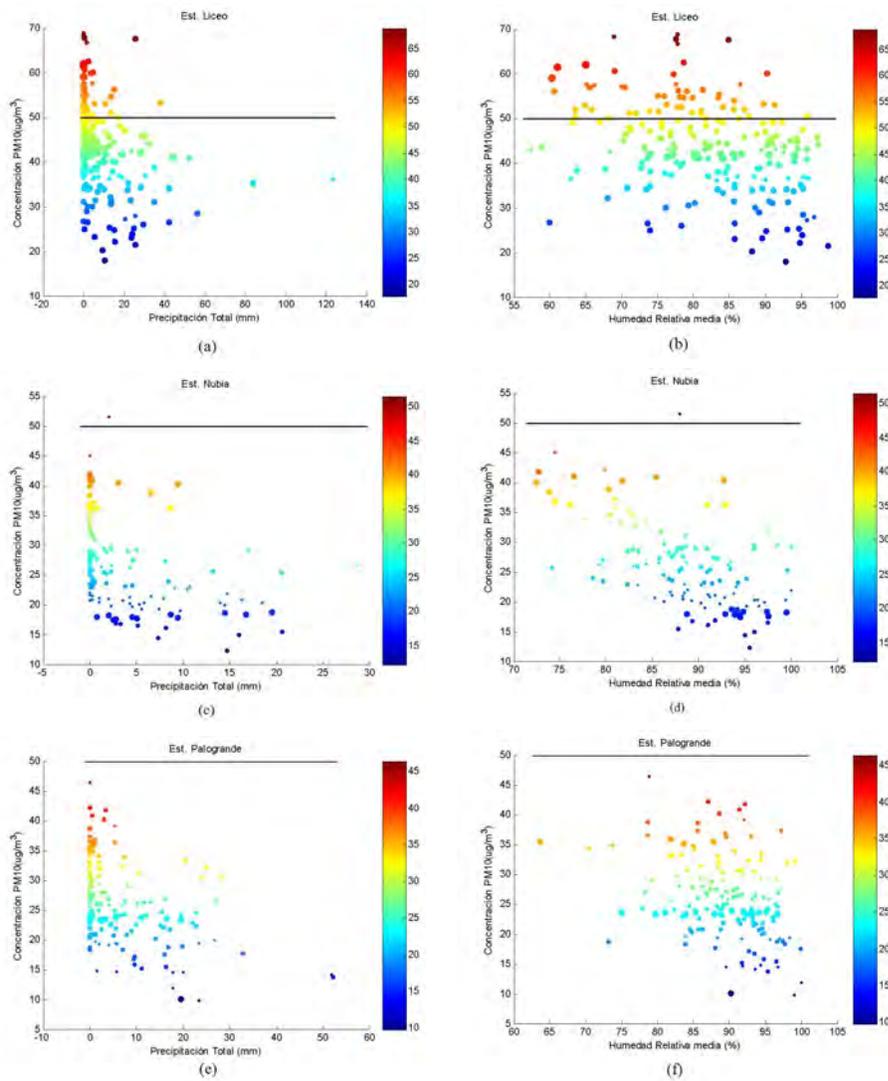


Figura 3.10: Gráficas que relacionan el nivel de  $PM_{10}$  con una variable meteorológica para las tres estaciones: a), c) y e)  $PM_{10}$  respecto a la Precipitación Total; b), d) y f)  $PM_{10}$  respecto al Porcentaje de Humedad Relativa. La escala de color hace referencia a los niveles de concentración de  $PM_{10}$ : El color rojo indica mayor valor de concentración y el color azul indica menor valor de concentración.

No obstante, las fuentes de emisión antropogénicas no son las únicas generadoras de estos niveles de partículas en la ciudad. A 28 km de Manizales, se encuentra el Volcán Nevado del Ruiz, el cual comenzó a registrar emisiones de gases y partículas más intensas y continuas durante los meses del primer semestre del 2012. En la Tabla 3.8, se observa el incremento mensual de los niveles de  $PM_{10}$ , a partir del mes de enero de 2012, en comparación con los promedios de los últimos meses del 2011. Esta información, también es soportada a partir de la Tabla 3.7, para la cual se evidencia el aumento del promedio anual del  $PM_{10}$  entre el 2011 y 2012, y a partir de los reportes del Observatorio Vulcanológico y Sismológico de Manizales [Observatorio, 2013], con los cuales se informó a la comunidad el cambio de actividad desde un Nivel III (amarilla: Cambios en el comportamiento del volcán) a un Nivel II (naranja: Erupción probable).

En las estaciones de calidad del aire, fueron detectados algunos valores máximos de  $PM_{10}$  en este intervalo de tiempo. Para Liceo, algunos valores máximos de  $PM_{10}$  fueron  $69 \mu g/m^3$  y  $68 \mu g/m^3$  registrados el 28 de Marzo y el 18 de Abril de 2012, respectivamente; para la Nubia,

fue de  $52 \mu\text{g}/\text{m}^3$ , registrado el 5 de Junio de 2012; y para Palogrande, fueron de 46 y  $42 \mu\text{g}/\text{m}^3$  registrados el 24 de Febrero, 16 de Marzo y 30 de Mayo de 2012, respectivamente. Estas variaciones pueden ser atribuidas a los aportes de partículas provenientes de esta fuente natural, las cuales logran ser detectadas o no por las estaciones de calidad del aire debido a la dinámica del viento que direcciona las emisiones.

Tabla 3.8: Promedios mensuales de  $PM_{10}$  para las tres estaciones desde Noviembre de 2011 hasta Junio de 2012

Mes	Promedios Mensuales ( $\mu\text{g}/\text{m}^3$ )		
	Estación Palogrande	Estación Nubia	Estación Liceo
Nov. 2011	19	21	46
Dic. 2011	23	22	44
Enero 2012	25	30	44
Febrero 2012	38	27	50
Marzo 2012	30	25	46
Abril 2012	26	30	43
Mayo 2012	29	33	43
Junio 2012	29	35	47

En la Figura 3.11a-f, se presenta la visualización de los datos de  $PM_{10}$ , dados por la escala de color, respecto a dos variables meteorológicas. En estas gráficas, se relacionan simultáneamente variables como radiación solar y velocidad del viento que, en términos de contaminación atmosférica, son de vital importancia para la comprensión de la estabilidad atmosférica de la ciudad.

Las bajas velocidades del viento y los bajos valores de radiación solar presentes en la ciudad, provocan una baja dispersión de la contaminación y, por lo tanto, mayores concentraciones de  $PM_{10}$ .

Las condiciones de estabilidad e inestabilidad atmosférica están determinadas por la capacidad de dispersión o dilución de los contaminantes en la baja atmósfera. De esta forma, las condiciones de radiación y velocidad del viento determinan las categorías de Pasquill que van desde la A (Extremadamente inestable) hasta la F (Extremadamente estable) [Santa Cruz, 2000]; [Zoras, 2006]; [Moragues, 2012]. A las condiciones de Manizales, le corresponde la categoría C. Esta se encuentra definida por la radiación solar promedio inferior a  $350 \text{ W}/\text{m}^2$  y la velocidad del viento promedio de 0 a  $0.8 \text{ m}/\text{s}$ .

Las Figuras 3.11a, 3.11c y 3.11e, muestran el comportamiento de estas dos variables en las tres estaciones. De acuerdo, a la categoría C de Manizales, se esperaría una agrupación de los datos en la zona inferior de las gráficas que cumplieran con las condiciones mencionadas de radiación solar y velocidad del viento para esta categoría, tal como lo muestra el esquema conceptual de la Figura 3.12.

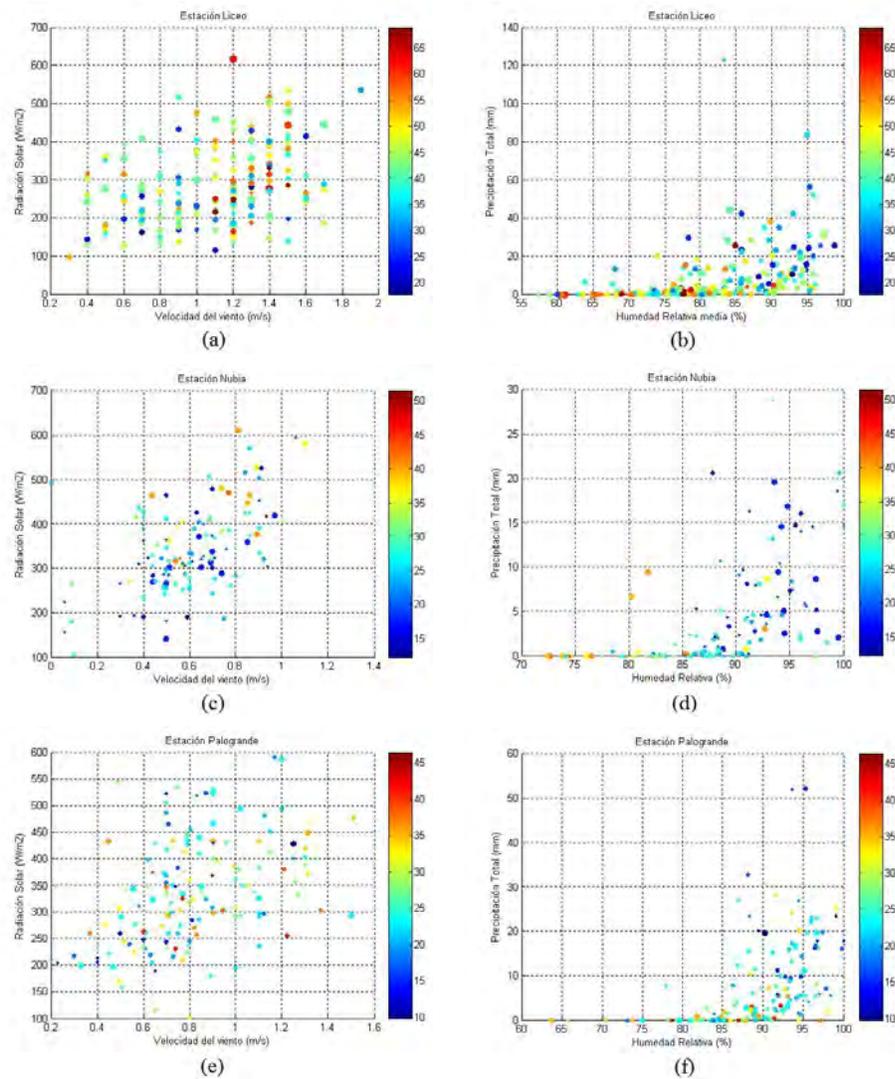


Figura 3.11: Gráficas que relacionan el nivel de  $PM_{10}$  con dos variables meteorológicas para las tres estaciones: a), c) y e) Relación del  $PM_{10}$  respecto a la Radiación Solar y la Velocidad del Viento; b), d) y f) Relación del  $PM_{10}$  respecto a la Precipitación y el Porcentaje de Humedad Relativa. Escala de color hace referencia a los niveles de concentración de  $PM_{10}$ : El color rojo indica mayor valor de concentración y el color azul indica menor valor de concentración.

Realmente, los datos de las tres estaciones cumplen con esta tendencia. Cada estación refleja la aglomeración de más del 50% de los datos dentro de esta categoría: para Liceo el 75%, para Nubia el 56% y para Palogrande el 60%. Con esto, estadísticamente los datos reflejan el comportamiento ligeramente inestable de la baja atmósfera en la ciudad.

En general, las gráficas que relacionan más de dos variables, ofrecen una introducción, en forma visual, al análisis estadístico multivariado (MVA) de este conjunto de datos.

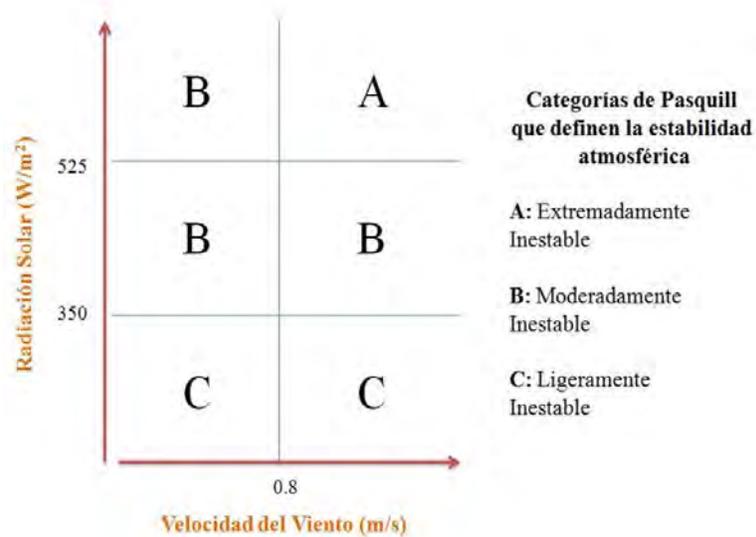


Figura 3.12: Esquema conceptual de la estabilidad atmosférica en Manizales

### 3.5.2.2. Aplicación de regresiones lineales a los datos de $PM_{10}$ respecto a las variables meteorológicas (Técnica de regresión)

En la clasificación y regresión, la identidad de la distribución se conoce a priori y la meta es encontrar aquellas variables que mejor explican la distribución, ya sea con fines descriptivos o predictivos. La regresión lineal clásica es una buena técnica para lograr este objetivo y encontrar la mejor ecuación lineal que defina la relación entre una respuesta numérica y las variables independientes [Gibert, 2008].

La Tabla 3.9, presenta el resumen de los resultados obtenidos de la aplicación de una regresión lineal simple a cada conjunto de datos de  $PM_{10}$  y cada una de las variables meteorológicas. Estas regresiones fueron realizadas utilizando la herramienta *cfTool*.

Los valores resaltados corresponden a las regresiones cuyos resultados indican un mejor ajuste. Esta condición está dada por parámetros estadísticos como el SSE (Suma del Cuadrado del Error) y el RMSE (Raíz del Error Cuadrático Medio) que son medidas del error, cuyo valor debe ser mínimo, y el  $R^2$  (Coeficiente de Determinación) que corresponde al nivel de ajuste de la regresión y cuyo valor debe ser máximo en un rango de 0 a 1.

A partir de la aplicación de esta herramienta, se observa que la estación Liceo y Nubia tienen los mejores parámetros de ajuste lineal con la Humedad Relativa, y la estación Palogrande con la Precipitación.

Tabla 3.9: Parámetros estadísticos para las regresiones lineales realizadas entre las concentraciones de  $PM_{10}$  y cada variable meteorológica de las tres estaciones de monitoreo

Parámetro estadístico	Temperatura	Velocidad del viento	Presión barométrica	Humedad relativa	Precipitación	Radiación
<b>Estación Liceo</b>						
SSE	1.831e+4	2.044e+4	2.033e+4	<b>1.79e+4</b>	1.857e+4	1.815e+4
$R^2$	0.1156	0.01259	0.01795	<b>0.1355</b>	0.1029	0.02331
RMSE	9.405	9.937	9.91	<b>9.298</b>	9.472	9.748
<b>Estación Nubia</b>						
SSE	5411	6371	6546	<b>4353</b>	5709	5812
$R^2$	0.1918	0.0484	0.02234	<b>0.3499</b>	0.1473	0.1319
RMSE	6.66	7.227	7.325	<b>5.973</b>	6.841	6.902
<b>Estación Palogrande</b>						
SSE	6384	7067	6911	6364	<b>5895</b>	7280
$R^2$	0.1232	0.02928	0.05077	0.1259	<b>0.1903</b>	4.879e-5
RMSE	6.524	6.864	6.788	6.513	<b>6.269</b>	6.967

Los datos resaltados en la Tabla 3.9, pese a que tienen los mejores valores de correlación, no indican una fuerte dependencia lineal entre el  $PM_{10}$  y dichas variables meteorológicas. Esto también se evidencia claramente al observar la distribución de los datos en las Figuras 3.10 y 3.11.

Lo anterior, muestra que este conjunto de datos es altamente no lineal y, por esto, los parámetros resultantes del ajuste a una línea recta no tienen valores estadísticamente representativos. El estudio de [Elminir, 2005], también muestra una tendencia no lineal al relacionar los niveles de contaminantes, como  $O_3$ ,  $SO_2$ ,  $PM_{10}$ ,  $CO$  y  $CO_2$ , respecto a una variable meteorológica.

En general, el análisis de sistemas ambientales, enfrenta una complejidad derivada de varias fuentes: la multidisciplinariedad, el dominio no lineal, alta dimensionalidad, multiescalaridad, heterogeneidad, cualidad no estacionaria, baja manejabilidad, incertidumbre, información imprecisa, comportamiento cíclico, entre otros, lo que implica la aplicación de técnicas diferentes a las convencionales para mejorar la comprensión del comportamiento de los datos y sus relaciones específicas [Gibert, 2008]. Una de estas técnicas corresponde al PCA.

### 3.5.2.3. Aplicación del Análisis de Componentes Principales - PCA (Técnica de reducción de datos y proyección)

El PCA es un método estadístico empleado para encontrar la proyección a un plano – o más generalmente a un subespacio lineal – el cual preserva tanto como es posible la varianza original del dato [van der Heijden, 2004]; [Berthold, 2010]. El PCA proporciona una guía sobre cómo reducir un conjunto de datos complejos a uno de dimensión inferior para revelar las relaciones entre variables que no son observadas a simple vista [Shlens, 2009].

En términos generales, el método toma una matriz de  $n$  datos y  $p$  variables, o atributos, los cuales pueden ser correlacionados y resumidos en ejes no correlacionados (componentes principales o ejes principales), que son combinaciones lineales de las  $p$  variables originales. Los primeros componentes muestran tanto como sea posible la variación entre los datos [Pla, 1986]; [Peña, 2002]; [Orozco, 2005].

El objetivo de PCA es principalmente girar los ejes de este espacio  $p$ -dimensional a las nuevas posiciones (ejes principales) que tienen las siguientes propiedades:

- El ordenamiento consiste en que el eje principal 1 tiene la más alta varianza, el eje 2 tiene la siguiente varianza más alta, y el eje  $p$  tiene la menor varianza.
- La covarianza entre cada par de los ejes principales es cero (los ejes principales no están correlacionados).

El PCA ha sido utilizado ampliamente en estudios de contaminación atmosférica para el análisis de la presencia de congéneres de Compuestos Orgánicos Persistentes (COPs), como dioxinas (PCDDs), furanos (PCDFs), dioxin like PCBs (dlPCBs), y éteres bifenílicos polibromados (PBDEs) en España [Martínez, 2010]; [Rivera-Austrui, 2011]), China ([Li, 2011]; [Ren, 2011], Brasil [de Assuncao, 2005], entre otros países, para los cuales se han utilizado diferentes software en sus estudios. El PCA en MATLAB, también ha sido usado en estudios de contaminación atmosférica de compuestos como *VOCs* y *O<sub>3</sub>*, y su relación con las condiciones climáticas y meteorológicas en España y Turquía [Pérez-Rial, 2010]; [Ozbay, 2011], respectivamente. Estos han proporcionado información importante acerca de las posibles fuentes de generación y distribución de los mismos.

Para este estudio, el PCA fue aplicado a los registros de las 6 variables meteorológicas utilizando los comandos *pca*, *printcomp* y *pcacov* del *PRTools*. Los niveles de *PM<sub>10</sub>* no fueron incluidos como parte de estos datos. En su lugar, fueron utilizados como etiquetas subdivididas en 3 clases (niveles de concentración), con el fin de detectar algún agrupamiento (*clustering*) o comportamiento semejante de las variables meteorológicas vinculadas a los altos o bajos niveles de *PM<sub>10</sub>*. Estas clases, fueron definidas en cada estación de acuerdo a la concentración máxima y mínima registradas, siguiendo la siguiente ecuación:

$$RCE_i = \frac{Concen.Max. - Concen.Min.}{3} \quad (3.1)$$

Donde RCE es el Rango de cada clase en la estación *i*.

Previo al análisis, fue necesario el escalamiento de los datos utilizando la desviación estándar, debido a la diferencia significativa que hay entre la magnitud de las unidades de medida de las variables en cuestión: esto es el *PM<sub>10</sub>* tiene magnitud de 10 a 68  $\mu\text{g}/\text{m}^3$  y la Radiación Solar de 100 a 600  $\text{W}/\text{m}^2$ . De esta manera, el escalamiento de los datos permite normalizar la magnitud de medida entre cada par de variables y así evitar que tomen más relevancia las variables de magnitud mayor. Es importante tener en cuenta cualquier tratamiento matemático que se le realice a los datos que serán ingresados en el PCA, ya que este es un método sensible a los cambios de escala [Pla, 1986].

La aplicación del PCA, requirió el manejo de procedimientos de cálculo estadístico como matrices de covarianzas, matrices de valores propios (*Eigenvalues*), proporciones de varianzas, porcentajes de varianza acumulada, porcentajes de varianza explicada, los cuales desempeñaron criterios analíticos a la hora de interpretar los resultados que esta técnica arroja.

Con base en algunos de estos cálculos estadísticos, varios autores sugieren el uso de algunas reglas para la selección del número de componentes [Peña, 2002]; [Orozco, 2005]:

- Realizar un gráfico de los valores propios frente al número de componentes principales (Gráfico Scree), buscando el “codo” o “quiebre” en el gráfico a partir del cual los valores propios son aproximadamente iguales.
- Seleccionar componentes hasta cubrir una proporción determinada de varianza, como del 80% o el 90%. Esta regla es arbitraria y debe aplicarse de acuerdo a cada caso.
- Desechar aquellos componentes asociados a valores propios inferiores a una cota, que suele fijarse como la varianza media. En particular, cuando se trabaja con la matriz de correlación, el valor medio de los componentes es 1, y conlleva a seleccionar los valores propios mayores que la unidad. Esta regla también es arbitraria y debe aplicarse de acuerdo a cada caso.

Al aplicar PCA a los datos de este estudio, se obtuvo que, de acuerdo a los valores de las varianzas y a los porcentajes de varianza acumulada para las tres estaciones en cada uno de los componentes principales (Tabla 3.10), las tres primeras componentes principales retienen más del 80% de la variabilidad original de los datos: Liceo y Nubia con el 81% y Palogrande con el 83%. Además, pese a que solo las dos primeras componentes tienen una varianza superior al promedio, se incluyó la tercera porque el 70% de la varianza acumulada, con dos componentes, puede no ser suficiente para representar la variabilidad de cada sistema. Esto se confirma al observar las Gráficas Scree (Figura 3.13a-f), en donde el “codo” o “quiebre” de las mismas ocurre a partir del segundo componente principal.

Tabla 3.10: Componentes Principales evaluados para los datos de las tres estaciones

Componentes principales	Liceo		Nubia		Palogrande	
	V.	%V.A.	V.	%V.A.	V.	%V.A.
1	<b>2.879</b>	<b>49</b>	<b>3.010</b>	<b>50</b>	<b>3.079</b>	<b>52</b>
2	<b>1.161</b>	<b>68</b>	<b>1.028</b>	<b>67</b>	<b>1.088</b>	<b>70</b>
3	<b>0.789</b>	<b>81</b>	<b>0.815</b>	<b>81</b>	<b>0.799</b>	<b>83</b>
4	0.600	91	0.572	90	0.549	92
5	0.330	97	0.325	96	0.3026	97
6	0.176	100	0.250	100	0.150	100
Promedio	0.989		1.000		0.994	

*V: Varianzas*

*%V.A.: Porcentaje de varianza acumulada*

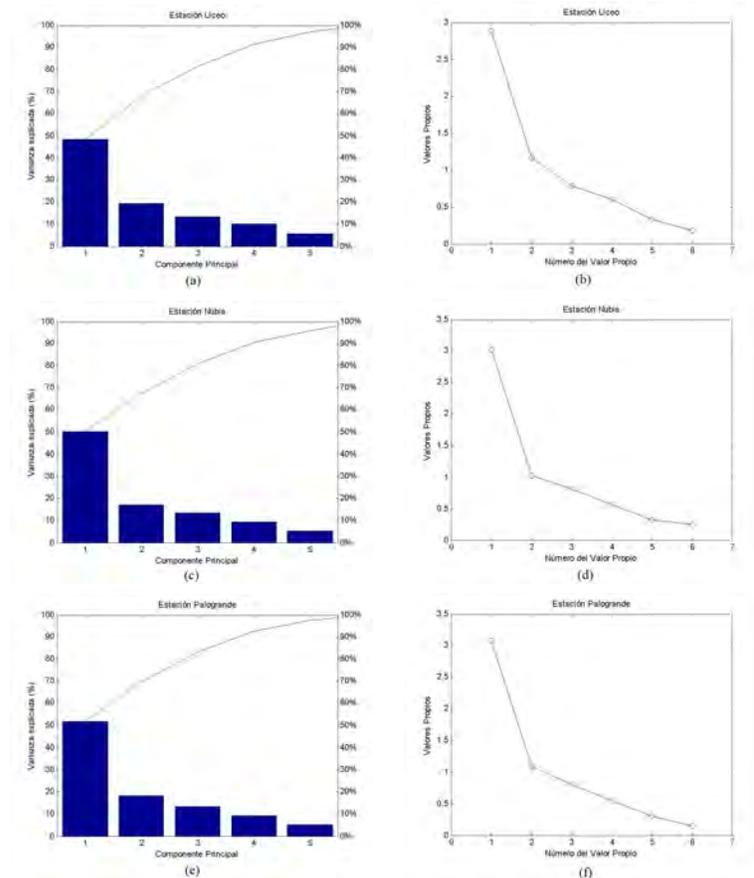


Figura 3.13: Gráficas Scree para las tres estaciones de monitoreo

La visualización de la proyección de los datos en dos y tres componentes principales (Figura 3.14a-f), con los datos de  $PM_{10}$  etiquetados en 3 clases o rangos de concentración, no proporciona información acerca de agrupamientos especiales de los mismos.

La alta no linealidad de este sistema, tal como se mostró en los resultados de la regresión lineal, genera una gran tendencia dispersiva de los datos, lo que impide el agrupamiento y aislamiento de datos con características similares que muestren información relevante a partir de estas representaciones gráficas. Esto significa que las variables meteorológicas no vinculan los niveles de  $PM_{10}$  etiquetados por rangos de concentración mayor, media o menor, con patrones característicos comunes.

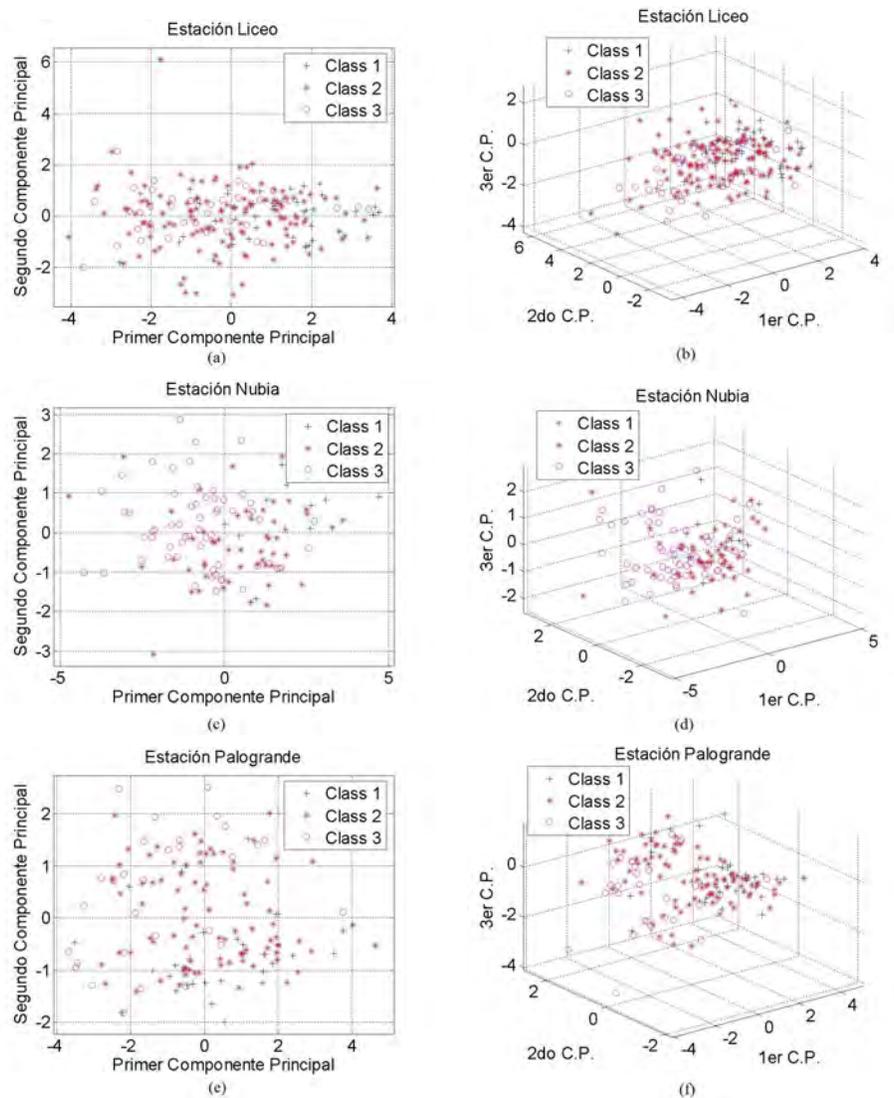


Figura 3.14: Proyección de los datos en dos y tres componentes principales de las tres estaciones de monitoreo por clases. a) y b) Estación Liceo; Clase 1: Concentraciones mayores de  $52 \mu\text{g}/\text{m}^3$ , Clase 2: Concentraciones entre  $35$  y  $52 \mu\text{g}/\text{m}^3$ , Clase 3: Concentraciones menores a  $35 \mu\text{g}/\text{m}^3$ . c) y d) Estación Nubia; Clase 1: Concentraciones mayores de  $39 \mu\text{g}/\text{m}^3$ , Clase 2: Concentraciones entre  $25$  y  $39 \mu\text{g}/\text{m}^3$ , Clase 3: Concentraciones menores a  $25 \mu\text{g}/\text{m}^3$ . e) y f) Estación Palogrande; Clase 1: Concentraciones mayores de  $34 \mu\text{g}/\text{m}^3$ , Clase 2: Concentraciones entre  $22$  y  $34 \mu\text{g}/\text{m}^3$ , Clase 3: Concentraciones menores a  $22 \mu\text{g}/\text{m}^3$

Sin embargo, la inspección de la matriz de valores propios para las 3 primeras componentes principales seleccionadas (Tabla 3.11), entregan información de las cargas (*loadings*) que dan más peso a las variables con más influencia en el sistema. La presencia del signo negativo en los valores propios, indica una relación inversa con las demás variables meteorológicas del mismo componente principal (CP).

Para la estación Liceo, en la 1ª CP, se observa que tiene la mayor correlación positiva (valores en verde) con la humedad relativa y la precipitación, mientras que la mayor correlación negativa (valores en rojo) la tiene con la temperatura, la radiación solar y la velocidad del viento. En la 2ª CP, la mayor correlación positiva la tiene con la presión barométrica, mientras que la correlación negativa la tiene con la precipitación y la velocidad del viento. En la 3ª CP, la mayor correlación positiva la tiene con la precipitación, la humedad relativa, la radiación solar y la velocidad del viento. Los demás valores, siendo cercanos a cero, no representan ninguna correlación con las CPs. Así, la misma interpretación se cumple para las dos estaciones restantes, Nubia y Palogrande.

Tabla 3.11: Cargas de las variables meteorológicas en cada Componente Principal (Matriz de valores propios).

Variables meteorológicas	Liceo			Nubia			Palogrande		
	1ª C.P.	2ª C.P.	3ª C.P.	1ª C.P.	2ª C.P.	3ª C.P.	1ª C.P.	2ª C.P.	3ª C.P.
Temperatura	-0.53	0.09	-0.02	-0.47	0.11	-0.26	-0.51	0.10	-0.04
Velocidad del Viento	-0.35	-0.44	0.28	-0.36	-0.48	0.39	-0.39	-0.36	-0.04
Presión Barométrica	0.05	0.75	0.62	0.20	-0.70	-0.69	-0.06	0.88	-0.33
Humedad Relativa	0.56	-0.06	0.01	0.51	0.05	0.18	0.53	-0.01	0.04
Precipitación	0.28	-0.48	0.67	0.38	-0.41	0.46	0.32	-0.27	-0.84
Radiación Solar	-0.45	-0.07	0.31	-0.45	-0.32	0.25	-0.44	-0.12	-0.43

*C.P.: Componente Principal*

En términos estadísticos, las 1ª CPs de todas las estaciones (que representa entre el 49% y el 52% de la varianza), confluyen en su fuerte relación con la humedad relativa y la precipitación en la dinámica de cada estación, e igualmente fuerte para su relación (inversa) con la temperatura, la radiación solar y la velocidad del viento.

Ahora, en términos de contaminación atmosférica, la precipitación (implícita la humedad relativa) juega un papel de gran importancia como variable meteorológica generadora del fenómeno de depuración de contaminantes en la atmósfera mediante la acción de la lluvia (*scavenging*), como se ha descrito a lo largo del capítulo. Lo cual corresponde a la acción climática inversa de la radiación solar, ya que, comúnmente la generación de nubosidad a la que preceden los eventos de lluvia, impide en gran medida el paso de la radiación solar a la tierra. Además, discurre el hecho de la participación de la radiación solar (implícitamente la temperatura) y la velocidad del viento en la definición de estabilidad atmosférica de la ciudad. La precipitación no solo tiene importancia en la 1ª CP, directa o inversamente, esta variable tiene influencia en todos los componentes de las tres estaciones.

En la 2ª CP de todas las estaciones, la presión barométrica tiene el mayor valor de correlación (positivo o negativo), lo que indica un efecto de la altura en donde están ubicadas las estaciones, como se observa en la Tabla 3.3; esto es a mayor altura, menor presión. La magnitud de los valores propios (independientemente del signo) coincide en orden descendente con la altura, en m.s.n.m., a la cual se encuentran los equipos de medición. De este modo, Palogrande (0.879 y 2166 m.s.n.m.) > Liceo (0.75 y 2156 m.s.n.m.) > Nubia (0.697 y 2104 m.s.n.m.).

Este comportamiento, es atribuido a las características topográficas de la ciudad. Manizales se encuentra situada en las laderas occidentales de la cordillera central de los Andes, a 2150 m.s.n.m., y su desarrollo urbano se extiende en un área de fuertes pendientes.

Para el mismo CP en todas las estaciones, la velocidad del viento también tiene una relevancia importante en los sistemas estudiados, debido a que la circulación del aire diluye los contaminantes en la atmósfera. Para esta variable, los valores de correlación lo tienen, en orden descendente, Nubia (0.48) > Liceo (0.44) > Palogrande (0.36), los cuales pueden ser asociados con los valores de correlación de la presión barométrica, ya que los movimientos de las masas de aire se mueven desde las regiones de presiones más altas a las de presiones más bajas y, precisamente, la Nubia (menor altura) tiene mayor presión y mayor peso de la variable velocidad del viento, lugar desde donde provienen las masas de aire en ciertas franjas horarias del día, hasta llegar a Palogrande (mayor altura) con menor presión barométrica y menor peso en la variable velocidad del viento. Esto hace a la presión barométrica y la velocidad del viento, variables que se correlacionan inversamente dadas las condiciones topográficas de la ciudad.

Para la 3ª CP, también en todas las estaciones, la radiación solar y la presión barométrica volvieron a obtener significancia en el análisis.

Al no mencionar en el análisis algunos valores resaltados, no significa que fueran menos importantes. Antes bien, algunas variables meteorológicas incluyen implícitamente la influencia de otra de ellas. Al igual que los CP que conservan menores porcentajes de varianza (superiores a la 3ª CP), al no ser mencionados, no implica que no pertenezcan al estudio, sino que involucran variables y relaciones ya mencionadas.

El análisis de PCA elaborado en MATLAB, contiene muchas más herramientas y comandos que permiten un análisis más detallado de esta técnica estadística, tales como la identificación de agrupamientos encontrados automáticamente por el software a partir de los datos ingresados, cálculo de otros parámetros estadísticos, entre otros, que serán ampliados en estudios posteriores.

Así, el manejo de las herramientas de MATLAB, facilita el estudio estadístico de un conjunto de datos, que para este caso corresponden a registros de contaminación atmosférica y meteorología en Manizales, logrando la generación de conocimiento a partir de la elaboración de gráficos, regresiones lineales simples y análisis multivariados aplicados a estos.

### 3.6. Conclusiones

El análisis de los patrones de vientos para las estaciones que poseen datos horarios, permitió corroborar la existencia de dos patrones de dirección en la ciudad de Manizales. Un patrón para el periodo diurno, donde se tienen direcciones predominantes provenientes del norte (N), nor-noroeste (NNW) y noroeste (NW); en comparación con el periodo nocturno, para el cual prevalecen vientos provenientes del este (E), este-sureste (ESE) y sureste (SE).

Los datos de calidad del aire recolectados en las estaciones de monitoreo presentan uniformidad y se encuentran dentro de un rango estadísticamente satisfactorio según el análisis de cajas, teniendo en cuenta que la mayoría de los datos obtenidos se encuentran entre la distribución de las cajas y sus cuartiles superior e inferior.

Los valores medios anuales de  $PM_{10}$  obtenidos para las estaciones de monitoreo están por debajo del límite anual establecido en la Resolución 610 de 2010 de  $50 \mu\text{g}/\text{m}^3$ , no obstante, todas las estaciones están por encima del límite promedio anual recomendado por la OMS de  $20 \mu\text{g}/\text{m}^3$ . Una mayor distribución temporal de  $PM_{10}$  en Liceo, en comparación con las estaciones Palogrande y Nubia, es un indicativo de la influencia de las emisiones provenientes de fuentes móviles en los niveles de  $PM_{10}$  en la ciudad.

La aplicación de estas herramientas, evidenció el fenómeno de lavado atmosférico (*scavenging*) asociado a la depuración atmosférica de  $PM_{10}$  por acción de la lluvia en Manizales, mediante la relación gráfica y estadística de las concentraciones de  $PM_{10}$  con los valores de precipitación registrados en las estaciones de calidad del aire y meteorológicas. Fue verificada la estabilidad atmosférica de la ciudad (ligeramente inestable) mediante la distribución de los niveles de  $PM_{10}$  respecto a la radiación solar y la velocidad viento. Esta última variable meteorológica, a pesar de presentar valores bajos (promedio) en la ciudad, está sujeta a una dinámica diurna y nocturna diferente debido al cambio de dirección predominante del viento del día a la noche.

La aplicación de herramientas para el análisis de datos de calidad del aire, como es el caso del lenguaje R (con su librería Openair) y MATLAB, son de gran utilidad para la ciudad de Manizales, ya que garantizan la obtención de forma ágil y sistemática de datos estadísticos y gráficos que permiten un análisis profundo de la dinámica de contaminantes como el  $PM_{10}$ . Estas herramientas y los resultados obtenidos son de utilidad tanto para los grupos de investigación como para la autoridad ambiental.

Específicamente con la aplicación de Openair al análisis de datos de  $PM_{10}$  en las tres estaciones de monitoreo objeto de estudio, se puede concluir que esta herramienta es útil en el análisis de series temporales, comparación entre estaciones de monitoreo y el estudio de la influencia de variables meteorológicas sobre el comportamiento del material particulado en la ciudad de Manizales, así como la elaboración de rosas de vientos para la ciudad.

Por otra parte, la aplicación de las herramientas de MATLAB, permite la implementación de técnicas de visualización, técnicas de clasificación y regresión, técnicas de reducción de datos y proyección, técnicas de agrupación, y técnicas aún más avanzadas para el estudio de contaminantes de importancia en el monitoreo de calidad del aire y su relación con las variables meteorológicas.

## Agradecimientos

Los autores expresan sus agradecimientos a la Vicerrectoría de Investigación y la Dirección de Investigaciones de Manizales (DIMA), de la Universidad Nacional de Colombia sede Manizales, por el apoyo brindado a través del "Programa de fortalecimiento de capacidades conjuntas para el procesamiento y análisis de información ambiental, código: 12677" y del proyecto "Grupo de Trabajo Académico en Ingeniería Hidráulica y Ambiental, código: 16075". Este trabajo también es apoyado por el "Programa Jóvenes Investigadores e Innovadores 2010, convenio especial de cooperación No. 146 de 2011 suscrito entre la Fiduciaria Bogotá S.A. como vocera del patrimonio autónomo denominado Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación Francisco José de Caldas y la Universidad Nacional de Colombia", por su financiación en recursos humanos, y a la Corporación Autónoma Regional de Caldas (CORPOCALDAS) por el apoyo brindado y la información suministrada.



# Referencias Bibliográficas

- [Benavides, 2010] Benavides, H., A. L. (2010). Análisis descriptivo de variables meteorológicas que influyen en la calidad del aire de los principales centros industriales del país. nota técnica del ideam. *Nota Técnica del IDEAM. Bogotá, Colombia*.
- [Berthold, 2010] Berthold, M.R., B. C.-H. F. K.-F. (2010). Guide to intelligent data analysis. *Computer Science*, 42.
- [Carslaw, 2012] Carslaw, D. C., . B. S. D. (2012). Characterizing and understanding emission sources using bivariate polar plots and k-means clustering. *Environmental Modelling & Software*, 40:325–329.
- [Carslaw, 2011] Carslaw, D.C., R. K. (2011). Openair-an r package for air quality data analysis. *Environmental Modelling & Software*, 27–28:52–61.
- [Chaloulakou, 2003] Chaloulakou, A., K. P. S. N. D.-P. K. P. (2003). Measurements of pm10 and pm2.5 particle concentrations in athens, greece. *Atmospheric Environment*, 37:649–660.
- [CONPES, 2008] CONPES (2008). Lineamientos para la formulación de la política integral de salud ambiental con énfasis en los componentes de calidad de aire, calidad de agua y seguridad química en colombia. *Atmospheric Environment*.
- [CORPOCALDAS, 2008] CORPOCALDAS (2008). Lineamientos para la formulación de la política integral de salud ambiental con énfasis en los componentes de calidad de aire, calidad de agua y seguridad química en colombia. *Línea Base de Indicadores para Caldas*.
- [Cortés, 2010] Cortés, A. (2010). Análisis de la variabilidad espacial y temporal de la precipitación en una ciudad de media montaña andina. caso de estudio: Manizales. Master’s thesis, Universidad Nacional de Colombia Sede Manizales.
- [Cárdenas, 2010] Cárdenas, H. (2010). Consideraciones del material particulado en bogotá. alternativas tecnológicas de medición de la calidad del aire. *Tecnura*, 25:107–118.
- [de Assuncao, 2005] de Assuncao, J.V., P. C. B. R.-C. L. (2005). Dioxins and furans in the atmosphere of sao paulo city, brazil. *Chemosphere*, 58:1391—1398.
- [Ellison, 2013] Ellison, R., G. S. H. D. (2013). Five years of londons low emission zone: Effects on vehicle fleet composition and air quality. *Transportation Research Part D: Transport and Environment*, 23:25–33.
- [Elminir, 2005] Elminir, H. (2005). Dependence of urban air pollutants on meteorology. *Science of the Total Environment*, 350:225—237.
- [Environmental Protection Agency, 2009] Environmental Protection Agency, E. (2009). Air quality index. a guide to air quality and your health. office of air quality planning and standards.
- [ERG, 2013] ERG (2013, Consultado el 3 de junio de 2013). Open air project. *Consultado el 15 de septiembre de 2013*.

- [Follos, 2012] Follos, F. (2012). Lenguaje r aplicado al análisis de datos de calidad del aire, manual de uso de r y openair.
- [Gibert, 2008] Gibert, K., S. J. S.-M. M.-A. N. C. J. (2008). *Data Mining for Environmental Systems*, chapter 12, pages 205–228. Elsevier, Amsterdam.
- [González, 2012] González, C. (2012). Calidad del aire en la zona centro y oriente de la ciudad de manizales: Influencia del material particulado (pm10) y lluvia ácida. Master’s thesis, Universidad Nacional de Colombia Sede Manizales.
- [Harrison, 2012] Harrison, M., L. D. M.-S. L.-K. (2012). Processes affecting concentrations of fine particulate matter (pm2.5) in the uk atmosphere. *Atmospheric Environment*, 46:115–124.
- [Lee, 2011] Lee, S., H. C. C.-Y. (2011). High-pm10 concentration episodes in seoul, korea: Background sources and related meteorological conditions. *Atmospheric Environment*, 45:7240–7247.
- [Li, 2011] Li, X., L. Y. Z.-Q. W.-P. Y. H. J. G. W. F. (2011). Evaluation of atmospheric sources of pcdd-fs, pcbs and pbdes around a steel industrial complex in northeast china using passive air samplers. *Chemosphere*, 84:957–963.
- [Martínez, 2010] Martínez, K., R. A. J.-J. E.-A. M. R. J. A. E. (2010). Assessment of the emission of pcdd-fs and dioxin-like pcbs from an industrial area over a nearby town using a selective wind direction sampling device. *Environmental Pollution*, 158:764–769.
- [Mathworks, 2013] Mathworks (2013). Overview. *Consultado el 5 de Junio de 2013*.
- [Mead, 2013] Mead, M. I., P. O. A. M.-S. G. B. L. P. C. M. H. M. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203.
- [Ministerio de Ambiente, 2006a] Ministerio de Ambiente, Vivienda y Desarrollo Territorial, M. (2006a). Decreto 244 del 2006. república de colombia.
- [Ministerio de Ambiente, 2006b] Ministerio de Ambiente, Vivienda y Desarrollo Territorial, M. (2006b). Resolución 601, 2006. república de colombia.
- [Ministerio de Ambiente, 2008] Ministerio de Ambiente, Vivienda y Desarrollo Territorial, M. (2008). Resolución 910, 2008. república de colombia.
- [Ministerio de Ambiente, 2010a] Ministerio de Ambiente, Vivienda y Desarrollo Territorial, M. (2010a). Resolución 610, 2010. república de colombia.
- [Ministerio de Ambiente, 2010b] Ministerio de Ambiente, Vivienda y Desarrollo Territorial, M. (2010b). Resolución 650, 2010. república de colombia.
- [Ministerio del Medio Ambiente, 1995] Ministerio del Medio Ambiente, M. (1995). Decreto 948 de 1995. república de colombia.
- [Moragues, 2012] Moragues, J. (2012, Consultado 25 del Julio de 2012). Clasificación de estabilidad. capas de mezcla.
- [Muñoz, 2007] Muñoz, A., P. J. Q. C. (2007). Efectos de la contaminación atmosférica sobre la salud en adultos que laboran a diferentes niveles de exposición. universidad de antioquia, medellín. *Revista Nacional de Salud Pública*, 25:85–94.
- [Munir, 2013] Munir, S., C. H. R. K. (2013). Quantifying temporal trends in ground level ozone concentration in the uk. *Science of The Total Environment*, 458–460:217–227.
- [Observatorio, 2013] Observatorio, V. (2013). *Consultado el 7 de agosto de 2013*.
- [Organización Panamericana de la Salud, 2005] Organización Panamericana de la Salud, O. (2005). Curso de autoinstrucción. conceptos básicos sobre la meteorología de la contaminación del aire. *Consultado el 14 de Junio de 2013*.

- [Orozco, 2005] Orozco, M. (2005). Selección efectiva de características wavelet en la identificación de bioseñales 1-d y 2-d usando algoritmos genéticos. Master's thesis, Universidad Nacional de Colombia Sede Manizales.
- [Ozbay, 2011] Ozbay, B., K. G. D. S. A. S. (2011). Multivariate methods for ground-level ozone modeling. *Atmospheric Research*, 102:57–65.
- [Peña, 2002] Peña, D. (2002). Análisis de datos multivariantes. *Atmospheric Research*.
- [Pérez-Rial, 2010] Pérez-Rial, D., L.-M. P. T. R. (2010). Investigation of the source composition and temporal distribution of volatile organic compounds (vocs) in a suburban area of the northwest of Spain using chemometric methods. *Atmospheric Environment*, 44:5122–5132.
- [Pla, 1986] Pla, L. (1986). Análisis multivariado: Método de componentes principales. *Secretaría General de la Organización de los Estados Americanos Programa Regional de Desarrollo Científico y Tecnológico. Serie Matemática. Monografía N. 27*.
- [Quijano, 2010] Quijano, A., Q. M.-H. M. (2010). Caracterización fisicoquímica del material particulado fracción respirable pm2.5 en Pamplona-norte de Santander-Colombia. *Bistua: Revista de la Facultad de Ciencias Básicas*, pages 1–20.
- [Ramanathan, 2009] Ramanathan, V., F. Y. (2009). Air pollution, greenhouse gases and climate change: global and regional perspectives. *Atmospheric Environment*, 43:37–50.
- [Ren, 2011] Ren, Z., Z. B.-L. P. L. C. G. L. Z. M. (2011). Characteristics of air pollution by polychlorinated dibenzo-p-dioxins and dibenzofurans in the typical industrial areas of Tangshan city, China. *Journal of Environmental Sciences*, 23:228—235.
- [Rivera-Austrui, 2011] Rivera-Austrui, J., B.-M. M. K. A. M. A. M. V. B. B. R. J. A. E. (2011). Assessment of polychlorinated dibenzo-p-dioxin and dibenzofuran emissions from a hazardous waste incineration plant using long-term sampling equipment. *Chemosphere*, 82:1343–1349.
- [Santa Cruz, 2000] Santa Cruz, A. (2000). Difusión de contaminantes gaseosos en la atmósfera. *Universidad Tecnológica Nacional. Facultad Regional Rosario. Departamento de Ingeniería Química*.
- [Seinfeld, 2006] Seinfeld, J.H., P. S. (2006). *Atmospheric chemistry and physics: From air pollution to climate change*. John Wiley & Sons, Inc, Hoboken, New Jersey, 2 edition.
- [Shlens, 2009] Shlens, J. (2009). A tutorial on principal component analysis. *Systems Neurobiology Laboratory, Salk Institute for Biological Studies*.
- [Universidad de Santiago de Compostela, 2013] Universidad de Santiago de Compostela, U. (2013). Tutorial matlab.
- [van der Heijden, 2004] van der Heijden, F., D. R. T. D. d. R. D. (2004). *Introduction to PRTools*, chapter Apéndice E, pages 405–416. John Wiley & Sons Ltd, England.
- [Velasco, 2006] Velasco, M. (2006). Calidad del aire en la ciudad de Manizales. *Corporación Autónoma Regional de Caldas, CORPOCALDAS*.
- [von Schneidmesser, 2010] von Schneidmesser, E., M. P. S. P.-D. C. (2010). Global comparison of VOC and CO observations in urban areas. *Atmospheric Environment*, 44:5053–5064.
- [WHO, 2006] WHO (2006). WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulphur dioxide. summary of risk assessment.
- [Zoras, 2006] Zoras, S., T. A. D. D. (2006). Atmospheric stability and pm10 concentrations at far distance from elevated point sources in complex terrain: Worst-case episode study. *Journal of Environmental Management*, 80:295—302.



## Capítulo 4

# Análisis multidimensional de datos ambientales

*Néstor Darío Duque Mendez*<sup>1</sup>,

*Mauricio Orozco Alzate*<sup>2</sup>

*Jorge Julián Vélez Upegui*<sup>3</sup>

*1* Grupo de Ambientes Inteligentes Adaptativos - Departamento de Informática y Computación - Facultad de Administración

*2* Grupo de Procesamiento y Reconocimiento de Señales - Departamento de Informática y Computación - Facultad de Administración

*3* IDEA - Departamento de Ingeniería Civil - Facultad de Ingeniería Universidad Nacional de Colombia - Sede Manizales

*... Los datos en bruto sólo son la recopilación de hechos, el procesamiento y análisis permite obtener conocimiento y dan valor a lo almacenado. ...*

## Resumen

La gran cantidad de datos obtenidos desde las estaciones hidroclimatológicas se convierte en una gran oportunidad para el análisis y obtención de conocimiento. Pero el almacenamiento de datos eficiente y la efectiva manipulación es un requisito previo en la investigación meteorológica e hidroclimatológica y requiere de estrategias para su captura, transporte, almacenamiento y procesamiento, garantizando la calidad y consistencia de los datos.

Este capítulo presenta el desarrollo de un modelo en estrella para una bodega de datos (data warehouse) ambientales, utilizada para el almacenamiento estructurado y aprovechada para el análisis multidimensional de datos hidroclimatológicos históricos. Se presentan conceptos básicos relacionados con el modelo multidimensional de almacenamiento de datos y un enfoque de análisis de los mismos. Se expone la propuesta de un modelo en estrella para una bodega de datos ambientales y se aprovecha sus características para aplicar análisis dimensional a los datos almacenados, provenientes de diversas estaciones de la red de estaciones hidrometeorológicas telemétricas de la ciudad de Manizales, Colombia, que en algunos casos conserva medidas que sobrepasan los 60 años.

Se aplican técnicas OLAP (On-Line Analysis Processing) para descubrir relaciones ocultas que permiten describir comportamientos y pueden servir de soporte para predicciones en la región. El modelo presentado es una alternativa al almacenamiento y procesamiento de infor-

mación y las salidas obtenidas fueron utilizadas tanto para la evaluación de la calidad de los datos como para la visualización de relaciones entre variables diversas en casos específicos. Para estos fines se diseñó y construyó MADE, una aplicación desarrollada para el procesamiento y el análisis de datos ambientales y que puede ser usada en otros escenarios.

**Palabras Clave:** Datos medioambientales, minería de datos, predicción.

## 4.1. Introducción

El aprovechamiento para análisis y excavación de conocimiento del gran número de registros de las mediciones obtenidas en tiempo real y en intervalos de tiempo muy pequeños, de diferentes variables provenientes de las estaciones hidrológicas, es un tema que reviste gran interés actualmente. Este enfoque permite superar las tradicionales aplicaciones usadas, únicamente, para reflejar los valores actuales o pasados de variables físicas y presentar algunas páginas de información básica a la comunidad. Pero para poder obtener beneficios de este alto volumen de datos es preciso diseñar e implementar esquemas de almacenamiento apropiados y aplicar herramientas que obtengan información y descubran conocimiento oculto a través de relaciones y tendencias y en el mejor de los casos poder predecir comportamientos futuros.

El poder beneficiarse del gran número de registros de fenómenos físicos se ha cristalizado en una serie importante de trabajos, entre los cuales se pueden mencionar: En [Ma et al.(2010)] se propone el diseño e implementación de una bodega de datos (data warehouse) meteorológica y un esquema de reportes, basados en Microsoft SQL Server, orientado al análisis e investigación meteorológica apoyado en procesamiento analítico en línea (OLAP) y reportes multidimensionales, los cuales son publicados en Web. Con la visión de los almacenes de datos, [Shang et al.(2010)] analizan múltiples fuentes de datos meteorológicos, proponen el diseño e implementación basado en la plataforma Oracle-OWB, al igual que el modelo de datos de destino y el proceso ETL, reportando muy buenos resultados. [Tan(2006)] propone una bodega de datos para pronósticos climáticos y desarrollan 4 esquemas de análisis basados en análisis multidimensional, determinando las combinaciones válidas entre hechos y dimensiones. [Bartok et al.(2012)] describen la investigación de modelos parametrizados y los métodos para la detección y predicción de fenómenos meteorológicos significativos cubriendo los métodos para la integración de los datos meteorológicos distribuidos necesarios para el funcionamiento de los modelos de predicción, los modelos de formación y luego la extracción de datos con el fin de ser capaces de predecir de manera eficiente y rápida, incluso al azar, los fenómenos que ocurren. [Ansell y Gutierrez(2010)] presentan una nueva técnica de combinar fuentes de información histórica con estaciones modernas de una manera jerárquica, en un modelo de reducción de escala única, modeladas mediante redes bayesianas. [Cortez y Morais(2007)] exploran técnicas de minería de datos para predecir posibles incendios en áreas forestales en Portugal, a partir de datos meteorológicos. En [Dominguez et al.(2008)] se valora la importancia para la sociedad de conocer el estado del tiempo y ofrecen una herramienta que muestra en forma gráfica, medible y detallada los cambios de temperatura en un periodo de tiempo. Se aprovecha una bodega de datos con registros de temperatura, viento, fechas y horas, que permite analizar la información, mediante la herramienta Power Play Transformer. Por su parte en [Lemus et al.(2009)] se presenta un proceso KDD (Knowledge Discovery in Databases) utilizando tareas de selección de atributos y regresión para analizar dependencias entre parámetros meteorológicos y estimar secundarios. En trabajo previo [Duque, Orozco e Hincapié(2010)] aplican técnicas de minería de datos sobre datos históricos recopilados de manera sistemática por estaciones de la red de estaciones hidrometeorológicas telemétricas, obteniendo un primer acercamiento para establecer tendencias en el comportamiento de algunas variables.

La revisión realizada deja en claro que para implementar un sistema para análisis de datos meteorológicos se debe determinar, por un lado la definición de los esquemas de almacenamiento y por el otro las técnicas de procesamiento de los datos recolectados para obtener información y conocimiento oculto.

El trabajo presentado en este capítulo recoge parte de los resultados obtenidos en el marco del programa de investigación financiado por la Universidad Nacional de Colombia, orientado a la acciones de grupos interdisciplinarios para sistematizar y aplicar técnicas de análisis a los datos obtenidos en la red de estaciones hidrometeorológicas telemétricas en la ciudad de Manizales en Colombia. Se diseñó e implementó un modelo estrella para una bodega de datos y se aplicaron técnicas OLAP, buscando describir el comportamiento de algunas variables en el tiempo y obtener conocimiento mediante el análisis multidimensional. La propuesta fue implementada en un sistema que se denominó MADE, por las iniciales de *Multidimensional Analysis of Data Environmental* y que ofrece una interfaz amigable para la administración y explotación de los datos y la obtención de salidas de acuerdo a los intereses del analista ambiental. Estos resultados pueden servir de base para la toma de decisiones e inversiones en este tema, a la vez que apuntan a establecer las tendencias en el comportamiento de dichas variables y en lo posible a la predicción de las mismas.

## 4.2. Almacenamiento y análisis de datos hidroclimatológicos

Obtener conocimiento de los datos tomados de los fenómenos físicos, requiere la definición, diseño y construcción de diversos componentes, lo cual se traducirá en resultados en el funcionamiento y disponibilidad del sistema.

### 4.2.1. Diseño de la bodega de datos

Una bodega de datos (data warehouse) es un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de la administración (W.H. Inmon, considerado como el padre del data warehouse) [Harjinder y Prakash, 1996]. Su orientación principal es la de servir de soporte a la toma de decisiones, pero por su alto volumen de datos se convierte en la base para la aplicación de diferentes técnicas de análisis de datos con fines descriptivos o predictivos. Data warehouse (DWH) es un concepto relativamente nuevo, orientado al manejo de grandes volúmenes de datos, provenientes de diversas fuentes, de muy diversos tipos. Estos datos cubren largos períodos de tiempo, lo que trae consigo que se tengan diferentes esquemas de las bases de datos fuentes.

En este trabajo se explotaron los beneficios que se derivan del almacenamiento de datos hidroclimatológicos en un modelo multi-dimensional y la aplicación de técnicas OLAP, permitiendo la generación de salidas multidimensionales y así obtener información relevante para el análisis histórico, actual e incluso para predicciones meteorológicas. Para el diseño e implementación de cualquier esquema de almacenamiento se requiere el conocimiento y análisis de los datos a almacenar, pues de sus características dependerá el enfoque seleccionado, lo que a la vez es condición para el éxito del sistema.

### 4.2.2. Datos disponibles desde las estaciones hidroclimatológicas

Los datos obtenidos desde las estaciones de medición provienen de dos redes de telemetría de estaciones hidrométricas y meteorológicas, que transmiten datos a través de señales de radio en tiempo real, a un servidor que los almacena tal como se reciben.

La Tabla 4.1 muestra algunas estaciones con su tipo, ubicación y los periodos disponibles. Las medidas en estas estaciones, en su gran mayoría, se registran cada 5 minutos.

Tabla 4.1: Características de algunas Estaciones de la Red. Fuente: propia.

Estación	Tipo	Municipio	Altitud	InicioFunc
Río Rioclaro	Hidrometereológica	Villamaría	1807	02/2011
Emas	Metereológica	Manizales	2060	08/1997
Cumanday	Metereológica	Villamaría	5072	2006
Cenicafé	Metereológica	Chinchiná	1310	01/1942
La Francia	Precipitación	Chinchiná	1267	01/1977
Agronomía	Metereológica	Manizales	2088	01/1956
Santagueda	Metereológica	Palestina	1026	01/1964
...	...	...	...	...

Cómo se aprecia existe una gran variedad en los tipos de datos, los periodos de obtención y en las condiciones de las zonas donde están ubicadas las estaciones y esta diversidad puede ser explotada mediante la correcta aplicación de técnicas de análisis de datos.

### 4.2.3. Modelo multidimensional - bodegas de datos

Existen dos modos de almacenar los datos en data warehouse: Bases de datos relacional y bases de datos multi-dimensional. El almacenamiento de datos eficiente y la correcta manipulación son problemas y requisitos previos para el éxito en el ámbito meteorológico y la climatología [Dimri and Gunwant(2012)]. La tarea de OLAP se fundamenta en análisis multidimensional, donde una dimensión puede ser cualquier aspecto en la vista de datos y el propósito principal es explorar los datos para encontrar relaciones o patrones y no sólo ver su comportamiento aislado. El modelo multidimensional en las bodegas de datos es una técnica de diseño lógico que busca presentar los datos de una forma intuitiva y con alto desempeño. Cada modelo dimensional se compone de una tabla con múltiples llaves foráneas, llamada tabla de hechos (fact table) y un conjunto de tablas más pequeñas, llamadas tablas de dimensión (dimension table). Los atributos de las tablas de dimensión determinan las restricciones de búsqueda en las consulta de la bodega de datos. Existen dos enfoques en el diseño del Modelo Multidimensional en las Bodegas de Datos: Modelo en estrella (star schema) y modelo de copo de nieve (snowflake schema). En el modelo estrella se tiene una tabla de hechos y en ella llaves foráneas a cada una de las tablas de dimensión. Cada tabla de dimensión está directamente relacionada con la tabla de hechos. Una estructura simple en estrella tiene solo una tabla de hechos. Los hechos son medidas de las variables que se consideran y normalmente están asociados a valores numéricos que soportan cálculos. Las dimensiones o atributos de tipo texto que describen cosas, se usan para definir restricciones y los nombres de las columnas de los reportes. Al agregar restricciones a una búsqueda se está haciendo un *drilling down*, es decir se está logrando mayor nivel de detalle y si se hace en forma eficiente mezcla atributos de las diferentes dimensiones, para hacer reportes robustos. Las llaves en la bodega de datos deben ser subrogadas, es decir no significan nada respecto al sistema y no se usan las llaves de las fuentes de datos. La granularidad representa el nivel de detalle en las unidades de los datos de la bodega [Darmawikarta(2007)].

En el enfoque copo de nieve una dimensión con campos de baja cardinalidad son removidos a tablas separadas y unidas a la tabla original con llaves foráneas. La tabla de hechos no tendrá llaves foráneas a todas las demás tablas. Las nuevas tablas no estarán conectadas con la tabla de hechos sino con las dimensiones establecidas. Por su complejidad no es muy usado [Darmawikarta(2007)].

Después de evaluar la estructura de los datos recolectados y de las posibilidades ofrecidas por los dos enfoques, para la implementación de la bodega de datos ambiental, se ve más natural y simple el modelo en estrella. La Figura 4.1 muestra gráficamente el diseño de la bodega de datos propuesto y los componentes tanto de la única tabla de hechos, como de las dimensiones que intervienen.

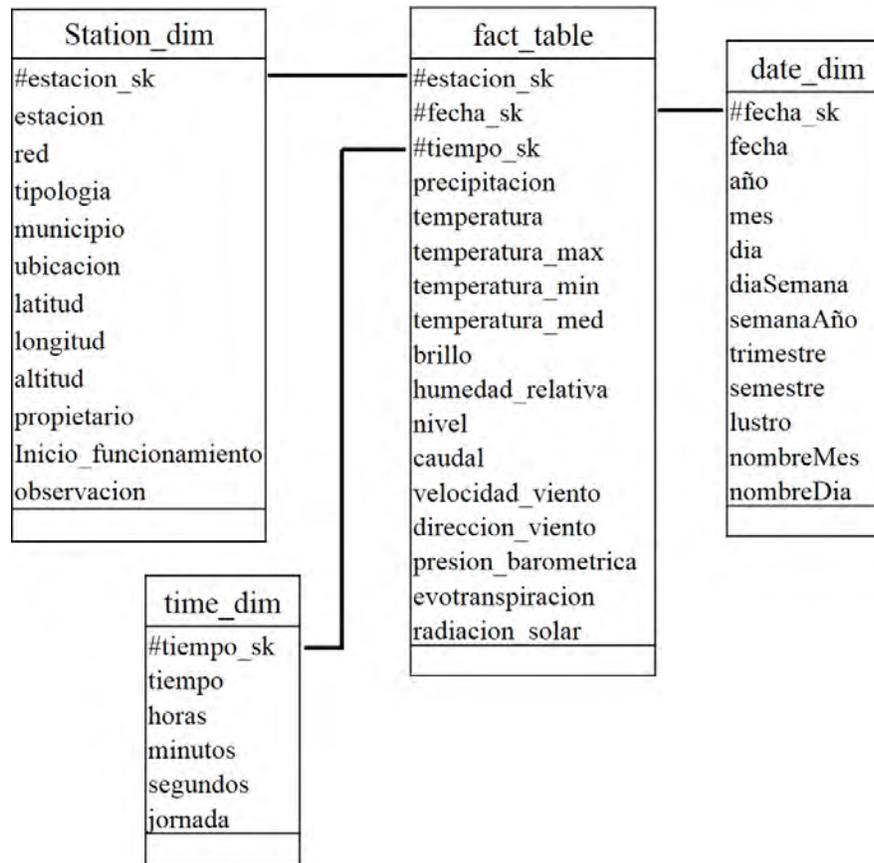


Figura 4.1: Diseño de la Bodega de Datos - Modelo Estrella. Fuente: propia.

El núcleo del modelo es la fact\_table, de la que dependen las dimensiones. En el caso de estudio, los atributos de la tabla de hechos son: estacion\_sk, fecha\_sk, tiempo, precipitación, temperaturas local, mínima, media y máxima, brillo solar, humedad relativa, caudal medio, nivel medio, velocidad y dirección del viento, presión barométrica, evotranspiración y radiación solar media. Las dimensiones básicas son estación, fecha y tiempo. Para el caso de fecha se tienen varios niveles de granularidad: año, mes, día, trimestre, semestre, lustro y década, lo que permite ampliar las vistas multidimensionales. Para la dimensión estación se almacena el identificador, ubicación (municipio y coordenadas) y se busca otros niveles de granularidad como latitud, región, etc.

### 4.3. Proceso KDD para estaciones hidroclimatológicas

El Descubrimiento de Conocimiento en Bases de Datos ,KDD, es un área de la computación que intenta explotar la ingente cantidad de información, extrayendo conocimiento oculto que pueda asistir a los humanos para llevar a cabo tareas de forma más eficiente y satisfactoria. Puede definirse KDD como el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos [Hernández et al.(2004)].

KDD un proceso que cubre diferentes fases, como puede verse en la Figura 4.2.

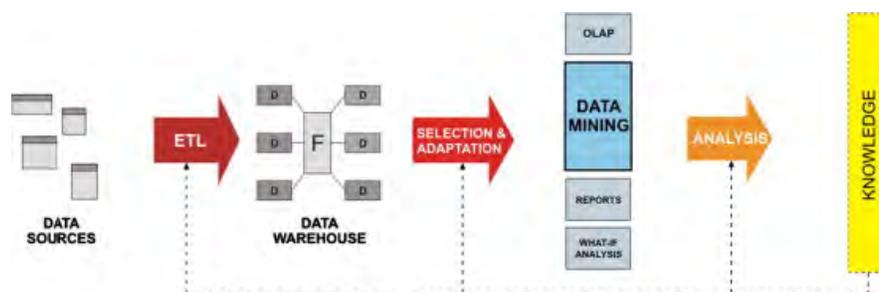


Figura 4.2: Fases de KDD con un modelo multidimensional de bodega de datos. [Zubcoff (2009)]

### 4.3.1. Proceso ETL

Además de definir la arquitectura de la bodega de datos y debido al volumen de los datos, a los diferentes esquemas en qué están almacenados y a los diferentes periodos a los que pertenecen, es usual la presencia de ruido, de datos inconsistentes o redundantes, por lo que se requiere aplicar mecanismos para el proceso previo de extracción, transformación y carga (ETL - Extract, transform, load) a los datos originales (operacionales) antes de ser llevados a la bodega de datos. Las actividades de ETL son particulares para cada conjunto de datos a tratar y tienen el objetivo de reforzar la calidad y consistencia del bloque de datos para que al aplicar técnicas OLAP se generen resultados que representen la realidad y entreguen vistas relevantes e implica adecuarlos a los formatos requeridos para el procesamiento y análisis.

Los datos fuente para este estudio presentaban inconvenientes en cuanto a la organización y formatos, pues al tratarse de estaciones con muchos años de funcionamiento, los datos se habían consignado en diferentes archivos, formatos de fechas diferentes, datos ausentes por falta de transmisión o valores por fuera de los permitidos o esperados.

Para algunos pasos de extracción y carga se empleó el software libre Talend Open Studio. En los datos en formularios inapropiados fue necesario realizar algunas actividades manuales para ubicar los valores según las variables reportadas. Para el caso de los faltantes estos valores fueron colocados en 'NULL' lo que garantiza que no se tengan en cuenta en las operaciones sumarias. Para el caso de valores inconsistentes, por ejemplos caudales negativos o temperaturas excesivas, en compañía de los expertos, se decidió si descartar los valores, tomar los valores promedios o deducirlos de las medidas cercanas. Para este fin se definieron una serie de filtros por cada estación y variable, lo que permitió automatizar el proceso.

Para obtener los valores de los campos de la dimensión fecha fue necesario generar una tabla que incluyera todas las fechas desde la más antigua hasta la más reciente y con los intervalos de los datos suministrados (cada 5 minutos). Para las columnas asociadas a diferentes niveles de granularidad, si no se cuenta con herramientas que lo hagan en forma automática, se pueden aplicar instrucciones SQL para generar los valores, por ejemplo: `update date_dim set nombreMes='Enero' where mes=1;` `update date_dim set nombreDia='Martes' where diaSemana=3;` `update date_dim set trimestre= 4 where mes between 10 and 12.`

El proceso de migrar los datos originales a la tabla de hechos (fact\_table) no es una tarea trivial y requiere que se mantengan las fechas respectivas y se actualicen los valores medidos según las dimensiones involucradas. Esto se logró utilizando la herramienta mencionada y en otros casos con consultas especializadas SQL.

Al final se obtuvo el modelo de la Figura 4.1 completamente poblado y con las posibilidades de aplicar las técnicas OLAP.

### 4.3.2. Herramientas OLAP para datos hidroclimatológicos

El concepto de procesamiento analítico en línea (OLAP) fue introducido desde 1970, pero sólo hasta mediados de la década de 1990 toma relevancia, aunque aún hoy muchas organizaciones no lo aprovechan. OLAP aplicado en bodegas de datos se ha convertido en la 'inteligencia' de la inteligencia de negocios. OLAP se presenta como una gran salida al diseño relacional tradicional y permite que los datos generen valor, se almacenen y se accedan de la manera más eficiente, permitiendo a los usuarios finales ver los datos desde los bordes de un 'cubo' hipotético de muchas dimensiones [Nanda(2013)]. OLAP permite a los usuarios acceder a la información de los almacenes de datos multidimensionales casi al instante, para ver la información de la manera que más les gusta y además realizar cálculos elevados [Thomsen(2002)]. Aunque ya están disponibles muchas herramientas y productos OLAP comerciales, OLAP es todavía una tecnología compleja y difícil de dominar.

### 4.4. Sistema *MADE* - Multidimensional Analysis of Data Environmental

Con el fin de cubrir los diferentes componentes de las fases requeridas para el análisis de los datos se diseñaron e implementaron diferentes módulos, con diferentes funcionalidades y tecnologías. Estos módulos se integran en un sistema que se denominó *MADE Multidimensional Analysis of Data Environmental* y que ofrece un punto de acceso centralizado y amigable para que los usuarios realicen, según su perfil, las diferentes tareas para el almacenamiento y procesamiento de los datos. La Figura 4.3 esquematiza el sistema *MADE*.

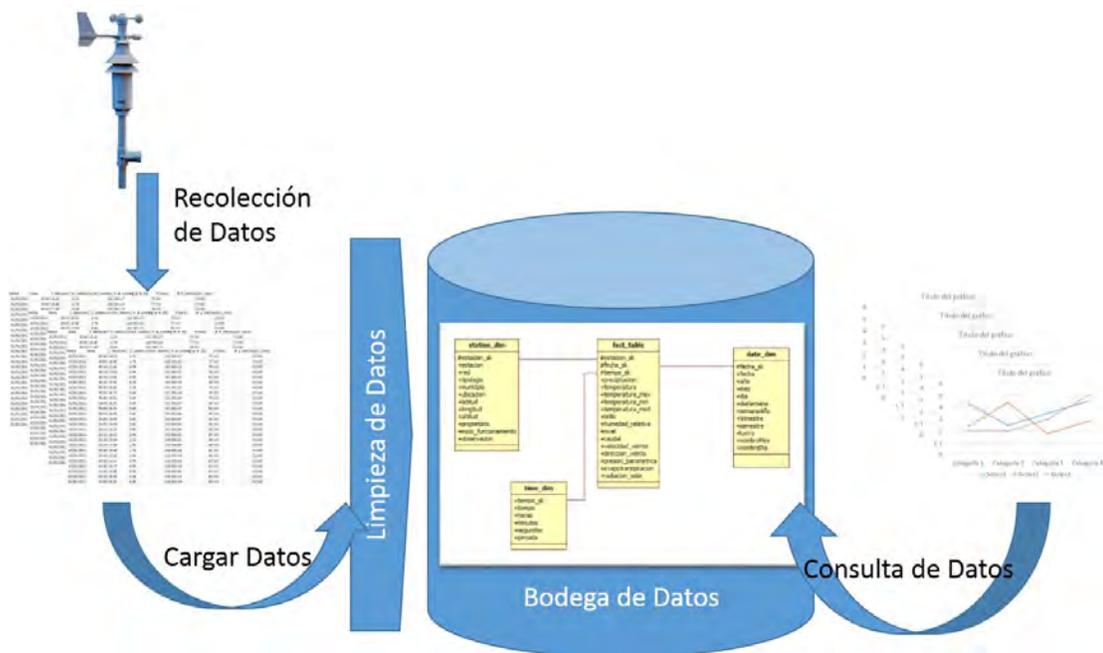


Figura 4.3: Sistema *MADE* - Multidimensional Analysis of Data Environmental. Fuente: propia.

Como se aprecia en la Figura el sistema incluye varios subsistemas que interactúan para lograr el objetivo planteado. A continuación se exponen los módulos del sistema *MADE*

#### 4.4.1. Bodega de datos Made-DWH

La bodega de datos fue construida en el Sistema Gestor de Bases de Datos **PostgreSQL** versión 9.1.3. Como ya se dijo para algunos casos del proceso de ETL se utilizó el software libre Talend Open Studio. Para el análisis inicial se aprovechó las posibilidades ofrecidas por SQL.

#### 4.4.2. Herramienta para OLAP - Made-OLAP

Las técnicas de OLAP se han utilizado ampliamente en las finanzas, ventas y marketing pero sus aplicaciones en los estudios científicos son relativamente recientes y la propuesta aquí presentada se puede ver, aún, como novedosa.

Para aplicar análisis multidimensional se desarrolló una aplicación en PHP, sobre sistema operativo Linux, que permite en forma interactiva definir medidas, dimensiones y periodos para obtener resultados desde la gran cantidad de datos almacenados. Se utilizó JQuery, JavaScript y las librerías de highcharts, escritas en HTML5/JavaScript puro.

La Figura 4.4, muestra la interfaz de consulta.



Figura 4.4: Interfaz de consulta en la herramienta MADE-OLAP. Fuente: propia.

Las posibilidades ofrecidas con la aplicación de OLAP son variadas, en particular los resultados obtenidos en el proyecto son valiosos tanto para evaluar la calidad de los datos como para las relaciones entre medidas físicas.

### 4.5. Análisis Multidimensional de datos hidroclimatológicos de Caldas

Recogiendo la propuesta de [Tan(2006)] y aprovechando el sistema MADE, se realizaron diferentes análisis multidimensionales a los datos de las estaciones hidroclimatológicas de Caldas. A continuación se presentan algunos ejemplos de los resultados obtenidos.

## ■ Relaciones.

Evaluar las relaciones de diferentes variables contando sólo con valores numéricos, tomados directamente de las estaciones, es una tarea complicada, dispendiosa y que no garantiza resultados certeros. Este mismo análisis aprovechando los procesos de agregación se hace más productivo y si además se acompaña con la representación grafica facilita mucho el entendimiento de las variables analizadas. Algunos ejemplos se muestran a continuación:

La Figura 4.5 presenta el comportamiento de las variables flujo medio y nivel medio para tres estaciones, con datos de hasta 50 años. En esta gráfica se debería apreciar la correlación existente entre el nivel y el caudal, y al no detectarse se puede inferir un posible problema en la curva de la calibración de la estación Montevideo, lo cual es un indicativo de la necesidad que los expertos revisen y corrijan esta situación. Algo similar sucede con las otras estaciones.

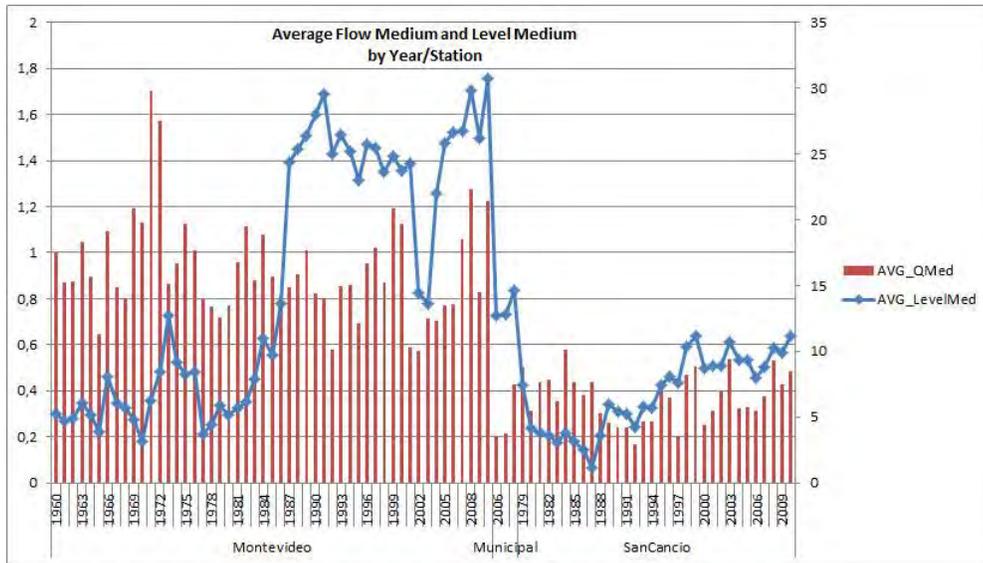


Figura 4.5: Relación entre flujo y nivel medio en tres estaciones. Fuente: propia.

Por otro lado, la Figura 4.6 permite apreciar las relaciones y tendencias de flujo y nivel en los 15 lustros anteriores para tres estaciones.

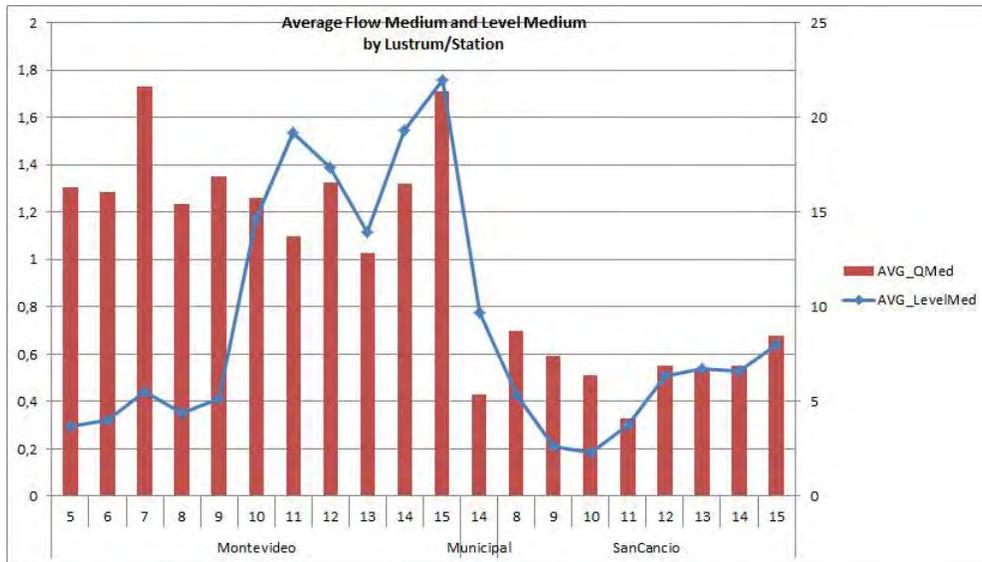


Figura 4.6: Relación entre Flujo y Nivel medio para 15 lustros. Fuente: propia.

■ **Análisis multi-escala.**

Las dimensiones fecha y tiempo en el modelo estrella desarrollado y coherente con la filosofía del enfoque multidimensional, permiten que se obtengan resultados en diferentes escalas de tiempo (niveles de granularidad) permitiendo análisis desde datos cincominutales hasta agregación por quinquenios. Todas estas ventajas están disponibles para los usuarios con solo hacer las selecciones respectivas.

La Figura 4.7 es un ejemplo de datos obtenidos para un día, con medidas cada 5 minutos para dos variables diferentes.

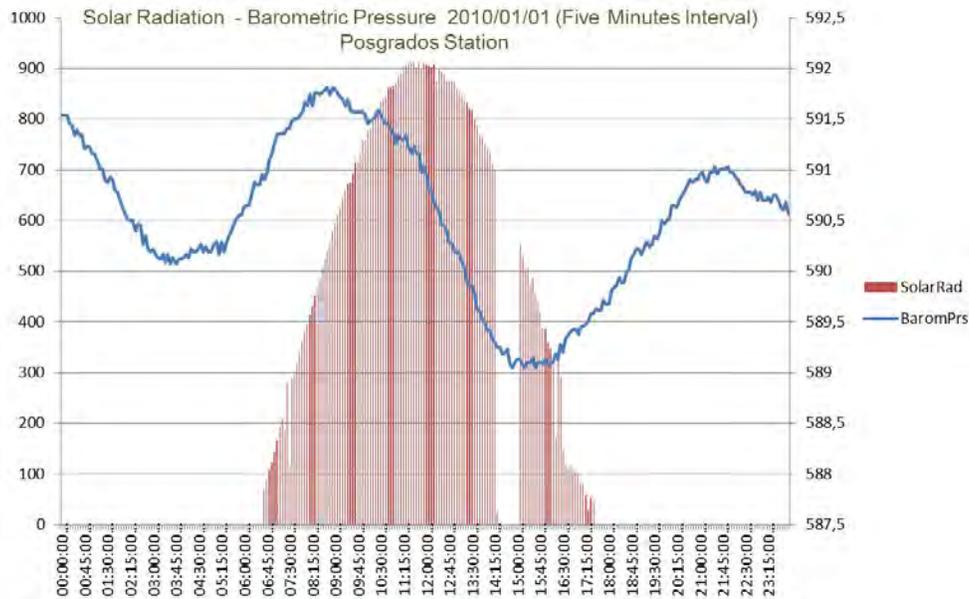


Figura 4.7: Radiación solar - Presión barométrica estación posgrados. Fecha: 01/01/2010 Intervalo: 5 Minutos. Fuente: propia.

Una funcionalidad que valoran los investigadores ambientales es la posibilidad de obtener valores acumulados a partir de datos instantáneos. En la Figura 4.8, se consigna la suma de precipitación en dos años para un mes específico en diferentes estaciones. Esta gráfica permite comparar el comportamiento de las lluvias en diferentes periodos.



Figura 4.8: Lluvia acumulada en año/mes para diferentes estaciones. Fuente: propia.

Para detectar tendencias o cambios en comportamientos el análisis con intervalos de tiempo mayor puede ser más representativo. Un ejemplo de ello es la información mostrada en la Figura 4.9 donde se aprecia el incremento de las lluvias en los años 1990-2001, causados por el fenómeno de *La Niña*.

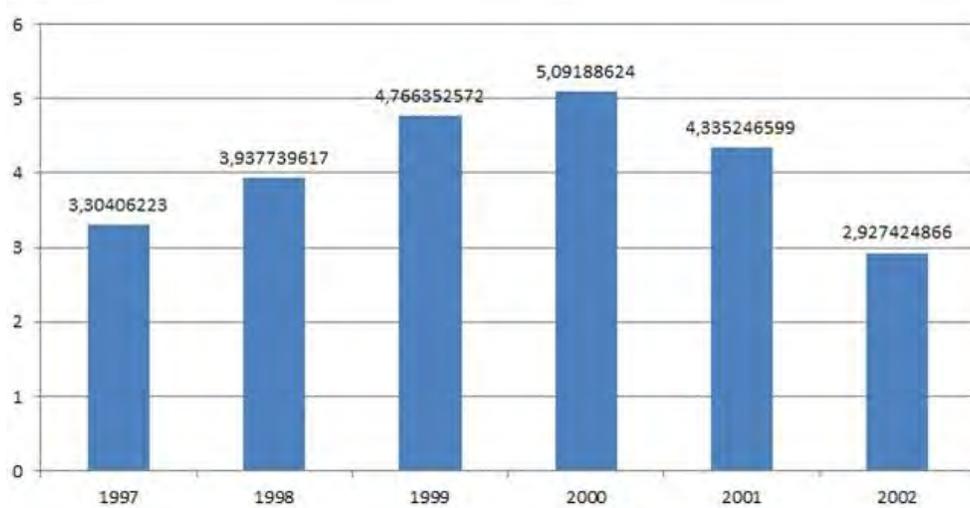


Figura 4.9: Promedio de lluvias en el periodo de los fenómenos de *El niño* y *La niña*. Fuente: propia.

Gráficamente se puede ver el comportamiento de todas las estaciones en un periodo de tiempo, como se presenta en la Figura 4.10, para el periodo comprendido entre 1956 y 1961, con lluvias acumuladas por año.

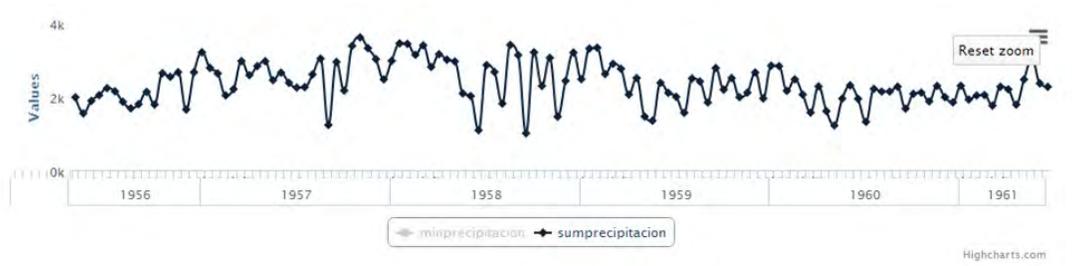


Figura 4.10: Lluvias acumuladas para el periodo 1956-1961. Fuente: propia.

■ **Versatilidad en el análisis.**

Como valor agregado, la versatilidad en la selección de variables, medidas, períodos y detalle para el análisis, dota de amplias posibilidades a los usuarios e investigadores, permitiendo que con pocas acciones puedan cambiar las variables, los tiempos, las estaciones, las operaciones de agregación de las medidas y la forma de presentación de los datos y así obtener resultados en forma numérica y gráfica de acuerdo a sus necesidades e intereses.

La Figura 4.11 es una mezcla de resultados obtenidos con sólo cambiar las selecciones de los parámetros y demuestra la versatilidad el modelo.

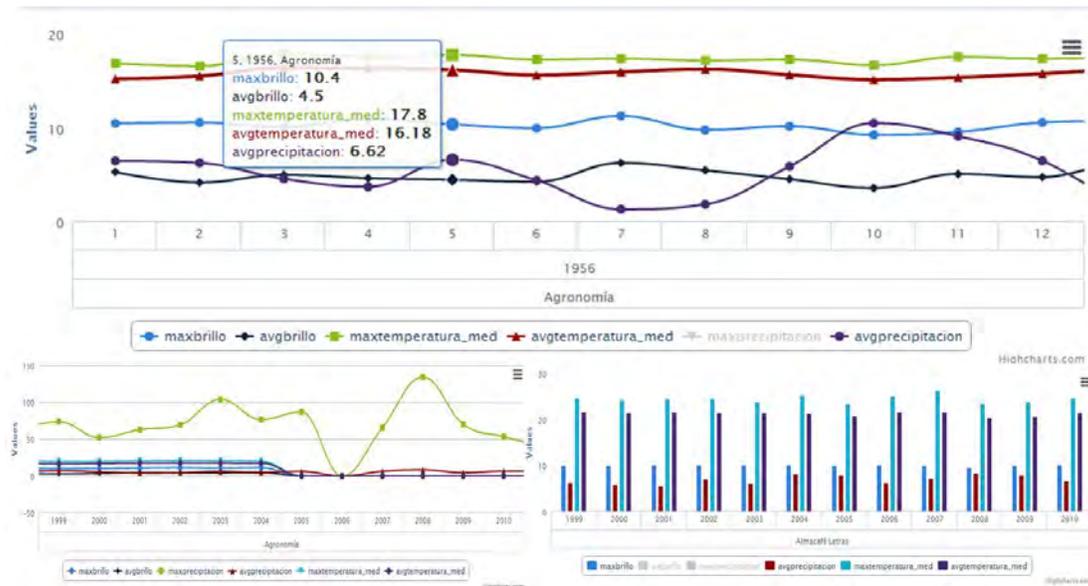


Figura 4.11: Diferentes escalas y variables para análisis. Fuente: propia.

■ **Tendencias.**

La Figura 4.12 registra mensualmente el promedio de la temperatura y de la precipitación media en la estación Cenicafé en los últimos 30 años y se aprecia el comportamiento con una leve tendencia al ascenso en la temperatura.

La Figura 4.13 representa el climograma (promedio de la precipitación y la temperatura media por año) de la estación Java por 11 años, donde el comportamiento es relativamente estable.

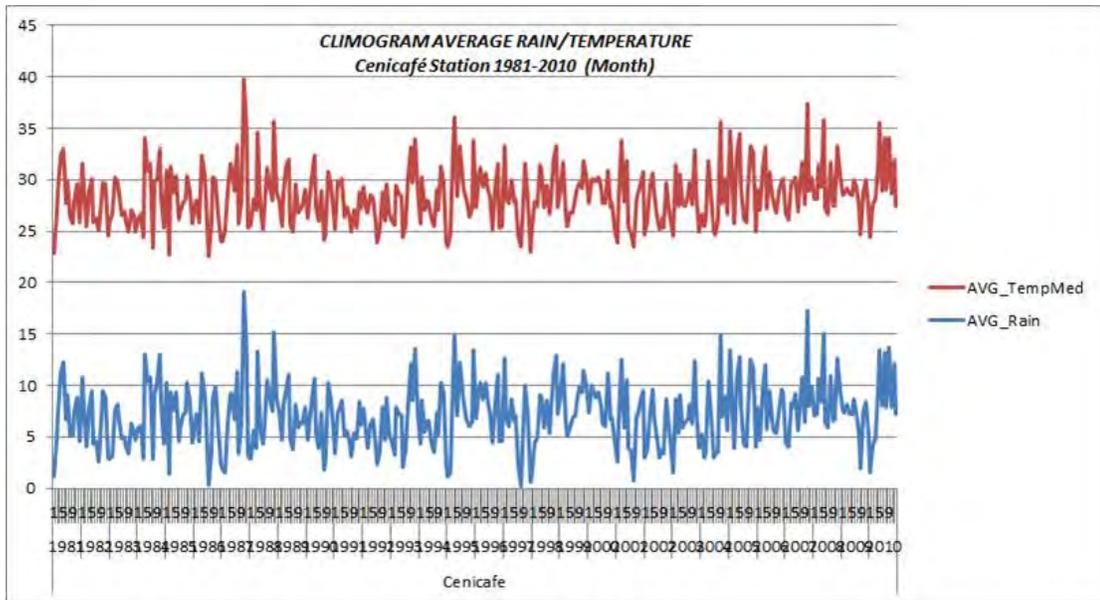


Figura 4.12: Tendencia para la estación Cenicafé. Fuente: propia.



Figura 4.13: Climograma para la estación Java 1999-2010. Fuente: propia.

■ **Validación de la calidad de los datos almacenados.**

Por diferentes razones, desde fallas en los instrumentos, errores en la transmisión o en el almacenamiento, el registro de las medidas físicas son susceptibles de errores e inclusive en muchos casos ser inexistentes. La presentación gráfica permite detectar en forma rápida estas situaciones, que pueden introducir errores en el análisis posterior.

Aunque OLAP no tiene como objetivo fundamental este tipo de salidas, las herramientas permiten la obtención de resultados de diferentes medidas y dimensiones con el propósito de evaluar la calidad de los datos. A continuación se presentan algunos ejemplos de este enfoque.

La Figura 4.14 permite apreciar que hay un periodo de tiempo donde de los valores del brillo son cero (0), por lo tanto se puede suponer que son datos ausentes para el caso de la estación Agronomía.



Figura 4.14: Datos faltantes en la medida brillo para la estación Agronomía en un período de tiempo. Fuente: propia.

Igual situación se aprecia en el climograma de la estación Santágueda, Figura 4.15, donde existe un fuerte decremento de la temperatura media en el año 2008, lo que requiere una revisión para determinar si son datos erróneos o un cambio producto de algún fenómeno climático.

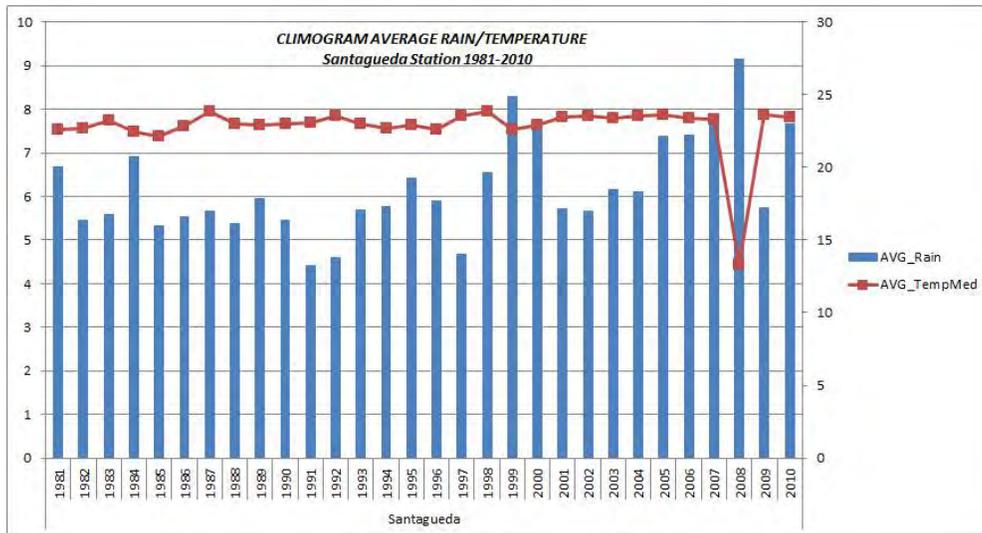


Figura 4.15: Climograma de la estación Santagueda. Fuente: propia.

Es posible con una rápida mirada detectar periodos de tiempo donde no existen registros de las medidas de algunas variables, lo cual sin significar un problema, porque puede ser debido a situaciones justificadas (equipos que no registran estas medidas, etc.), en ocasiones puede ser una alarma de problemas de adquisición, transmisión o almacenamiento. La Figura 4.16, que expone las medidas en la Estación La Argentina, permite ver claramente que no hay datos de la humedad relativa antes del año 2005, a partir del cual aparecen los registros.



Figura 4.16: Medidas almacenadas de la estación La Argentina - Ausencia de datos. Fuente: propia.

Esta funcionalidad en MADE-Olap es una buena ayuda en la percepción de la calidad de los datos y debe ser complementada con las pruebas estadísticas de rigor que demuestren las correlaciones, los cambios de media, las tendencias y la consistencia de los datos.

Esta sección demuestra las amplias posibilidades que el modelo implementado abre para el aprovechamiento del volumen de datos almacenados en lo referente a calidad de los datos, dinámica de algunas variables a nivel del departamento, por regiones o por estaciones; e incluso puede ser la base para realizar algunas predicciones. MADE proporciona una interfaz sencilla para que los investigadores interactúen con los datos y obtengan resultados en forma rápida y adecuada a sus intereses. Como un valor adicional permite determinar la consistencia de los datos y apoya la limpieza y mejoramiento si se requiere.

## 4.6. Conclusiones y trabajos futuros

Sistemas que registran datos que representan instantáneas de fenómenos, en nuestro caso físicos, generan volúmenes de datos cada vez crecientes, que deben ser gestionados tanto en su almacenamiento como en su procesamiento. Los datos almacenados *per se* no garantizan obtener conocimiento, pero para que puedan ser explotados el diseño del esquema de almacenamiento es fundamental y debe estar acorde con el enfoque del análisis a realizar.

El resultado presentado en este capítulo es una bodega multidimensional en estrella cuyo diseño potenció favorablemente la aplicación de técnicas OLAP, con ventajas para procesos de evaluación de calidad de los datos, el análisis multiescala temporal y para obtener relaciones entre medidas y un primer acercamiento a las tendencias subyacentes. El software MADE desarrollado permite al usuario realizar valiosas tareas en forma intuitiva y versátil. La automatización de los procesos de carga de datos provenientes de las estaciones hacia la bodega de datos, mediante filtros, es un factor fundamental para lograr la constante actualización de la bodega de datos.

Como trabajos en marcha se están incorporando nuevas estaciones y se ampliara el modelo para recibir otras medidas y si es requerido, definir otras escalas de dimensiones. Además de implementar otras técnicas de minería de datos que permitan clasificación, clustering y reconocimiento de los patrones en los datos almacenados.

## **Agradecimientos**

A la Convocatoria Nacional de Investigación y de Creación Artística de la Universidad Nacional de Colombia 2010 - 2012, en la modalidad fortalecimiento a grupos de investigación, por el apoyo al Programa de Fortalecimiento de Capacidades Conjuntas para el Procesamiento y Análisis de Información Ambiental, con código 12677.

# Referencias Bibliográficas

- [Anzell y Gutierrez(2010)] Anzell, R. y Gutierrez, J.M.(2010). Downscaling dense rain gauge networks with short historical records using Bayesian networks. *Journal of Hidrology*.
- [ASCE Task Committee(2000)] ASCE Task Committee.(2000). Artificial neural networks in hydrology - I: preliminary concepts. *Journal of Hydrologic Engineering. ASCE*, 5. pp. 115-123.
- [Bartok et al.(2012)] Bartok, J., Habala, O., Bednar, P., Gazak, M., and Hluchý, L. (2012). Data Mining and Integration for Predicting Significant Meteorological Phenomena., *Procedia Computer Science* pp. 37-46.
- [Beran et al.(2012)] Beran, B., Van Ingen, C., Zaslavsky I. and Valentine, D. (2012). OLAP Cube Visualization of Environmental Data Catalogs. *Microsoft Technical Report MSR-TR-2008-70* <http://research.microsoft.com/pubs/70582/OLAPreport.pdf>.
- [Carlson et al.(1970)] Carlson R.F., MacCormick, A.J.A., and Watts, D.G.(1970). Application of linear random models to four annual streamflow series. *Water Resources Research*, 6 (4). pp. 1070-1078.
- [Cortez y Morais(2007)] Cortez, Paulo and Morais, Aníbal (2007), A data mining approach to predict forest fires using meteorological data, *New trends in artificial intelligence : proceedings of the 13th Portuguese Conference on Artificial Intelligence (EPIA 2007)*.
- [Chen et al.(2010)] Chen, D.Q, Wang W.Y. and Yang H.K. (2010). Application Research on Data Warehouse of Hydrological Data Comprehensive Analysis. *IEEE* . pp. 1070-1078.
- [Darmawikarta(2007)] Darmawikarta, Djoni. (2007), Dimensional Data Warehousing with MySQL: A Tutorial, *BrainySoftware*, 448 pages.
- [Dibike et al.(2001)] Dibike, Y.B. Velickov, S. Solomatine, D. and Abbott, M.B. (2001). Model induction with support vector machines: introduction and applications. *Journal of Computing in Civil Engineering*, 15 (3), pp. 208-216.
- [Dimri and Gunwant(2012)] Dimri, P., Gunwant, H. (2012), Conceptual Model For Developing Meteorological Data Warehouse In Utta -Rakhand- A Review., *Journal of Information and Operations management* 107-110.
- [Dominguez et al.(2008)] Domínguez Lugo, Alma Jovita, Torres Ovalle, Sócrates, Alba Cisneros, Dalia Margarita, Silva Ávila, Alicia E.(2008), Medición y Análisis de Datos Meteorológicos, utilizando Bodega de Datos, *Simposio de Metrología*,.
- [Duque, Orozco e Hincapié(2010)] Duque M., N.D., Orozco A., M., Hincapié, L. (2010), Minería de Datos para el Análisis de Datos Meteorológicos, *Tendencias en Ingeniería de Software e Inteligencia Artificial, volumen 3*.
- [Han et al.(2001)] Han E., Merwade V.and Heathman G.C. (2001). Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model. *Journal of Hydrology. Vol. 416-417, pp. 98-117*

- [Harjinder y Prakash, 1996] Harjinder, S. Gill. Prakash, C. Rao. Data Warehousing. La Integración de Información para la Mejor Toma de Decisiones. Prentice Hall. México, 1996.
- [Hernández et al.(2004)] Hernández Orallo, José, Ramírez Quintana, María José y Ramírez Ferrí, César. (2004), Introducción a la Minería de Datos. *Pearson, Prentice Hall, Madrid*.
- [Jang(1993)] Jang. J. S. R. (1993). ANFIS: adaptive-network-based fuzzy inference systems. *IEEE Transactions on Systems, Man and Cybernetics*, 23 (3), pp. 665-685.
- [Lemus et al.(2009)] Lemus Gil, C., Rosete Suárez, A., Turtós Carbonel, L., Zerquera Diaz, R., Morales Montejó, A.,(2009), Estimación de parámetros meteorológicos secundarios aplicando Minería de Datos. *Instituto Cujae. Cuba*.
- [Ma et al.(2010)] Ma, N., Yuan, M., Bao, Y., Jin, Z., and Zhou, H. (2010), The Design of Meteorological Data Warehouse and Multidimensional Data Report, *Second International Conference on Information Technology and Computer Science 280-283*.
- [Mejía y Botero(2009)] Mejía F. y Botero B. A. (2009), Monitoreo Hidrometeorológico de los glaciares del Parque Nacional Natural Los Nevados. *Glaciares, Nieves y Hielos de América Latina. Cambio Climático y Amenazas. Colección Glaciares, Nevados y Medio Ambiente. Editores: C.D López y Ramírez J. Instituto Colombiano de Geología y Minería, Bogotá*.
- [Nanda(2013)] Nanda, Arup. Oracle Database 11g: The Top Features for DBAs and Developers.. *Sitio Oficial de Oracle. Disponible en <http://www.oracle.com/technetwork/articles/sql/11g-dw-olap-100058.html>. Consultado 6 agosto 2013*.
- [Ravindrana et al.(2010)] Ravindrana, N. Liang, Y. and Liang, X. (2010). A labeled-tree approach to semantic and structural data interoperability applied in hydrology domain. *Information Sciences*. 180, (24), pp. 5008-5028.
- [Shang et al.(2010)] Shang, Xinna ; Yuan, Mei ; Youwen Bao ; Xiaowei Liu. (2010). OWB-based construction of meteorological data warehouse. *International Conference on Computer Application and System Modeling (ICCASM), volume 8*.
- [Tan(2006)] Tan, Xiaoguang.(2006), Data Warehousing and its Potential Using in Weather Forecast, *Proc. 22nd Int. Conf. on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology, Atlanta, GA*.
- [Teegavarapu et al.(2009)] Teegavarapu R. , Tufail, M., and Ormsbee L. (2009). Optimal functional forms for estimation of missing precipitation data. *Journal of Hydrology*, 374, pp. 106-115.
- [Thomsen(2002)] Thomsen,Erik. (2002). OLAP Solutions: Building Multidimensional Information Systems. *Wiley. 2a ed.*
- [Wang(1991)] Wang, Q.J. (1991). The genetic algorithm and its application to calibrating conceptual rainfall-runoff models. *Water Resources research*, 27 (9), pp. 2467-2471.
- [Wang et al.(2009)] Wang W.C., Chau,K.W., Cheng, C.T. and Qiu L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374, pp. 294-306.
- [Whigam and Crapper(2001)] Whigam, P.A. and Crapper P.F. (2001). Modelling rainfall-runoff relationships using genetic programming. *Mathematical and Computer Modelling*, 33, pp. 707-721.
- [Zaslavsky et al.(2007)] Zaslavsky, I.; Rodriguez, M.; Beran, B.; Valentine, D.; van Ingen, C.; Wallis, J. C. (2007). OLAP Cube Visualization of Hydrologic Data Catalogs. *Proceedings in American Geophysical Union, AGU Fall Meeting*.
- [Zubcoff(2009)] Zubcoff, J., Pardillo, J., and Trujillo, J. (2009), A UML profile for the conceptual modelling of data mining with time series in data warehouses. *Information and Software Technology pp. 977-992*.

Entendimiento de fenómenos ambientales mediante el análisis de datos  
Impreso en la Sección de Publicaciones e Imagen  
de la Universidad Nacional de Colombia, Sede Manizales, febrero de 2015.  
Tamaño: 170 x 240 mm.  
Tiraje: 150 ejemplares.  
Fuentes tipográficas: Times y Courier new  
Páginas interiores Bond 75 gramos.  
Carátula Propalcote 300 gramos.





9 789587 752823