

Información Importante

La Universidad de La Sabana informa que el(los) autor(es) ha(n) autorizado a usuarios internos y externos de la institución a consultar el contenido de este documento a través del Catálogo en línea de la Biblioteca y el Repositorio Institucional en la página Web de la Biblioteca, así como en las redes de información del país y del exterior con las cuales tenga convenio la Universidad de La Sabana.

Se permite la consulta a los usuarios interesados en el contenido de este documento para todos los usos que tengan finalidad académica, nunca para usos comerciales, siempre y cuando mediante la correspondiente cita bibliográfica se le de crédito al documento y a su autor.

De conformidad con lo establecido en el artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, La Universidad de La Sabana informa que los derechos sobre los documentos son propiedad de los autores y tienen sobre su obra, entre otros, los derechos morales a que hacen referencia los mencionados artículos.

BIBLIOTECA OCTAVIO ARIZMENDI POSADA
UNIVERSIDAD DE LA SABANA
Chía - Cundinamarca

**EVALUACIÓN DE REDES PLUVIOMÉTRICAS APLICANDO ENTROPÍA Y
REDES NEURONALES ARTIFICIALES. CASO DE ESTUDIO: REGIÓN
BOGOTÁ - CUNDINAMARCA**



**Universidad
de La Sabana**

AUGUSTO RAFAEL GARRIDO ARÉVALO

**UNIVERSIDAD DE LA SABANA
FACULTAD DE INGENIERÍA
MAESTRÍA EN DISEÑO Y GESTIÓN DE PROCESOS
CHÍA, CUNDINAMARCA
2014**

**EVALUACIÓN DE REDES PLUVIOMÉTRICAS APLICANDO ENTROPÍA Y
REDES NEURONALES ARTIFICIALES. CASO DE ESTUDIO: REGIÓN
BOGOTÁ - CUNDINAMARCA**

AUGUSTO RAFAEL GARRIDO ARÉVALO

Proyecto de Grado

Directores:

**PhD. Nelson Obregón Neira
MSc. Mauricio Agudelo Otálora**

**UNIVERSIDAD DE LA SABANA
FACULTAD DE INGENIERÍA
MAESTRÍA EN DISEÑO Y GESTIÓN DE PROCESOS
CHÍA, CUNDINAMARCA
2014**

Nota de aceptación:

Firma del Director del Proyecto

Firma del Jurado

Firma del jurado

Chía, Mayo de 2014

A Dios – mi Todo –
A mi familia– mi apoyo –

AGRADECIMIENTOS

Debo agradecer de manera muy especial al ingeniero Nelson Obregón Neira, quien ha sabido orientarme en el desarrollo de este proyecto, brindándome su ayuda de manera desinteresada. Él me ha enseñado, además, que la ciencia y la investigación tienen un lado humano y que ellas son un buen espacio para fortalecer la amistad y ejercitar las virtudes asociadas a ésta.

Agradezco a la Universidad de La Sabana, por la oportunidad de participar en el programa Asistente Graduado, durante tres semestres. De manera particular, agradezco al ingeniero Mauricio Agudelo por su colaboración.

Mención especial merecen los ingenieros Federico González, Ana Carolina Santos y Miguel Ángel Vanegas, quienes me colaboraron en distintas fases del proyecto con el desarrollo de los códigos de Matlab y recomendaciones para el tratamiento de la información, compartiendo sus conocimientos y brindando una ayuda oportuna.

Quiero agradecer a mis compañeros de clase: Alain Paez, Martín Díaz, Andrés Álvarez, Camilo García, Lady Correa, Lina Bohorquez y Yulieth Montaña, por su apoyo en este proceso.

Son muchos los amigos que, directa e indirectamente, han colaborado para llevar a feliz término el proyecto, entre ellos: Felipe Ángulo, Luis Álvarez, Nelson Berdugo, Nazly Cottiz, Daisy Saldarriaga, Mary Elajaik, Christian Castellón, Cristian Vega, Felipe Bustamente, Carlos Llano, Juan Zuleta, Gustavo López, Daniel Aristizabal, Luis Millán, Carolina Pérez y un largo etcétera.

Tabla de contenido

Introducción	1
Capítulo 1	2
1.1 Planteamiento del Problema	2
1.2 Pregunta de Investigación	2
1.3 Objetivos	2
1.3.1 Objetivo General	2
1.3.2 Objetivos Específicos	3
1.4 Alcance y Relevancia del Proyecto	3
1.5 Estado del Arte	3
1.6 Marco Teórico	7
1.6.1 La precipitación	7
1.6.2 Medida de la precipitación	8
1.6.3 Redes neuronales artificiales	14
1.6.4 Clustering: métodos y aplicaciones	15
1.6.5 Redes de Kohonen	17
1.6.5 Arquitectura de las redes de Kohonen	19
1.6.6 La Entropía en la teoría de la Información	20
Capítulo 2	24
2.1 Localización de la zona de estudio	24
2.2 Aspectos Ambientales	24
2.2.1 Cuenca Río Sumapaz	26
2.2.2 Cuenca Río Bogotá	26
2.2.3 Cuenca de la vertiente oriental del Río Magdalena	27
2.2.4 Cuenca Río Negro	27
2.2.5 Cuenca Río Minero	28
2.2.6 Cuenca Ríos Ubaté y Suárez	28
2.2.7 Cuenca Río Blanco	28
2.2.8 Cuenca Río Machetá	29
2.2.9 Cuenca Río Gachetá	29
2.3 Red Hidroclimática	30



2.4 Selección de las estaciones para el estudio	33
Capítulo 3	38
3.1 Fase 1: Clasificación las estaciones pluviométricas de la región de estudio aplicando redes neuronales artificiales.....	39
3.1.1 Recolección de la información	39
3.1.2 Tratamiento de la Información	39
3.1.3 Clustering	39
3.2 Fase 2: Evaluación del desempeño de la red pluviométrica de la región Bogotá-Cundinamarca	43
3.2.1 Entropía de la información.....	43
3.2.1 Recomendaciones para el rediseño de la red	45
Capítulo 4	46
4.1 Variables de entrada	46
4.2 Proceso de clasificación	47
4.2.1 Tipología No. 1	47
4.2.2 Tipología No. 2	51
4.2.3 Tipología No. 3	57
4.2.3 Análisis de la clasificación	62
Capítulo 5	64
5.1 Datos de entrada.....	64
5.2 Cálculo de la Entropía y de la Transinformación.....	65
5.2.1 Estación E163 con E77.....	65
5.3 Resultados por grupos.....	71
5.3.1 Grupo No. 1	71
5.3.2 Grupo No. 2	73
5.3.2 Grupo No. 3	73
5.3.4 Grupo No. 4	73
5.3.5 Grupo No. 5	73
5.3.6 Grupo No. 6	73
5.3.7 Grupo No. 7	74
5.3.8 Grupo No. 8	74



5.3.9 Grupo No. 9	74
5.3.10 Grupo No. 10	74
5.3.12 Grupo No. 12	75
5.3.13 Grupo No. 13	75
Conclusiones y Recomendaciones	76
Trabajos citados	79

LISTADO DE TABLAS

Tabla 1. Densidad mínima recomendada de estaciones pluviométricas (OMM citado por (OEA, 2010))	13
Tabla 2. Clasificación de las principales medidas de distancia (Vidal Díaz, 1998)	17
Tabla 3. Cuencas hidrográficas de la región de estudio. (CAR, 2011)	26
Tabla 4. Clasificación de las estaciones en la zona de estudio. Fuente: SILCA	32
Tabla 5. Estaciones seleccionadas para el estudio.....	37
Tabla 6. Transformación de las variables (Gwo-Fong Lin, 2005)	40
Tabla 7. Características de las tipologías.....	40
Tabla 8. Clasificación de la Transinformación (Mishra & Coulibaly, 2009)	45
Tabla 9. Variables de entrada normalizadas	46
Tabla 10. Matriz de Hits	49
Tabla 11. Distribución de estaciones por grupos para la Tipología No. 1	50
Tabla 12. Matriz de Hits de la Tipología No. 2	54
Tabla 13. Distribución de estaciones por grupos para la Tipología No. 2	55
Tabla 14. Grupos finales para la Tipología No.2.....	56
Tabla 15. Matriz de Hits de la Tipología No. 3	59
Tabla 16. Distribución de estaciones por grupos para la Tipología No. 3	60
Tabla 17. Grupos finales para la Tipología No.3.....	61
Tabla 18. Resumen de la clasificación por tipología	63
Tabla 19. Combinación de estaciones comunes en las tipologías	63
Tabla 20. Estaciones pluviométricas del Grupo No. 9.....	64
Tabla 21. Precipitación mensual – Grupo No. 9.....	65
Tabla 22. Transinformación Serie 1 (E163)	71
Tabla 23. Transinformación Serie 2 (E77)	71
Tabla 24. Transinformación para las estaciones del Grupo No. 1.....	72
Tabla 25. Transinformación para las estaciones del Grupo No. 2.....	73
Tabla 26. Transinformación para las estaciones del Grupo No. 4.....	73
Tabla 27. Transinformación para las estaciones del Grupo No. 5.....	73
Tabla 28. Transinformación para las estaciones del Grupo No. 7.....	74
Tabla 29. Transinformación para las estaciones del Grupo No. 8.....	74
Tabla 30. Transinformación para las estaciones del Grupo No. 9.....	74
Tabla 31. Transinformación para las estaciones del Grupo No. 10.....	75
Tabla 32. Transinformación para las estaciones del Grupo No. 12.....	75
Tabla 33. Transinformación para las estaciones del Grupo No. 13.....	75

LISTADO DE FIGURAS

Figura 1. Registro pluviográfico típico (Shengwei Wang, 2011).....	9
Figura 2. Algunos modelos de pluviómetros (OEA, 2010).....	10
Figura 3. Red Pluviométrica de Colombia (IDEAM, 2002).....	11
Figura 4. Modelo de una neurona artificial (Yegnanarayana, 2006).....	15
Figura 5. Arquitectura de las redes de Kohonen (González, 2012).....	20
Figura 6. Ejemplos de Topologías de redes de Kohonen (González, 2012)	20
Figura 7. Área de estudio (CAR, 2012)	25
Figura 8. Cuencas de la región de estudio (CAR, 2012).....	30
Figura 9. Estaciones hidrométricas, climatológicas y de precipitación en la zona de estudio (CAR, 2012)	31
Figura 10. Estaciones instaladas por año	32
Figura 11. Longitud de la serie histórica	33
Figura 12. Estaciones con más del 80% de datos mensuales.....	34
Figura 13. Esquema del procedimiento	38
Figura 14. Ejemplo de Mapa de Hits	42
Figura 15. Ejemplo de Mapa de Distancia entre Neuronas.	43
Figura 16. Mapa de Hits para la Tipología No. 1	47
Figura 17. Mapa de Distancia entre Neuronas Tipología No. 1	48
Figura 18. Formación de grupos en la tipología No. 1	48
Figura 19. Distribución geográfica de los grupos para la Tipología No. 1	51
Figura 20. Mapa de Hits para la Tipología No. 2	52
Figura 21. Mapa de Distancia entre Neuronas Tipología No. 2	53
Figura 22. Formación de grupos en la tipología No. 2	53
Figura 23. Distribución geográfica de los grupos para la Tipología No. 2	57
Figura 24. Mapa de Hits para la Tipología No. 3	58
Figura 25. Mapa de Distancia entre Neuronas Tipología No. 3	58
Figura 26. Formación de grupos en la tipología No. 3	59
Figura 27. Distribución geográfica de los grupos para la Tipología No. 3	62
Figura 28. Pluviograma de las Estaciones E163 y E77	66
Figura 29. Histograma Serie 1 y 2.....	67
Figura 30. Serie 1 a partir de la Serie 2	67
Figura 31. Histograma Serie 1 Generada a partir de Serie 2.....	68
Figura 32. Comparación entre la Serie 1 Original y la Serie 1 generada	68
Figura 33. Serie 2 a partir de la Serie 1	69
Figura 34. Comparación entre la Serie 2 Original y la Serie 2 generada	69
Figura 35. Histograma Serie 2 Generada a partir de Serie 1	70
Figura 36. Histogramas bivariados	70

LISTA DE ANEXOS DIGITALES

Anexo 1: Datos de las estaciones y serie histórica de precipitación de la zona de estudio.

Anexo 2: Resultados detallado del Clustering.

Anexo 3: Resultados detallado del cálculo de la Transinformación.

Anexo 4: Documento Guía

RESUMEN

En este proyecto se evaluó el diseño de una red pluviométrica a partir de conceptos de entropía de la información y de redes neuronales artificiales. Para ello se analizó la red de estaciones registradas en la Corporación Autónoma Regional de Cundinamarca. La red estudiada está compuesta por 182 estaciones para la medida de la precipitación. Posee una serie histórica que va, en algunos casos, desde 1931 hasta 2012. Abarcó dos fases, en la primera se realizó una clasificación de las estaciones pluviométricas de la región de estudio mediante la aplicación de la técnica de Mapas Auto-organizados. En la segunda fase se evaluó el desempeño de la red mediante la aplicación de conceptos de entropía, teniendo en cuenta la distribución de las estaciones en cada uno de los grupos obtenidos previamente.

Para la clasificación, se plantearon tres escenarios denominados Tipologías, en los que se variaba el número de neuronas de la capa de salida. Se observó que al aumentar el número de estas neuronas, aumentó el número de grupos formados, esto viene a estar justificado en que, en un proceso de clustering un número alto de neuronas en la capa de salida puede ocasionar que el modelo se sobre-entrene, resaltando diferencias entre cada uno de los patrones. Resultó significativo que al realizar una comparación entre los resultados de las distintas tipologías, se encontró la permanencia de ciertos rasgos en la clasificación, lo que indica la validez de la clasificación. Posteriormente, se tomaron los resultados de la Tipología No. 1 y se evaluó la entropía de la serie histórica, calculando la transformación, lo que permitió concluir que la red se encuentra con déficit de información.

ABSTRACT

This project evaluated the rainfall network design from concepts of the information's entropy and the artificial neural networks. In order to do this we analyzed the stations' network registered in the "*Corporación Autónoma Regional de Cundinamarca*". The studied network comprises 182 stations for measuring precipitation. It has a series that goes in some cases through 1931 to 2012. It comprised two phases. In the first one, we did a classification of the rainfall stations in the studied region by applying the technique of Self-organizing Maps. In the second phase was evaluated the network performance by applying concepts of entropy, taking into account the distribution of the stations in each of the groups previously obtained.

For the classification, three scenarios were raised and they were called Types, in which were varying the number of neurons in the output layer. It was observed that by the increasing of the number of these neurons, also were increasing the number of the formed groups. This is justified because in a clustering process, a large number of neurons in the output layer can cause the model turns into an over-trained model, highlighting differences between each of the patterns. It was significant that when doing a comparison between the results of different types, we found the persistence of certain features in the classification, indicating the validity of the classification. Later, we took the results of the Type No. 1 and we evaluated the entropy of the time series, calculating the transformation, allowing conclude that the network has deficit in information.

Introducción

El concepto de entropía ha sido utilizado en áreas tan diversas como la economía (Vinod, 2006; J.P. Noonan, 2007), la ecología (Weber, 2010), la inferencia estadística (Payandeh Najafabadi, Hatami, & Najafabadi, 2011), la agricultura (Aurbacher & Dabbert, 2010), entre otras. Obteniendo resultados positivos que validan su aplicabilidad. En el campo de la hidráulica y de la hidrología, también ha sido utilizado con distintos fines, como por ejemplo, en la determinación de una metodología para el aforo de ríos (Calisto, 2002), para el análisis regional de precipitación (Daleziosa & Tyraskis, 2003) y para el estudio de la variabilidad de ésta en una zona determinada (Mishra, Özger, & Singh, 2009). Así mismo, el concepto de redes neuronales ha sido aplicado en campos tan variados como la electricidad (ani E. Asimakopoulou, 2013; A. Kheirkhah, 2013), la química (Hooman Adib, 2013), el reconocimiento de patrones visuales (Fukushima, 2013), la geotecnia (Riccardo Taormina, 2012), entre otros. De igual manera, esta técnica ha mostrado resultados positivos en el campo de la hidráulica y la hidrología (Anish C. Turlapaty, 2010; Jung-Woo Kim, 2010; Qian-Jin Liu, 2013).

En este proyecto se evaluó el diseño de una red pluviométrica a partir de conceptos de entropía de la información y de redes neuronales artificiales. Para ello se analizó la red de estaciones registradas en la Corporación Autónoma Regional de Cundinamarca. Se ofrece, como resultado de este proyecto, algunas recomendaciones generales para el rediseño de red pluviométrica en la región de estudio.

Este proyecto abarcó dos fases. En la primera se realizó una clasificación de las estaciones pluviométricas de la región de estudio mediante la aplicación de la técnica de Mapas Auto-organizados y en la segunda fase se evaluó el desempeño de la red pluviométrica de la región Bogotá-Cundinamarca a partir de conceptos de entropía. Se plantearon tres escenarios para la clasificación de las estaciones, obteniendo los siguientes resultados: en la Tipología No. 1 (100 neuronas en la capa de salida) se formaron 13 grupos, en la Tipología No. 2 (400 neuronas) se formaron 47 grupos y en la Tipología No. 3 (900 neuronas) se formaron 56 grupos. Posteriormente, se tomaron los resultados de la Tipología No. 1 y se evaluó la entropía de la información, calculando la transformación, lo que permitió concluir que la red se encuentra con déficit de información.

Capítulo 1

Aspectos Generales

1.1 Planteamiento del Problema

El comportamiento del clima depende de muchas variables, por lo que su predicción es un proceso complejo que requiere de información detallada. Las redes pluviométricas aportan información para el conocimiento y determinación del estado del clima y de las condiciones de caudal en una cuenca determinada. Con el fenómeno del cambio climático tal conocimiento cobra cada vez más importancia porque de su precisión puede depender la efectividad de las medidas tomadas para prevenir la afectación de la población frente a las condiciones extremas que se presenten. Aunque, si bien es cierto que, para el diseño de las redes existen diversos métodos, en la práctica éstas son ubicadas con criterios que no responden a requerimientos técnicos (Gontijo, 2007), debido a su complejidad. Este hecho se ve reflejado en que las estaciones de monitoreo son ubicadas en sitios redundantes, descuidando otras áreas.

1.2 Pregunta de Investigación

¿Cómo se pueden emplear los conceptos de entropía en la teoría de la información y de redes neuronales artificiales en la evaluación de la cobertura de redes pluviométricas?

1.3 Objetivos

1.3.1 Objetivo General

Evaluar la distribución de las estaciones de la red pluviométrica de la región Bogotá-Cundinamarca mediante la aplicación de conceptos de entropía de la información y de redes neuronales artificiales.

1.3.2 Objetivos Específicos

- Agrupar las estaciones pluviométricas de la región de estudio, de acuerdo a las zonas climatológicas homogéneas, mediante un análisis de frecuencia regional aplicando redes neuronales artificiales.
- Evaluar la distribución de las estaciones pluviométricas en cada una de los grupos obtenidos mediante un análisis de su registro histórico, aplicando conceptos de entropía de la información.
- Formular un documento guía con recomendaciones generales para el rediseño de red pluviométrica en la región de estudio.

1.4 Alcance y Relevancia del Proyecto

En este proyecto se evaluó la distribución de las estaciones de la red pluviométrica de la región Bogotá-Cundinamarca. Por tal región, entiéndase el área de jurisdicción de la Corporación Autónoma Regional de Cundinamarca (CAR). Aunque el número y cobertura de las estaciones analizadas estuvo supeditado a la disponibilidad gratuita de la información requerida.

La evaluación que se realiza de la red pluviométrica se limita a determinar la redundancia de la información aportada por las estaciones seleccionadas y a partir de esta valoración se proponen medidas para el rediseño de la red. Se hará una contribución específica al centrar el proyecto en el área de la región Bogotá-Cundinamarca.

El documento guía que se propone, recoge el procedimiento para la evaluación de la red de la región de estudio y las recomendaciones para el rediseño de la misma. Si bien, esta red ya existe y está operando, tal herramienta puede ser empleada para la mejora de la existente y para la proyección de una futura ampliación.

1.5 Estado del Arte

A continuación se hace un listado de los artículos que han servido como base al presente trabajo, a través de éste se demuestra que la novedad del proyecto radica en la combinación de técnicas de redes neuronales artificiales y entropía para la evaluación de

redes pluviométricas, además de ofrecer una novedad geográfica al tratarse de la región Bogotá-Cundinamarca.

En la Universidad de Brasilia se efectuó un proyecto titulado *Avaliação e redimensionamento de redes para o monitoramento fluviométrico utilizando o método sharp e o conceito de entropía* (Gontijo, 2007) cuyo objetivo principal fue proponer un procedimiento para la evaluación de redes de vigilancia fluviométrica, aplicando el método de Sharp y el concepto de entropía. Este trabajo se aplicó en la cuenca del río Sao Francisco. Este proyecto servirá en la presente investigación, en el desarrollo de la segunda fase de la metodología propuesta, puesto que aporta un procedimiento para la evaluación de redes a partir de entropía de la información. Aunque si bien, tal procedimiento fue pensado para una red fluviométrica se pueden realizar las adaptaciones correspondientes.

El concepto de entropía fue empleado para investigar la variabilidad asociada con series mensuales, estacionales y anuales de precipitación en el estudio titulado *An entropy-based investigation in to the variability of precipitation* (Mishra, Özger, & Singh, 2009). Se tomó como caso del estudio al estado de Texas y a partir de este proyecto se pudo hacer una caracterización de la precipitación en esta región, pudiendo plantear algunas sugerencias para el manejo eficiente del agua.

En el estudio *Application of an entropy-based Bayesian optimization technique to the redesign of an existing monitoring network for single air pollutants* (Ainslie, C. Reuten, Le, & Zidek, 2009) se emplea el método bayesiano basado en la entropía para el rediseño de una red de monitoreo de la contaminación en el aire. El trabajo demuestra la gran utilidad de la técnica de análisis, y también proporciona una guía estadísticamente defendible en el rediseño espacial de esta importante red de monitoreo. En este mismo sentido, fue empleado el concepto de entropía en el estudio *Modeling the space/time distribution of particulate matter in Thailand and optimizing its monitoring network* (Puangthongthub, Wangwongwatana, Kamens, & Serre, 2007) en el que se evaluó la distribución espacio temporal de PM10 en Tailandia, aplicando una modelación con el método de máxima entropía bayesiano a la red de monitoreo existente. A partir de los resultados se logró hacer ciertas recomendaciones para la optimización de la red. En 2011 se publicó el estudio *Design of ensemble neural network using entropy theory* (Zhao & Zhang, 2011) en

el que se usa el concepto de entropía para el diseño de redes neuronales, reduciendo el exceso de ajuste de las redes integrantes.

En 2007 se publicaron los resultados de una investigación titulada *On hydrologic calculation using artificial neural networks* (Fenga & Hongb, 2007) en el que se demuestra que la tecnología de redes neuronales artificiales proporciona una herramienta relativamente eficaz de resolver problemas en el cálculo hidrológico. Así mismo, el estudio *Hybrid neural network models for hydrologic time series forecasting* (Jain & Kumar, 2007), propone un sistema en el que se combinan las técnicas convencionales de para la predicción de series de tiempo y redes neuronales artificiales (RNA). Los resultados obtenidos sugieren que el método de combinar los puntos fuertes de las técnicas convencionales y las RNA proporciona un marco de modelado sólido capaz de capturar la naturaleza no lineal de las series de tiempo complejas y elaborar previsiones más exactas. Además, señala que aunque en este trabajo, los modelos híbridos se aplican a la hidrología, tienen un enorme margen para su aplicación en una amplia gama de áreas para lograr una precisión mayor en la predicción de series de tiempo.

En el estudio *Identification of homogeneous región for regional frequency analysis using the self-organizing map* presentado por Gwo-FongLin y Lu-Hsien Chen en 2005, se realiza una comparación entre los métodos de Ward, K-means y mapas auto organizados para la identificación de regiones homogéneas. Concluye que este último arroja mejores resultados. Luego, el mejor método es aplicado a datos reales de la red pluviométrica de Taiwán (Gwo-Fong Lin, 2005). En esta misma línea se encuentra el proyecto *Regional flood frequency analysis by combining self-organizing feature map and fuzzy Clustering* presentado en 2008 por Srinivas, ShivamTripathi, Rao y Govindaraju. En este proyecto se ha aplicado el método de mapas auto-organizados como técnica de agrupación para la regionalización hidrológica. Se demuestra que tal método puede servir como precursor útil para algoritmos de agrupamiento. En el proyecto se estudia la región de Indiana, estados Unidos (Srinivas, 2008). Estos dos últimos artículos resultan de gran ayuda en la primera fase del proyecto puesto que ofrecen una metodología de aplicación de redes neuronales artificiales (RNA) para la clasificación de zonas climatológicas.

En el trabajo titulado *Hydrometric network evaluation for Canadian watersheds* evalúa la red hidrológica de Canadá, identificando las principales estaciones de medición y las zonas donde la densidad de éstas sea deficiente. Para ello se utiliza la teoría de entropía. El método es aplicado a los datos de caudales diarios, a diferencia del presente estudio, en el cual la red estudiada es de precipitación, por tanto los parámetros empleados en la clasificación variarán, sin embargo este trabajo resulta útil porque presenta una metodología en la cual la entropía marginal conjunta y el índice de transinformación son utilizados para la identificación de las estaciones importantes, así como las áreas críticas en la cuenca. En este artículo se pone de manifiesto que debido a los impactos del cambio climático sobre los recursos hídricos, la presión sobre los recursos hídricos aumentará, y por tanto, también crece la necesidad de contar con una mejor red de información de tales recursos que permitan el diseño de proyectos orientados al desarrollo sostenible (A.K. Mishra, 2010).

En el trabajo *Modeling and forecasting river flow rate from the Melen Watershed, Turkey* presenta un nuevo enfoque utiliza redes neuronales artificiales para mejorar el rendimiento del pronóstico de la precipitación. Se desarrolló un estudio de caso de una subregión de Turquía que se encuentra cercana al Mar Negro. El objetivo de este proyecto fue estimar los valores perdidos en la serie histórica de precipitación que se poseía y generar pronósticos de la intensidad de la precipitación. Los resultados de este estudio prueban que las redes neuronales artificiales son un buen método para obtener pronósticos de precipitación con limitada cantidad de observaciones. (Muhammed Ernur Akiner, 2012)

En el artículo *Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques* publicado en *Journal of Hydrology* en 2012, se emplean las redes neuronales artificiales y el sistema neuro-fuzzy para la previsión del flujo de entrada mensual en un embalse. Este estudio es útil para comprender cómo a partir estos modelos se pueden generar medidas para el control de inundaciones, abastecimiento de agua para industria o la agricultura (A.K. Lohani, 2012).

La cuenca Abay Gilgel de Etiopía fue seleccionada como región de estudio del proyecto de investigación titulado *Using self-organizing maps and wavelet transforms for space-time pre-processing of satellite precipitation and runoff data in neural network based*

rainfall-runoff modeling. En éste, la técnica de Mapas Auto-organizados propuesta por Kohonen, la cual hace uso de redes neuronales artificiales, es empleada para identificar las agrupaciones espaciales homogéneas a partir de los datos satelitales de precipitación. En esta misma línea, en el estudio *Classification of microwater sheds based on morphological characteristics* publicado en 2011 en Journal of Hydro-environment Research, se emplean tres técnicas para la clasificación de 25 microcuencas de la gran cuenca Kherthal, Rajasthan. Las técnicas empleadas fueron *K-means Cluster Analysis (KCA)*, *Cluster Fuzzy Analysis (FCA)* y Redes Neuronales de Kohonen. Siendo ésta última, una de las empleadas en el presente proyecto. Se usaron diez parámetros, principalmente morfológicos, para el procedimiento. Este trabajo aporta, además, un método para la obtención del número óptimo de grupos, basándose en dos índices de validación de clúster, Daviese Bouldin y Dunn. (K. Srinivasa Raju, 2011).

Por otra parte, en el año 2006 el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) y la Alcaldía Mayor de Bogotá presentaron el informe titulado *Diseño de la ampliación de la red hidrometeorológica de Bogotá y cuenca alta del río Tunjuelo*. El objetivo principal de este estudio fue, mejorar el conocimiento del clima en el Distrito Capital, además proveer información hidrometeorológica para la elaboración de pronóstico de niveles de agua y condiciones climáticas en esta zona del país (IDEAM, 2007). Este documento aporta datos sobre cierto sector del área de estudio del presente proyecto, contribuyendo a la caracterización de la zona.

1.6 Marco Teórico

1.6.1 La precipitación

Precipitación es el término que se refiere a todas las formas de humedad venidas de la atmósfera y depositadas en la superficie terrestre, tales como lluvia, granizo, rocío, neblina, nieve o heleada.

Los elementos necesarios para la formación de las precipitaciones son los siguientes: Humedad atmosférica, radiación solar, mecanismos de enfriamiento del aire, presencia de núcleos, higroscópicos para que haya condensación y mecanismo de crecimiento de las partículas (Monsalve Sáenz, 1995).

Para que ocurra la precipitación, se requiere algún mecanismo que enfríe el aire lo suficiente para que llegue de esta manera a, o cerca, del punto de saturación. Los

enfriamientos de grandes masas, necesarios para que se produzcan cantidades significativas de precipitación, se logran cuando ascienden las masas del aire. Este fenómeno ocurre por medio de sistemas convectivos o convergentes que resultan de radiaciones desiguales, las cuales producen calentamiento o enfriamiento de la superficie de la tierra y la atmósfera, o por barreras orográficas (montañas). Sin embargo la saturación no necesariamente produce una precipitación (Kohler Max, 1977).

1.6.2 Medida de la precipitación

1.6.2.1 Medidas características

Las medidas de precipitación, que han de referirse a un intervalo de tiempo concreto, suelen expresarse en unidades de altura de agua sobre una superficie plana horizontal, generalmente en milímetros o centímetros, y más raramente en metros. Un milímetro es la altura que alcanza un litro de agua sobre una superficie de un metro cuadrado, por tanto 1 mm de lluvia equivale a un 1 l/m^2 .

Al ser la precipitación una variable discontinua en el espacio y en el tiempo, los valores medidos en una estación sólo son representativos para el punto de medida y su entorno. Es por esto que para caracterizar la precipitación de una zona es necesario contar con una red de pluviómetros. La densidad de la red pluviométrica depende de la orografía. La organización Meteorológica Mundial recomienda un pluviómetro por cada 250 km^2 en zonas de topografía suave (Martínez Alfaro, Martínez Santos, & Castaño, 2005). Algunas medidas en las que se expresa la precipitación son:

- **Altura pluviométrica:** Se expresa diariamente, mensualmente, anualmente, etc. Su unidad de medida es el milímetro.
- **Intensidad de precipitación:** Corresponde la razón entre la altura pluviométrica y un tiempo determinado. Se expresa en mm/hora.

$$i = \Delta h / \Delta t$$

- **Duración:** Es el período en horas, desde el inicio hasta el fin de la precipitación (Monsalve Sáenz, 1995)

En la Figura No. 1 se presenta un registro pluviográfico, en el cual, el eje de las ordenadas representa la precipitación mensual, mientras que el eje de las abscisas, el año correspondiente.

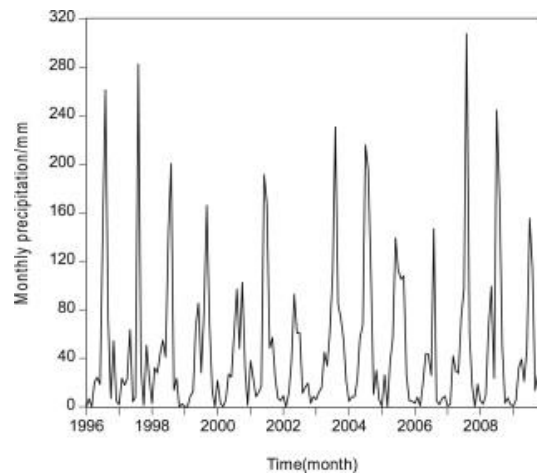


Figura 1. Registro pluviográfico típico (Shengwei Wang, 2011)

1.6.2.2 Los pluviómetros

Los pluviómetros son los aparatos destinados a medir la precipitación que se produce en un intervalo de tiempo determinado. En la Figura No. 2 se muestran algunos tipos de pluviómetros. Según el intervalo de tiempo que se considere hay:

Pluviómetros no registradores: registran la precipitación caída en un punto durante un intervalo de tiempo generalmente diario. Proporcionan el volumen total del agua precipitada, sin embargo no dan información sobre la distribución de la precipitación en el intervalo de tiempo considerado.

Pluviómetros registradores o pluviógrafos: proporcionan una gráfica, pluviograma, que es una curva continua de precipitación acumulada a lo largo del intervalo de tiempo que se considere. De un pluviograma puede obtenerse los siguientes datos:

- Hora de comienzo y fin de la lluvia.
- Volumen total de agua precipitada.
- Intensidad máxima de precipitación (volumen máximo de agua caída en la unidad de tiempo).
- Intensidad para cualquier intervalo de tiempo registrado en el pluviograma.
- Curvas de intensidad- duración- frecuencia.

Pluviómetros totalizadores: son pluviómetros no registradores de gran capacidad que se instalan en lugares de difícil acceso, por ejemplo en lugares que van a quedar aislados.

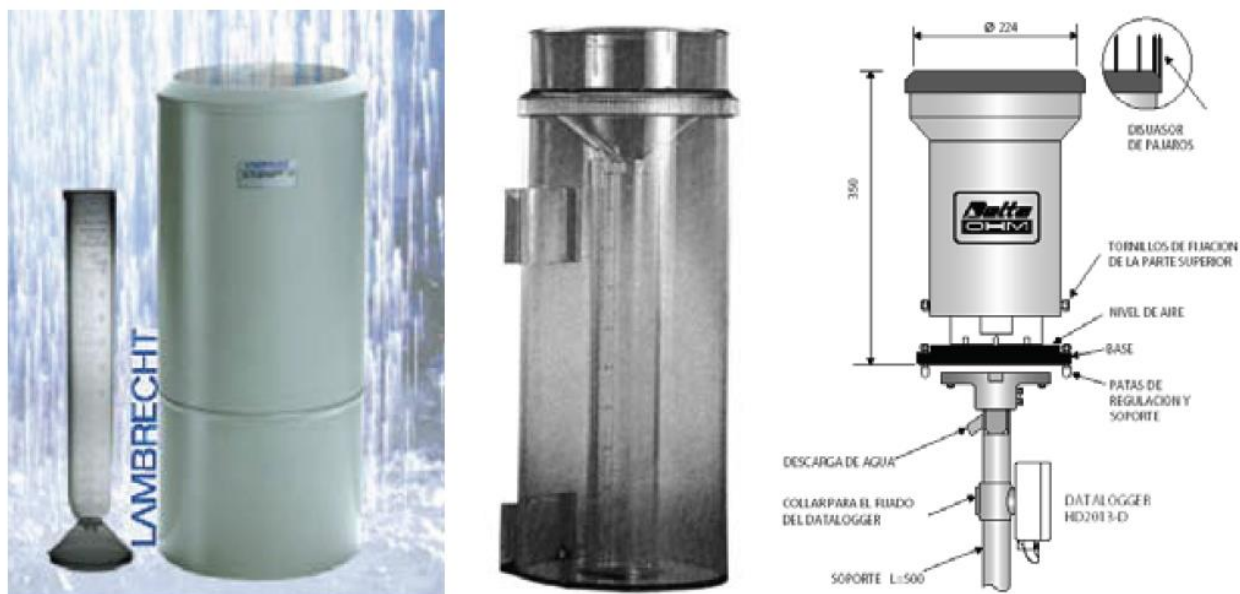


Figura 2. Algunos modelos de pluviómetros (OEA, 2010)

1.6.2.3 Red pluviométrica

Las redes pluviométricas deben ser planeadas de tal manera que se obtenga un cuadro representativo de la distribución espacial de la precipitación. No debe haber concentración de pluviómetros en áreas con mucha lluvia a costa de una densidad baja en áreas relativamente secas, o viceversa. Desafortunadamente, el costo de instalación y mantenimiento de una red y la accesibilidad al sitio del pluviómetro por un observador son siempre condiciones importantes. Los usos que se le dará a la información obtenida sobre la precipitación deberían determinar la densidad de la red. Una red de estaciones relativamente dispersa debería bastar para el estudio de grandes tormentas, o para determinar promedios de grandes áreas planas. Se necesita una red bastante densa para determinar el patrón de lluvia de un aguacero (Kohler Max, 1977). En la Figura No. 3 se muestra la Red Nacional Colombiana de estaciones hidrológicas y meteorológicas.

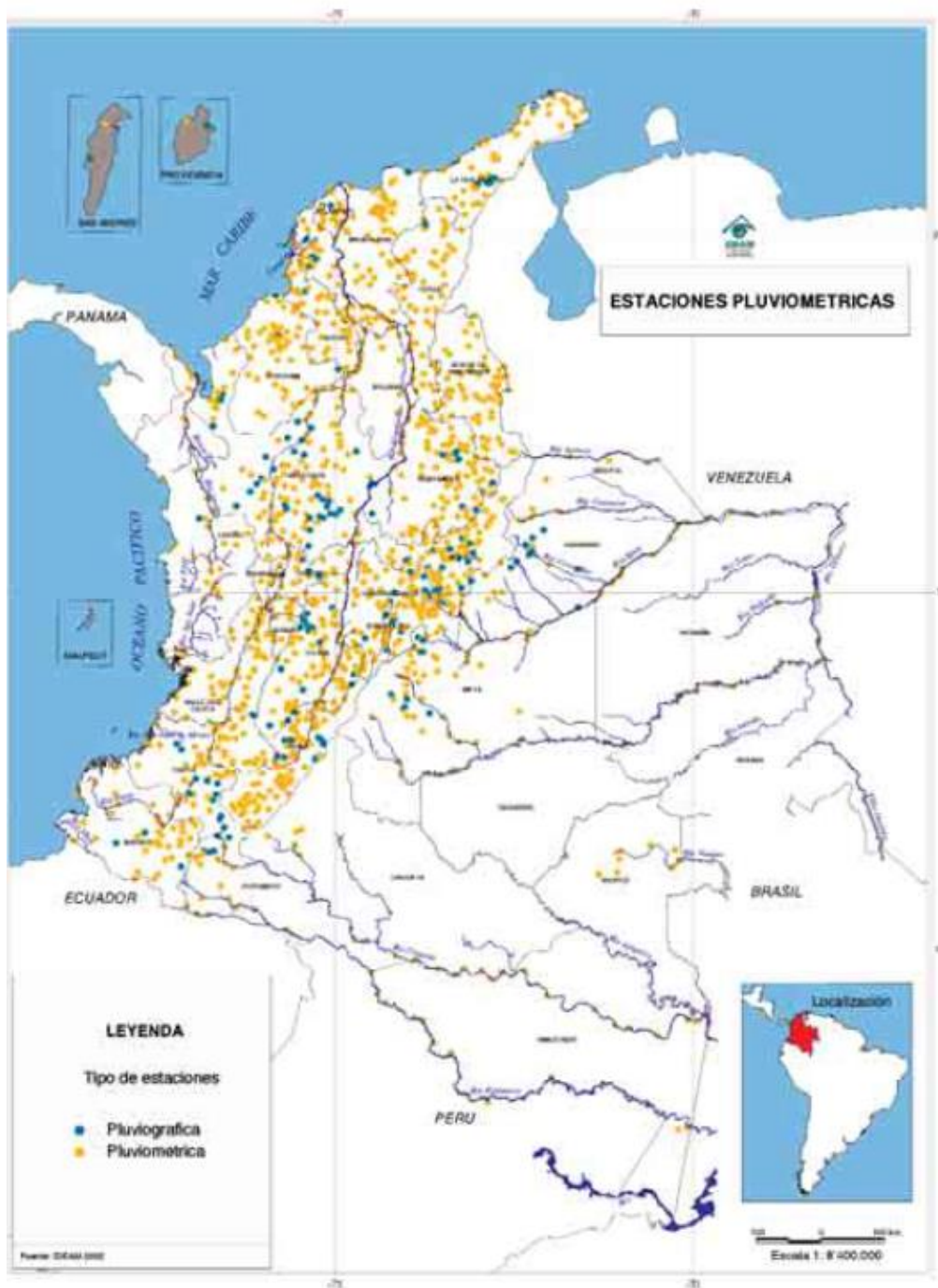


Figura 3. Red Pluviométrica de Colombia (IDEAM, 2002)

1.6.2.4 Diseño de la red pluviométrica

De acuerdo con el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), en general, los pluviómetros deberían estar tan uniformemente distribuidos como sea posible, teniendo en cuenta las necesidades prácticas de la información y la ubicación de los observadores voluntarios. En regiones montañosas la atención debe darse a la clasificación vertical utilizando aforos de almacenamiento para medir las precipitaciones en grandes altitudes. Así mismo, recomienda que exista coordinación con la ubicación de las estaciones de aforo de caudal, puesto que así se mejorará el sistema de información para la prevención de inundaciones. Es recomendable que los pluviómetros estén localizados de modo que la precipitación en la cuenca pueda estimarse por cada estación de aforo de caudal (IDEAM, 2007).

El Departamento de Desarrollo Sostenible de la Organización de los Estados Americanos, publicó en 2010 un *Manual para el diseño, instalación, operación y mantenimiento de sistemas comunitarios de alerta temprana ante inundaciones*, en el que se plantean las siguientes recomendaciones para la ubicación de equipos de medición de lluvias y de niveles del río:

- Elaborar el mapa cartográfico de la cuenca.
- Identificar los ríos principales, elevaciones principales. Se identifican las comunidades.
- Recabar toda la información de campo o entrevistas (principalmente con los comunitarios) sobre inundaciones históricas, régimen de lluvias, etc. Se hace un análisis preliminar de la información.
- Identificar las sub-cuencas, considerando los afluentes que trasladan más caudal.
- Utilizar los mapas de precipitación como ayuda para entender el régimen de las lluvias. Recolectar información sobre las estaciones hidrometeorológicas en la zona o áreas cercanas.
- Usar todo tipo de información, datos, mapas, documentos, etc. para apoyar el diseño del sistema (OEA, 2010).

En este mismo manual, se dan las siguientes recomendaciones sobre el sitio dónde se ubicará el pluviómetro:

- Que sea accesible a los voluntarios que harán las lecturas y que permita garantizar su seguridad física.
- Es preferible un sitio en la parte alta de la cuenca.
- Se debe cubrir toda la extensión de la cuenca (mínimo una estación).
- Buscar un área despejada de follajes y sin peligro de inundación.
- Sitio relativamente plano.
- El pluviómetro debe ubicarse a un metro sobre la superficie del suelo, a nivel, evitando cualquier inclinación, sobre una base vertical (OEA, 2010).

En cuanto al número mínimo de pluviómetros en una red, la Organización Meteorológica Mundial (OMM), propone un número de estaciones dependiendo unidad fisiográfica de la región donde se instalará la red. En la Tabla No. 1 se presenta la densidad mínima recomendada de estaciones pluviométricas.

Unidad fisiográfica	Densidades mínimas por estación (superficie en km ² por estación)	
	Sin registro	Con registro
Zonas costeras	900	9000
Zonas montañosas	250	2500
Llanuras interiores	575	5750
Zonas escarpadas/ondulantes	575	5750
Pequeñas islas	25	250
Zonas urbanas		10 a 20
Zonas polares y áridas	10000	10000

Tabla 1. Densidad mínima recomendada de estaciones pluviométricas (OMM citado por (OEA, 2010))

Criterios de Escalas de Representatividad

Los fenómenos meteorológicos no tienen todos las mismas dimensiones, existen diferencias significativas entre el tamaño de una tormenta de carácter local de tipo convectivo y las presentadas en cadena a lo largo de una línea de inestabilidad o un frente meteorológico. De la misma manera, si contemplamos cualquier otra variable, temperatura, viento, precipitación, etc., vamos a encontrar cambios generados así sea en

un entorno que se estima homogéneo, y más aún si las condiciones comienzan a tener comportamientos físico-geográficos diferentes, la variación de dichas variables va a ser mayor.

Por ello, al proyectar una red meteorológica, conviene tener en cuenta la magnitud de los fenómenos que se van a observar y medir, además la variabilidad climática de la zona, su aspecto y factores geográficos y orográficos, coberturas de vegetación, población, centros industriales, cuencas hidrográficas, etc. Lo anterior contempla entonces que se necesitan de ciertas escalas a saber:

- Micro fenómenos - Distancia en superficie hasta de 20 kilómetros.
- Meso fenómenos - Distancia en superficie entre 30 a 80 kilómetros.
- Macro fenómenos - Distancia en superficie hasta 150 kilómetros.

1.6.3 Redes neuronales artificiales

Una red neuronal artificial (RNA) es un procesador inspirado en las redes neuronales biológicas, que puede almacenar el conocimiento experimental. El procedimiento para el aprendizaje se conoce como un algoritmo de aprendizaje. Su función es modificar los pesos sinápticos de las redes con el fin de alcanzar un objetivo especificado de antemano. La neurona es la unidad fundamental para el funcionamiento de una red neuronal (Karabacaka & Cetin, 2013). Las Redes neuronales artificiales son programas informáticos que pueden simular el comportamiento de las redes neuronales humanas, teniendo mecanismos de auto -aprendizaje, y además de esto, también tienen la capacidad de crear relaciones entre la memorización y la información en sí (Dursun & Özden, 2014).

1.6.3.2 Modelo de una neurona artificial

En la Figura No. 4 se muestra el modelo de una neurona; éste este es el elemento básico de una red neuronal artificial. Aquí se pueden identificar los elementos constitutivos del modelo, a saber: enlaces de conexión (pesos sinápticos), punto de suma, valor de activación y señal de salida.

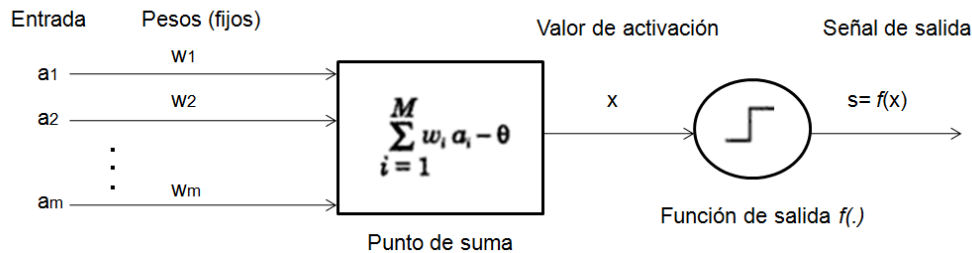


Figura 4. Modelo de una neurona artificial (Yegnanarayana, 2006)

1.6.4 Clustering: métodos y aplicaciones

1.6.4.1 Definición y aplicaciones

El análisis de clúster o clustering es el nombre genérico otorgado a una gran variedad de técnicas que tienen como objetivo la búsqueda de grupos en un conjunto de individuos. En general, todo método de clasificación parte de un conjunto de elementos singulares que deben ser clasificados en un número reducido de grupos o clúster, obtenidos por particiones sucesivas del conjunto original y en los que se respete la estructura relacional que en el mismo se mantenía (Fernández, 1991). Es un campo de análisis exploratorio que busca la organización de un conjunto de elementos en grupos, partiendo de supuesto de que los miembros de un grupo determinado tienen alto grado de similitud (Carvalho, 2008). Los métodos de la agrupación se pueden clasificar en métodos jerárquicos y los métodos de partición. Dentro de cada uno de los tipos, existen una gran cantidad de subtipos diferentes y de algoritmos para la construcción de los grupos (Aliguliyev, 2009).

A medida que avanza la tecnología de la información, grandes cantidades de datos son almacenadas en las bases de datos (Ching-Hsue Chenga, 2010), por lo que se requieren métodos más eficientes para la clasificación de éstos. Además, el problema de la clasificación aparece en muchos campos más, como la minería de datos, el reconocimiento de patrones, el análisis estadístico y la bioinformática (Tomi Kinnunen, 2011). Este hecho hace que el análisis exploratorio sea un tema frecuente de investigación, puesto que el procesamiento electrónico de datos, pone a disposición de los investigadores grandes cantidades de información, procedentes de fuentes diferentes de todo el mundo (Sambu Seo, 2004).

1.6.4.2 Etapas en un análisis de clúster

Selección de las variables que van a ser incluidas en el análisis

En el análisis de clúster se debe prestar singular atención a los datos seleccionados, puesto que los grupos resultantes estarán condicionados por éstos. Se resaltan tres aspectos a considerar en esta fase:

- Que las variables estén in-correlacionadas.
- Que la unidad de medida sea la misma para todas las variables analizadas (estandarizar).
- Que el número de variables no sea demasiado grande (Vidal Díaz, 1998).

Elección de la técnica de agrupamiento

Las técnicas de agrupamiento se dividen en: Jerárquicas y No jerárquicas. Los métodos de agrupamiento jerárquico unen o separan grupos de información similar sucesivamente. Si el método jerárquico es de agrupación, parte de un número de grupos igual al número de observaciones y los grupos similares se van uniendo hasta formar un solo grupo; es claro que la homogeneidad de cada grupo en este caso se va reduciendo a medida que disminuye el número de grupos. En cambio, si el método jerárquico es divisivo, el funcionamiento es contrario al de agrupación, es decir que se parte de un grupo que incluye todas las observaciones, el cual se va dividiendo en grupos cada vez más homogéneos hasta obtener un número de grupos igual al número de observaciones. El número de grupos en los que se divide la información analizada depende del nivel o jerarquía que se observe en el dendograma. Los métodos de agrupamiento no jerárquicos o particionales, asumen de entrada un número fijo de grupos a los cuales se les asigna aleatoriamente una de las observaciones o patrones como centroide. El proceso de agrupamiento consiste en que los patrones se van incluyendo al grupo del centroide más cercano y con base en el nuevo agrupamiento, el centroide se calcula como el valor medio de las observaciones que pertenecen a ese grupo. Este proceso se detiene cuando el valor del centroide no cambia. El método no jerárquico más empleado para el agrupamiento de información es el denominado K-medias (*K-means* en inglés) (González, 2012). El algoritmo de K-medias intenta encontrar los centros de clúster (c_1, c_2, \dots, c_K) de tal manera que la suma de las distancias al cuadrado de cada punto de datos (x_i) a su más cercano centro de agrupación (c_K) se reduzca al mínimo (Kalyani & Swarup, 2011).

La elección de que técnica va a ser utilizada es algo que depende del propio investigador, y tiene que ser compatible con su deseo de clasificación, la métrica de las variables utilizadas y la medida de similitud seleccionada (Vidal Díaz, 1998).

Selección de las medidas de proximidad y distancia

El objetivo de esta fase es medir «qué grado de similitud» o «de diferencia» tienen los individuos seleccionados. Para ello, se han desarrollado índices de distancia y de proximidad que permiten medir el grado de diferencia o semejanza entre dos individuos. Es importante recordar que la distancia cambiará en función de las variables seleccionadas y el coeficiente de ponderación (si es que existe) asignado a cada una. En la Tabla No. 2 se presentan algunas de las medidas más utilizadas (Vidal Díaz, 1998).

Datos binarios	Jaccrd
	Russel y Rao
	Sokal y Sneath
	Rogers y Tanimoto
VARIABLES DISCRETAS	Chi-Cuadrado
	Phi- Cuadrado
VARIABLES CONTINUAS	Distancia euclídea, euclídea al cuadrado
	Coficiente de correlación entre vectores (similitud)
	Coseno de vectores (similitud)
	Distancia métrica de Chebynev
	Distancia Manhattan o City-block
	Distancia de Minkowki

Tabla 2. Clasificación de las principales medidas de distancia (Vidal Díaz, 1998)

4.6.5 Redes de Kohonen

La Red de Kohonen es un tipo de Red Neuronal Artificial de aprendizaje no supervisado, es decir, no hay ninguna información previa sobre la existencia de clúster en el conjunto de datos durante el proceso de aprendizaje. Esta red está formada por un sistema de auto-organización capaz de proyectar los datos de alta dimensión en un espacio de dimensiones bajas, especialmente el espacio de dos dimensiones lo que facilita la interpretación del conjunto de datos sin perder la información original.

La visualización e interpretación de la Red de Kohonen permite a los usuarios suponer que todas las muestras indexadas en la misma neurona o en sus alrededores son similares de acuerdo a las características que se están evaluando. Este esquema permite la fácil detección de similitudes en las muestras y de valores atípicos (Ribeiro & Bezerra, 2014).

El algoritmo de aprendizaje utilizado para establecer los valores de los pesos de las conexiones entre las N neuronas de entrada y las M de salida es el siguiente:

1. Inicializar los pesos (w_{ij}) con valores aleatorios pequeños y fijar la zona inicial de vecindad entre las neuronas de salida.
2. Presentar una entrada en forma de vector $E_k = (e_1^{(k)}, \dots, e_N^{(k)})$, donde los componentes $e_i^{(k)}$ serán números reales.
3. Determinar la neurona vencedora de la capa de salida, la cual será la que tenga el valor más parecido al patrón de entrada E_k . Para ello, se calculan las distancias o diferencias entre ambos vectores, considerando una por una todas las neuronas de salida:

$$d_j = \sum_{i=1}^8 (e_i^k - w_{ij})^2 \quad 1 \leq j \leq M$$

$e_i^{(k)}$: Componente i-ésimo del vector k-ésimo de entrada.

w_{ij} : Peso de la conexión entre la neurona i de la capa de entrada y la neurona j de la capa de salida.

4. Una vez localizada la neurona vencedora (j^*), se actualizan los pesos de las conexiones entre las neuronas de entrada y dicha neurona, así como los de las conexiones entre las de entrada y las neuronas vecinas de la vencedora.

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)[e_i^k - w_{i^*j}(t)]$$

$\alpha(t)$: Parámetro de ganancia o coeficiente de aprendizaje, con un valor entre 0 y 1, el cual decrece con cada iteración.

5. El proceso se debe repetir, volviendo a presentar todo el juego de patrones de aprendizaje.

El mapa construido (SOM) es una representación plana de los vectores prototipo, imaginados como puntos localizados en el espacio de datos. La eficiencia de esta representación es medida por dos índices:

1. Error de Cuantización. Este error corresponde al promedio de la distancia Euclidiana de los vectores de datos a sus representantes más cercanos.
2. Error Topológico. Este error indica cual es la fracción de vecinos en el mapa, los cuales no tienen regiones de *Voronoi* en el espacio de datos (Niño & De Moya, 2005).

4.6.5 Arquitectura de las redes de Kohonen

Las Redes de Kohonen están conformadas por una capa de entrada en la cual se presentan los patrones de entrada al modelo y una capa de salida, allí los pesos de las neuronas se actualizan a partir de los patrones de entrada. La capa de salida es el espacio bidimensional que se “auto-organiza” en función de la estructura de los patrones de entrada.

Cada una de las neuronas de la capa de entrada se conecta con todas las neuronas de la capa de salida. Si bien físicamente no hay conexiones laterales entre neuronas de la misma capa, el concepto de vecindad y sobre todo la actualización de los pesos de la capa de salida en función de la vecindad de la neurona ganadora, generan un vínculo de similitud entre neuronas cercanas que finalmente se ve reflejado en el agrupamiento o auto-organización de neuronas con características similares.

En la Figura No. 5 se observa una representación de la arquitectura de este tipo de modelo de redes neuronales.

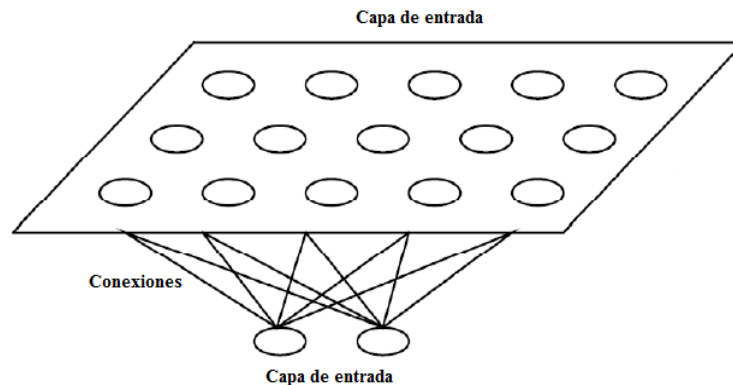


Figura 5. Arquitectura de las redes de Kohonen (González, 2012)

Generalmente, se emplea un arreglo bidimensional de neuronas en la capa de salida con el fin de facilitar la visualización de los resultados y la creación de los mapas topológicos. Las conexiones entre las capas van de la capa de entrada a la capa de salida, haciendo que la red se alimente hacia adelante (*feed forward*).

Existen varias topologías que condicionan la forma de la vecindad de cada neurona. En la Figura No. 6 se observan las tres topologías más usadas donde las neuronas se representan con el color rojo (González, 2012)

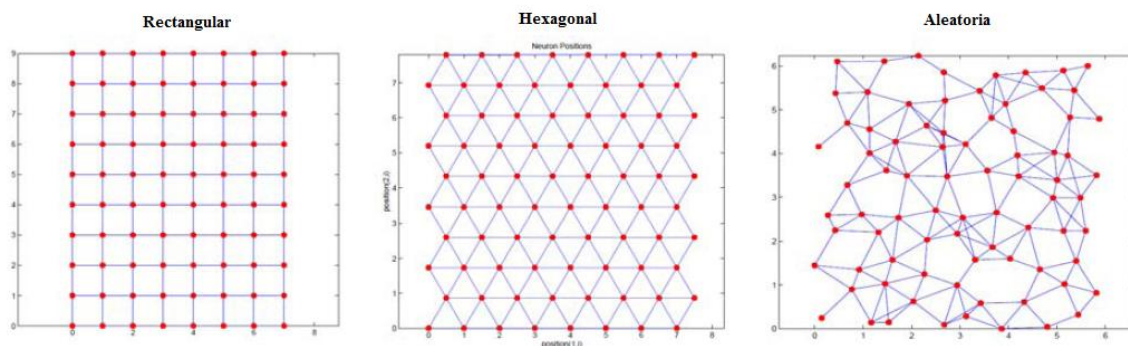


Figura 6. Ejemplos de Topologías de redes de Kohonen (González, 2012)

4.6.6 La Entropía en la teoría de la Información

Las bases de la teoría de la información fueron establecidas por Shannon en 1948. Ésta es definida como la medida del desorden o la peculiaridad de ciertas combinaciones

(Shanon, 1948). Fue concebida como una herramienta para el diseño de sistemas de comunicación. Pero, en la práctica, las ideas de la teoría de la información han sido útiles en aplicaciones diferentes del objetivo original. La información, en el sentido popular, se refiere al significado de ideas, signos, significado literal, porque interesa medir numéricamente la cantidad de información o, simplemente, la información contenida en un símbolo o mensaje.

La teoría de la información es una teoría matemática abstracta, en la cual se supone que hay una fuente de información con n símbolos s_1, s_2, s_n y que cada símbolo tiene una probabilidad de ocurrir: $p(s_1) = p_1, p(s_2) = p_2, \dots, p(s_n) = p_n$. Nada se supone en relación con los propios símbolos o sus significados, tan sólo que cada símbolo puede ser identificado individualmente. No es parte de la teoría definir las probabilidades p_i . La pregunta es: Cuánta información entrega la recepción de un símbolo s_i , si por ejemplo $p_1=1$ (todos los otros p_i serán entonces iguales a cero), la recepción de s_1 no causará "sorpresa", pues no hay información en un mensaje que afirma ocurrirá algo conocido. Pero, si todos los p_i son diferentes entre sí, la mayor sorpresa o información se tendrá al recibirse un símbolo poco probable o cuya probabilidad de ocurrir sea baja.

Por ejemplo, habrá información en una previsión de temperatura muy baja (evento poco probable) en plena estación de calor. Por otro lado, el aviso anticipado de temperatura alta (evento muy probable en la misma estación de calor) nos dará poca información.

Por consiguiente, la medida numérica de la información puede relacionarse de manera inversa con la probabilidad de ocurrencia del evento. Sea un evento y la transmisión de un mensaje (un símbolo s_i) comunicando su ocurrencia, la información transmitida para un receptor es definida como (Goldman, 1954, citado por (Calisto, 2002)):

$$\text{Información recibida} = I(s_i) = \text{Log} \left(\frac{p'_i}{p_i} \right)$$

Donde p'_i es la probabilidad del evento en el receptor después de la llegada del mensaje; p_i es la probabilidad del evento en el receptor antes de la llegada del mensaje.

Los mensajes recibidos pueden ser iguales a los transmitidos (transmisión sin ruido) o diferentes de los que se transmiten (transmisión con presencia de ruido). La situación sin ruido es la pertinente para el presente análisis.

En la transmisión sin ruido, el receptor tiene certeza de recibir la información correcta, pues todos los mensajes que obtiene son los mismos que se transmiten; por consiguiente, p_i . De este modo, la información depende del valor de la probabilidad anterior a la recepción del mensaje:

$$\text{Información recibida} = I(s_i) = -\text{Log}p_i$$

El uso del logaritmo hace aditiva la información de mensajes independientes, ya que la probabilidad de ocurrencia simultánea de un número de eventos independientes es igual al producto de las respectivas probabilidades individuales. En la teoría de la información se usa log, La unidad de información se obtiene cuando $p_i =$ porque: $-\log, = \log, = \text{bit}$.

Un sistema de transmisión de información es un sistema que envía una secuencia de símbolos en una clave y a una tasa determinadas. Determinar una medida numérica para la tasa media máxima a la cual un sistema de transmisión puede transmitir información es uno de los intereses del proyecto de sistemas de comunicación.

Un sistema convencional de transmisión de información debe considerar lo siguiente: una fuente de información, una clave para descifrar la información, un canal de transmisión, el descifrar y recuperar la información original a partir de la señal recibida y un receptor. Pero, ¿cómo descifrar, transmitir y recuperar la información original? Si bien la fuente de información es aleatoria, al cifrar la información se establecen mensajes específicos. Sin embargo quien diseña el sistema no tiene manera de anticipar el mensaje específico que será transmitido. Así, cualquier mensaje enviado es una muestra aleatoria de la población de todos los mensajes posibles y el sistema de transmisión debe ser diseñado para tratar cualquier mensaje posible. Por lo tanto, la teoría es esencialmente de probabilidad.

Al cifrar la información se forman secuencias ergódicas (controladas por probabilidades). Un proceso ergódico es un proceso estocástico cuando las secuencias ergódicas son puramente aleatorias (generadas por un proceso aleatorio). Un proceso estocástico es

ergódico cuando las propiedades estadísticas del proceso pueden ser determinadas a partir de una Única realización del proceso (son función de la muestra). Las secuencias ergódicas consideradas en la teoría de la información son idénticas a las series de tiempo estacionarias. Según la ley de los grandes números de la teoría de la probabilidad, cualquier secuencia suficientemente grande y cuya composición de frecuencias relativas esclarece las probabilidades de los símbolos del proceso aleatorio es una secuencia ergódica o típica del proceso. Se tiene una fuente de información ergódica que genera una secuencia grande den símbolos. Cada símbolo si conducirá hacia un receptor $I(s_i) = \log p_i$ unidades de información. De los n símbolos diferentes, el símbolo si podrá ocurrir $N_i = n p_i$ veces, con la condición de que n sea suficientemente grande. Luego, cada símbolo proporciona una cantidad total de información igual a $N_i (-\log p_i) = n p_i \log p_i$. Todos los símbolos son independientes entre sí; por lo tanto, la información total conducida por todos los n símbolos de la secuencia puede obtenerse al sumar la información de cada símbolo: $(-n \sum p_i (\log p_i))$ La información media será:

$$\text{Información media símbolo} = - \sum_{i=1}^n p_i \log p_i = \text{entropía} = H$$

Se observa, entonces, que la entropía de la teoría de la información es función solamente de la distribución de probabilidades p_i , porque no depende de los s_i . Esta característica permite que la entropía de la teoría de la información pueda ser útil en situaciones ajenas a los sistemas de comunicación (Calisto, 2002).

La Transinformación o información mutua es una medida de la información contenida en un proceso sobre otro proceso, está dada por:

$$T(X, Y) = H(X) + H(Y) - H(X, Y)$$

$H(X)$ corresponde a la entropía marginal de los datos reales, $H(Y)$ corresponde a la entropía marginal para la serie simulada y $H(X, Y)$ es la entropía conjunta (Mishra & Coulibaly, 2009).

Capítulo 2

Descripción de la zona de estudio

En este capítulo se presentan datos relevantes de la zona de estudio, la cual corresponde al área de jurisdicción de la Corporación Autónoma Regional de Cundinamarca. Se presenta la red pluviométrica disponible y sus características. Al final de este capítulo se muestran las estaciones pluviométricas que serán empleadas en el presente estudio.

2.1 Localización de la zona de estudio

La región de estudio abarca un área de 18.615 Km² aproximadamente, donde se encuentran 104 municipios: 98 pertenecientes al departamento de Cundinamarca, 6 al departamento de Boyacá y la zona rural del Distrito Capital de Bogotá D.C, corresponde al área de jurisdicción de la Corporación Autónoma regional de Cundinamarca (CAR, 2011). Este territorio en su mayor extensión es rural, con vocación agrícola y pecuaria, aunque en zonas como Soacha, Sabana Centro y Sabana Occidente a la par con el desarrollo agrícola y pecuario tienen un gran desarrollo industrial; igualmente en la zona de Ubaté se presenta un desarrollo minero considerable (CAR, 2009). En la Figura No. 7 se muestra un mapa de la zona de estudio.

Topográficamente, el 30% de la zona de estudio se ubica en alturas entre los 2.500 a 3.000 m.s.n.m. (Sabana de Bogotá y valle de Ubaté-Chiquinquirá con sus laderas colindantes), lo que configura la llamada Alta Montaña, mientras que cerca del 16% del área corresponde a las llamadas cumbres andinas (superiores a los 3.000 m.s.n.m) (CAR, 2012).

2.2 Aspectos Ambientales

La cuenca es la unidad más importante para la gestión ambiental, ya que permite el análisis ecosistémico del territorio en función del agua que es el eje estructural de la misma. La zona de estudio está dividida en nueve cuencas hidrográficas de segundo orden (Sumapaz, Bogotá, Vertiente Oriental del río Magdalena, Negro, Minero, Ubaté y Suárez, Blanco, Gachetá y Mchetá) y una cuenca de tercer orden (Tunjuelo). (CAR,

2011). En la Tabla No. 3 se muestran las áreas correspondientes a cada una de estas cuencas, mientras que en la Figura No. 7 se muestra su ubicación.



Figura 7. Área de estudio (CAR, 2012)

Vertiente	Cuenca	Área (km²)
Magdalena	Río Sumapaz	2527,2
	Río Bogotá	5671
	Río Magdalena - Vertiente Oriental	2191,4
	Río Negro	4238,6
	Río Minero	990,8
	Río Ubaté – Suárez	1965,4
	Orinoquía	Río Blanco
Río Gachetá		97,3
Río Machetá		508,7

Tabla 3. Cuencas hidrográficas de la región de estudio. (CAR, 2011)

A continuación se presentan algunos datos relevantes de cada una de estas cuencas:

2.2.1 Cuenca Río Sumapaz

La cuenca del Río Sumapaz se encuentra localizada en los departamentos de Cundinamarca y Tolima. En el área que corresponde al presente estudio, la cuenca tiene una extensión de 2.527 km², abarcando doce municipios con una población estimada de 225.000 habitantes, siendo el municipio de Fusagasugá el principal polo de desarrollo de la región. La cuenca del Río Sumapaz está conformada por diez cuencas de tercer orden: Ríos Alto Sumapaz, San Juan, Pilar, Medio Sumapaz, Negro, Cuja, Panches, Pagüey, Bajo Sumapaz y quebrada Negra (CAR, 2011).

2.2.2 Cuenca Río Bogotá

La cuenca del Río Bogotá se encuentra localizada en el departamento de Cundinamarca con una superficie total de 5.671km², cuenta con 7 millones de habitantes, distribuidos en 42 municipios, de los cuales 6.6 millones se localizan en grandes centros poblados, entre los que se destacan Bogotá, Girardot, Zipaquirá y Chía, está conformada por 19 subcuencas de tercer orden: Por los Ríos Bogotá (sector Soacha), Río Balsillas, Río Bajo Bogotá, Río Apulo, Río Neusa, Río Tunjuelo, Río Teusacá, Río Medio Bogotá (Sector Salto – Apulo), Alto Bogotá, Calandaima, Bogotá (Sector Sisga – Tibitóc), Frío, Bogotá

(sector Soacha –Salto), Soacha, Negro, por los embalses del Sisga, Chicú, Muña, y Tominé (CAR, 2011).

El Río Bogotá constituye la corriente principal de la cuenca recorriendo desde su nacimiento a los 3.300 msnm en el municipio de Villa Pinzón (sub-cuenca Río Alto Bogotá), hasta su desembocadura al Río Magdalena a los 280 msnm en el municipio de Girardot (sub-cuenca Río Bajo Bogotá Apulo – Girardot), recorriendo un total de 308 km.

El 30% de la cuenca se caracteriza por presentar un relieve fuertemente ondulado a fuertemente quebrado, con pendientes entre 12 y 25 %; le sigue en importancia el relieve fuertemente quebrado con pendientes entre 25 y 50%. Estas condiciones de relieve y pendiente se presentan de forma general en toda la cuenca, con excepción del área comúnmente denominada como sabana y en la parte baja de la cuenca, donde predominan las pendientes entre 0 y 7 % y los relieves planos, ligeramente plano, ligeramente inclinado y ligeramente ondulado (CAR, 2006).

2.2.3 Cuenca de la vertiente oriental del Río Magdalena

Localizada al oriente del departamento de Cundinamarca, presenta un área de drenaje de 2.191 km², en jurisdicción total o parcial de 11 municipios, alberga una población de 235.000 habitantes, sobresaliendo los centros poblados de Girardot y Puerto Salgar. En el área de estudio, el Río Magdalena tiene una longitud de 228 Km y drenan once subcuencas de tercer orden: De los ríos Magdalena sector Nariño, Río Seco, Río Magdalena sector Beltrán, Río Seco de Las Palmas, Chaguaní, Magdalena sector Guaduas, Seco Norte, Magdalena sector Puerto Salgar, Negrito, quebrada Seca y Doña Inés (CAR, 2011).

2.2.4 Cuenca Río Negro

La cuenca del Río Negro se localiza en el noroccidente del departamento de Cundinamarca y sur occidente de Boyacá, presenta una extensión de 4.238 km², teniendo como característica principal la alta oferta hídrica y las bajas densidades de población, destacándose los centros poblados de Villeta, Pacho y Yacopí. La cuenca del Río Negro está conformada por dieciséis cuencas de tercer orden: Ríos Bajo Negro, Guaguaquí, Terán, Macopay, Cambras, Guatachí, Guaduro, Medio Negro 1 y 2, Patá, Tobia,

Pinzaima, Murca, Alto Negro y quebradas Guaguaquí, Guatachí, Negra y Terama (CAR, 2011).

2.2.5 Cuenca Río Minero

La cuenca del Río Minero forma parte de la cuenca hidrográfica del Río Carare, tiene un área de drenaje de 7.523 km², en jurisdicción de los departamentos de Boyacá y Santander. En el territorio de la CAR la cuenca tiene una extensión de 991 km², abarcando cerca del 5.3% del área, se localiza al norte del departamento de Cundinamarca y occidente del departamento de Boyacá, presenta una población aproximada de 20.500 habitantes distribuidos en nueve municipios, se caracteriza por un paisaje de topografía de altas pendientes, bajas densidades de población a lo largo de la cuenca y problemas de remoción en masa (CAR, 2011). La mayor parte del área tiene una pendiente que oscila entre 25 –50% (Subcuencas Palenque, Piedras y Villamizar), le sigue en predominio pendientes entre > 50% (Subcuencas Guaquimay y Negro), 12 – 25% en varios sectores de las subcuencas y en menor proporción suaves entre el 0 – 7% en pequeños sectores en las márgenes de los cauces (CAR, 2009).

2.2.6 Cuenca Ríos Ubaté y Suárez

La cuenca del Río Suárez drena parte de los departamentos de Cundinamarca, Boyacá y Santander en una extensión total de 7.662 km², de los cuales 1.748 km² corresponden al área de jurisdicción de la CAR, ocupando cerca del 10% del territorio. En el área CAR la superficie de la cuenca es de 1.965 km², diferenciándose dos sistemas hidrográficos, la cuenca del Río Ubaté en la parte alta y la cuenca del Río Suárez en la parte plana aguas abajo de la laguna de Fúquene, abarcando en su totalidad 17 municipios con una población estimada de 206.000 habitantes, entre los que sobresalen los centros poblados de Chiquinquirá y Ubaté; la cuenca presenta una intensa actividad agropecuaria y minera que se desarrolla alrededor del sistema de origen lacustre conformado por las lagunas de Fúquene, Cucunubá, Palacio y Suesca (CAR, 2011).

2.2.7 Cuenca Río Blanco

La cuenca del Río Blanco se localiza en la parte alta de la cuenca hidrográfica del Río Guayuriba, tiene una extensión de 3.540 km², drena parte de los departamentos de Cundinamarca y Meta. Esta cuenca de segundo orden es afluente del río Meta, que drena a la gran cuenca del Orinoco (Ministerio de Ambiente y Desarrollo Sostenible, 2005). En el

sector de la CAR la cuenca del Río Blanco tiene una superficie de 471 km², abarca parte de los municipios de La Calera y la zona rural de Bogotá D.C., caracterizada por ser de baja densidad poblacional y de gran importancia ambiental dada su localización en las estribaciones del páramo de Sumapaz y posee gran potencial hídrico (CAR, 2011).

2.2.8 Cuenca Río Mchetá

La cuenca del Río Mchetá hace parte de la cuenca del río Garagoa la cual se localiza al oriente del departamento de Cundinamarca y suroccidente de Boyacá con una extensión total de 2.508 km². En la zona de interés, la cuenca del Río Mchetá presenta un área de drenaje 509 km² correspondientes al 2.6% del área, abarcando cinco municipios con una población estimada de 24.000 personas, destacándose los centros poblados de Mchetá, Manta y Tibirita (CAR, 2011).

2.2.9 Cuenca Río Gchetá

En el municipio de Guatavita se localiza la cuenca del Río Gchetá, la cual hace parte de la cuenca del Río Guavio, situada al oriente del departamento de Cundinamarca con una extensión total de 1.716 km². En el área de interés, la cuenca está conformada por las cuencas de tercer orden de los Ríos Sueva, Lagunero y Amoladero, con una extensión de 97 km², en una zona alta de baja densidad poblacional y de gran importancia hidrológica (CAR, 2011).

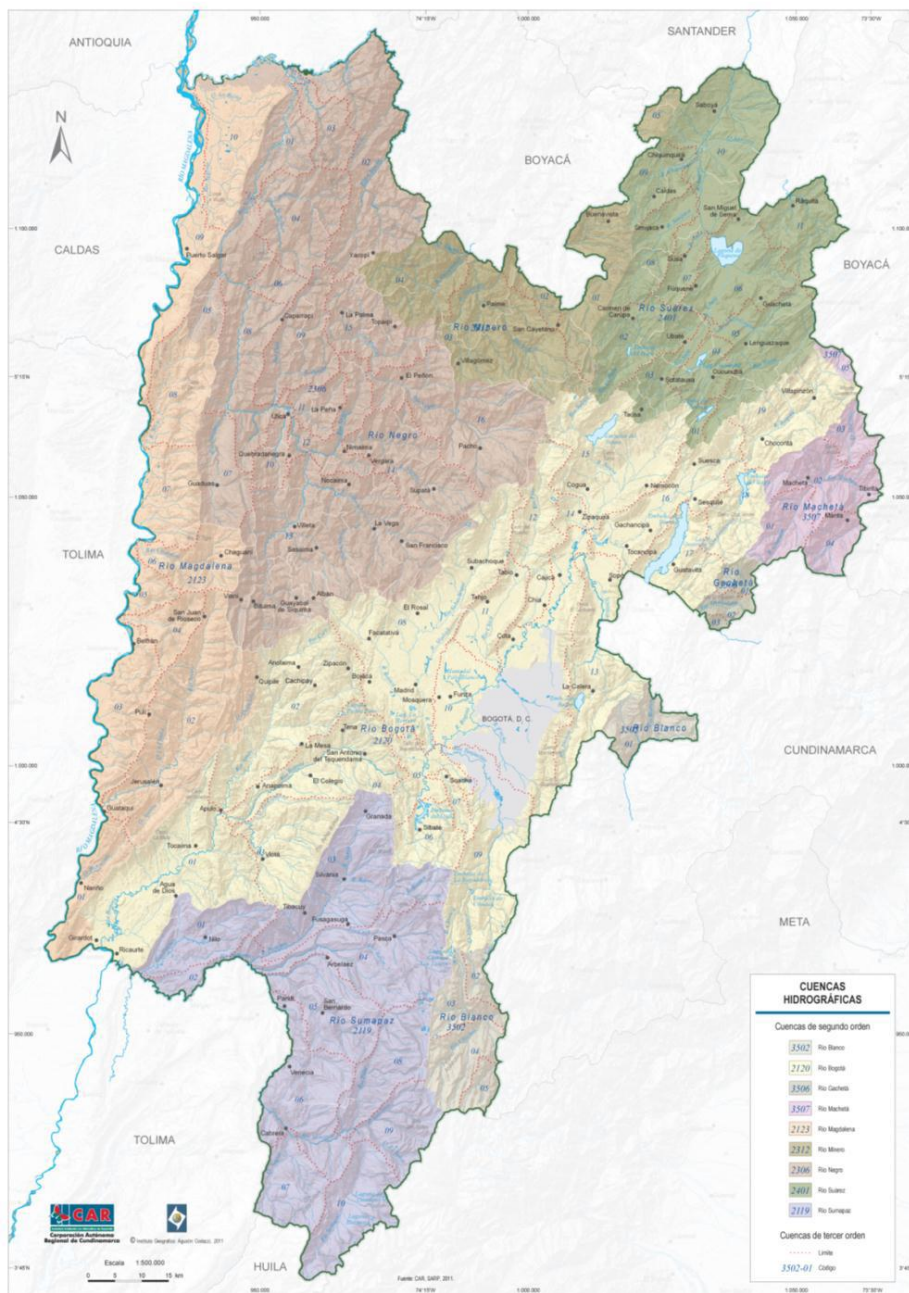


Figura 8. Cuencas de la región de estudio (CAR, 2012)

2.3 Red Hidroclimática

La red hidroclimática es una infraestructura instrumental para el monitoreo del recurso hídrico y del clima en una región. La Corporación Autónoma Regional de Cundinamarca (CAR) es la entidad encargada del manejo de esta red en el área de estudio del presente

De acuerdo con el Sistema de Información Climatológica e Hidrológica (SICLICA), del total de estaciones registradas, un 37% corresponde a estaciones Pluviométricas (PV), un 28% a estaciones Pluviográficas, un 15% a estaciones Climatológicas Principales (CP), un 14% a estaciones Climatológicas Ordinaria, y el 5% restante a estaciones automáticas y satelitales. En la Tabla No. 4 se muestra la clasificación de las estaciones.

Tipo de estación	Cantidad
Automática	6
Climatológica Ordinaria	26
Climatológica Principal	28
Pluviográfica	51
Pluviométrica	68
Satelitales	3

Tabla 4. Clasificación de las estaciones en la zona de estudio. Fuente: SILCA

La mayoría de las estaciones fueron instaladas en la década de los sesenta. Esta red está a cargo de la Corporación Autónoma Regional de Cundinamarca CAR desde 1961. En la Figura No. 10 se muestra el número de estaciones instaladas por año. Es importante resaltar que del total de estaciones mostradas, actualmente 47 han sido suspendidas, es decir, de estas estaciones sólo 135 se encuentran en funcionamiento.

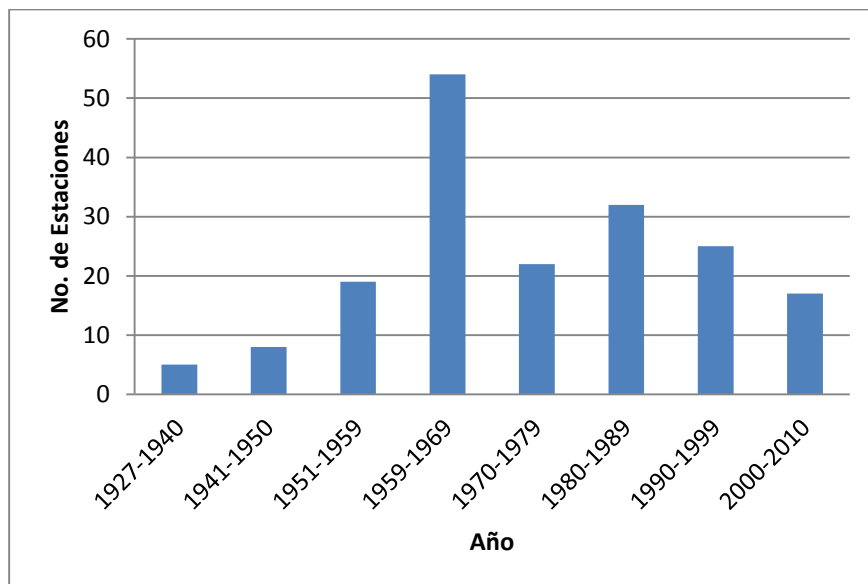


Figura 10. Estaciones instaladas por año

En la Figura No. 11 se muestra la longitud en años de la serie histórica del total de las estaciones que se conserva registro, comparado con la serie de las estaciones que actualmente se encuentran en funcionamiento. Se deduce que la mayor parte de las estaciones poseen serie histórica de más de 15 años; sin embargo, conviene aclarar que dentro de esa serie, en ciertos casos, faltan algunos meses. Éste será un problema que se enfrentará en la primera fase del proyecto, puesto que se deberá seleccionar un periodo en el que la mayor parte de las estaciones posean la mayor cantidad de datos.

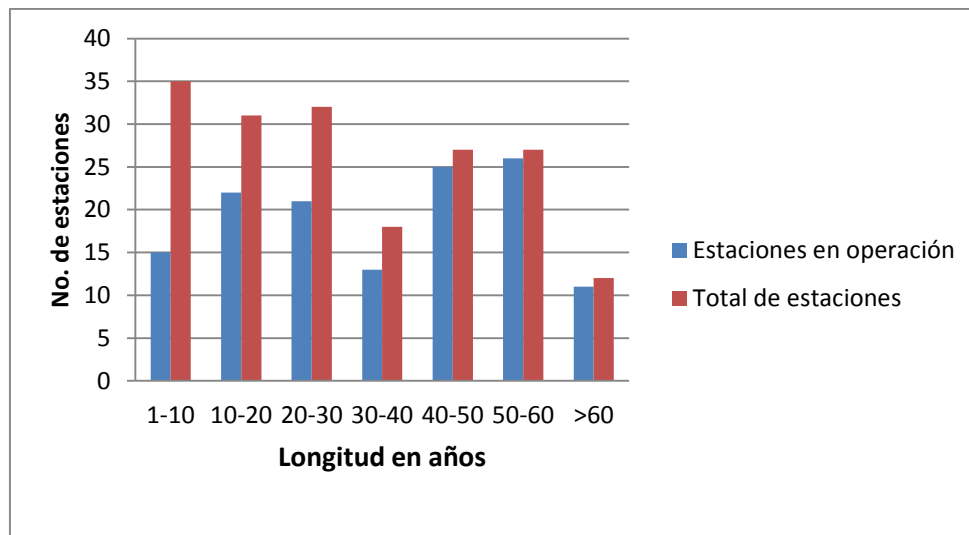


Figura 11. Longitud de la serie histórica

2.4 Selección de las estaciones para el estudio

Como se muestra en la Figura No. 11, el registro de la precipitación en las estaciones estudiadas no es homogéneo en cuanto a la cantidad de información; en algunos casos, la serie histórica es menor a diez años, mientras que en otros casos supera los 60 años de registro. Otra dificultad consiste en que la serie histórica tiene algunos vacíos en los datos mensuales, como se señaló anteriormente.

En este breve apartado se mostrará el procedimiento que llevó a la selección de las estaciones y del rango de años que se emplearán en la realización del proyecto. El objetivo, en este sentido, es seleccionar las estaciones que posean más del 80% de los

datos mensuales en un periodo no menor a 20 años, siguiendo las recomendaciones de la literatura (Mishra & Coulibaly, 2009).

Se enumeraron los meses, empezando por el más antiguo de la serie hasta el más reciente y verificó, uno a uno, cuáles de las estaciones poseían información sobre la precipitación en este mes en particular, se representó gráficamente esta información, como se muestra en la Figura No. 12. En ésta se identificó el periodo en el que el mayor número de estaciones posee al menos el 80% de los datos mensuales; este periodo está comprendido entre 1986 y 2008. Luego se verificó cuáles estaciones cumplían este criterio dentro del periodo seleccionado. En la Tabla No. 5 se muestran las 80 estaciones seleccionadas.

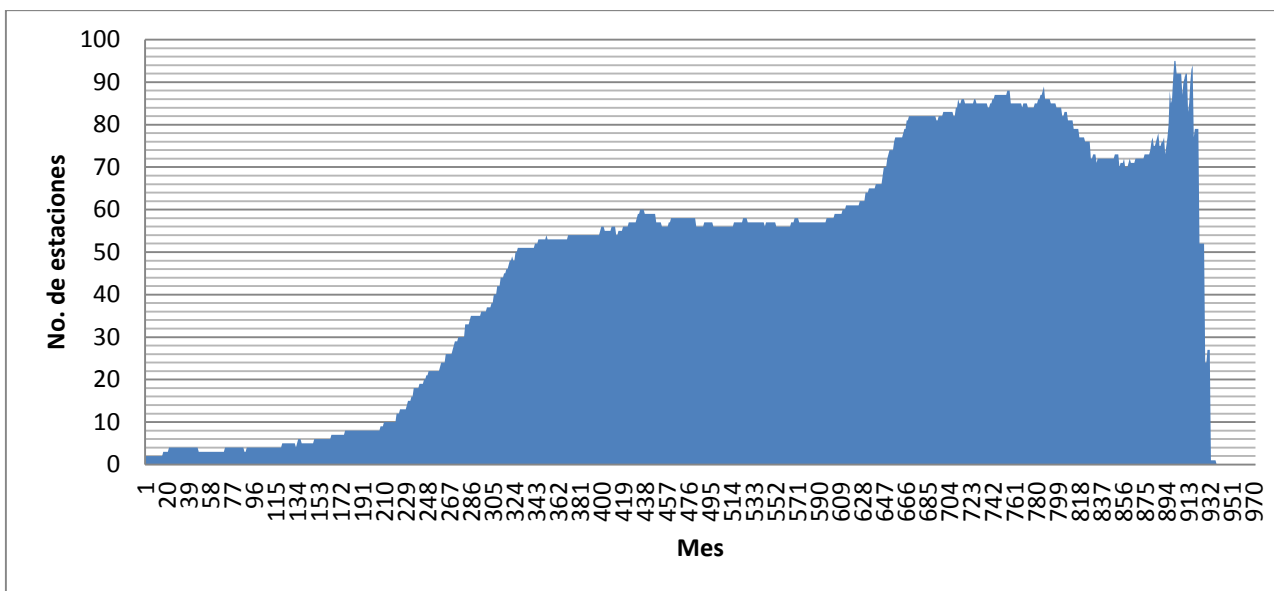


Figura 12. Estaciones con más del 80% de datos mensuales



No.	Código Estación	Nombre Estación	Fecha Instalación	Registro	
				Año de Inicio	Año de finalización
1	2120074	Zipaquirá	02/01/1960	1932	2012
2	2120637	Violetas Las	06/01/1989	1989	2012
3	2120640	Victoria La	12/01/1986	1987	2012
4	2120026	Ventalarga	06/01/1947	1947	2012
5	2120629	Venecia	12/01/1956	1957	2012
6	2120055	Unión La	01/01/1960	1960	2012
7	2401039	Triángulo El	01/01/1963	1963	2012
8	2401044	Tres esquinas	12/01/1965	1966	2012
9	2120077	Torca	02/01/1960	1970	2012
10	2120069	Tesoro	08/01/1931	1931	2012
11	2120565	Tabio	07/01/1970	1970	2012
12	2401029	Sutamarchán	05/01/1960	1960	2012
13	2120103	Sta Teresa	09/01/1970	1971	2012
14	2401046	Sta Sofía	12/01/1965	1965	2012
15	2312012	Sta Rita	04/01/1961	1961	2012
16	2120176	Sta Isabel	05/01/1968	1969	2012
17	2120136	Sta Inés	03/01/1980	1980	2012
18	2401037	Socotá	12/01/1962	1963	2012
19	2401513	Simijacá	11/01/1984	1958	2012
20	2401531	San Miguel de Sema	04/01/1961	1961	2012
21	2120080	San José	10/01/1960	1960	2012
22	2120172	San Jorge	04/01/1960	1960	2009
23	2120180	San Gregorio	02/01/1989	1989	2012
24	2120088	Salitre El	04/01/1955	1955	2012
25	2401538	Esclusa Tolón	05/01/1963	1945	2011
26	3502017	Puentes Los	01/01/1966	1966	1969
27	2401038	Puente El	12/01/1962	1963	2012
28	2120557	Primavera La	06/01/1965	1963	2012
29	2120044	Pradera La	09/01/1951	1951	2012
30	2123502	Placer El	12/01/1999	1999	2012
31	2312019	Pinos Los	12/01/1972	1973	2012
32	2401027	Pino El	03/01/1960	1960	2012
33	2120156	Picota La	06/01/1980	1981	2012
34	2120182	Peñas Blancas	02/01/1988	1988	2012
35	2120178	Pedro Palo	09/01/1988	1988	2012
36	2120134	Parque Sopó	03/01/1980	1980	2012
37	3509504	Olarte El	02/01/1984	1982	1997



No.	Código Estación	Nombre Estación	Fecha Instalación	Registro	
				Año de Inicio	Año de finalización
38	2120195	Montecillos	03/01/1996	1996	2012
39	2401036	Monserate	12/01/1962	1963	2012
40	2120646	Mesitas	09/01/1990	1990	2012
41	2120186	María La	06/01/1992	1993	2012
42	2120071	Margaritas Las	08/01/1959	1959	2012
43	2306039	Los Tiestos	04/01/2002	2007	2012
44	2120133	Local El	06/01/1980	1980	2012
45	2120635	Lagos Los	07/01/1988	1988	2000
46	2120184	Lag del Indio	04/01/1989	1989	2012
47	2401069	Juan Pablo II	09/01/1989	2009	2012
48	3509009	Hato Viejo el Rancho	09/01/1989	1990	1997
49	2401059	Hato No 8 El	10/01/1972	1973	1998
50	2401055	Hato No 4 El	10/01/1972	1973	1998
51	2401054	Hato No 3 El	10/01/1972	1973	1998
52	2401536	Hato el Hda.	03/01/1999	1998	2011
53	3502020	Gutiérrez	01/01/1966	1966	1969
54	2120214	Guerrero	09/01/1997	1998	2012
55	2401533	Guachetá	03/01/2000	2000	2010
56	2120166	Fute El	09/01/1959	1959	2012
57	2120647	Esperanza La	12/01/1992	1993	2012
58	2401518	Esclusa Tolón	05/01/1963	1945	2010
59	2401043	Esclusa Merchán	10/01/1965	1962	2012
60	2120641	Esc Samper Madrid	03/01/1987	1987	2012
61	2120058	Edificio CAR	17/03/2011	2008	2012
62	2120115	Edif. Manuel Mejía	01/01/1981	1981	2003
63	2119515	Dorado El	03/01/2000	2003	2007
64	2120194	Choche El	03/01/1996	1996	2008
65	2120089	Chisacá	05/01/1963	1946	1955
66	2120175	Charquito El	01/01/1988	1988	1990
67	2120213	Cedral El	12/01/1997	1998	2012
68	2401515	Carrizal	04/01/1960	1960	2012
69	2306519	Caparrapí	06/01/2006	2010	2011
70	2401026	Candelaria La	03/01/1960	1960	2012
71	2120173	Campobello	07/01/1985	1986	2012
72	2120552	Cachipay	17/03/2011	2009	2011
73	2120085	Bosque El	12/01/1962	1963	2012
74	2120661	Boquemonte	09/01/2010	2010	2012



No.	Código Estación	Nombre Estación	Fecha Instalación	Registro	
				Año de Inicio	Año de finalización
75	2120187	Bombas Sesquilé	01/01/1984	1984	2012
76	2119046	Batán	07/01/1997	1998	2012
77	2120633	Barrancas	08/01/1966	1966	2012
78	2401514	Balsa La	04/01/1960	1960	1963
79	2120644	Argelia	12/01/1991	1991	2012
80	3502016	Alto Caicedo	07/01/1966	1966	1969

Tabla 5. Estaciones seleccionadas para el estudio

Capítulo 3

Metodología

Este capítulo contiene los aspectos metodológicos establecidos para el cumplimiento de los objetivos. Abarca dos fases, en la primera se empleó el concepto de redes neuronales artificiales para la clasificación de las estaciones pluviométricas de la zona de estudio y, en la segunda, se evaluó el desempeño de esta red. A continuación se describen cada una de estas fases. Este proyecto de investigación fue de carácter cuantitativo, en el cual se sigue el método científico para abordar el problema planteado. En la figura No. 13 se muestra de manera general el procedimiento a desarrollar.

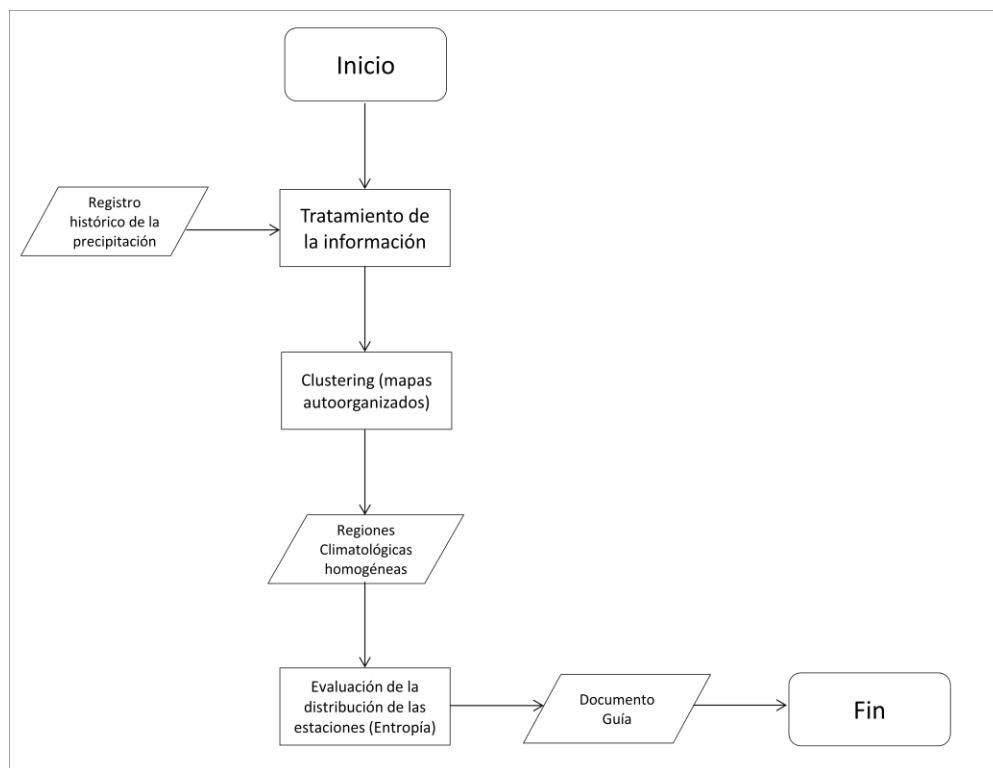


Figura 13. Esquema del procedimiento

3.1 Fase 1: Clasificación las estaciones pluviométricas de la región de estudio aplicando redes neuronales artificiales.

3.1.1 Recolección de la información

En esta etapa se gestionó con la Corporación Autónoma Regional de Cundinamarca (CAR) la información sobre la red climatológica de su área de jurisdicción. Los datos de interés en esta etapa fueron la ubicación de la estación (coordenadas), su estado (en operación o fuera de operación) y su registro histórico (precipitación mensual). En el Anexo No. 1 se muestran los datos de todas las estaciones del área de estudio.

3.1.2 Tratamiento de la Información

Teniendo en cuenta que las estaciones de la red analizada han entrado en funcionamiento en momentos distintos, se seleccionó un período en el cual la mayor cantidad posible de estaciones posea información y de éstas se seleccionaron las que al menos tengan el 80% de los datos de precipitación mensual en el periodo establecido (Mishra & Coulibaly, 2009). Las estaciones seleccionadas se muestran en la Tabla No. 5.

3.1.3 Clustering

Teniendo en cuenta que el método de mapas auto-organizados (SOM) permite identificar las regiones homogéneas con más precisión que dos de los métodos de clasificación más usados, *K-means* y el método de Ward (Gwo-Fong Lin, 2005), se empleó para efectuar dicha clasificación, la cual se realizó mediante el aplicativo Neural Network Toolbox del software Matlab.

3.1.3.1 Normalización de las variables

Las variables de entrada para la clasificación fueron: latitud (m), longitud (m), elevación (m), precipitación media anual (mm), desviación estándar de la precipitación media anual y precipitación media mensual (mm) de cada una de las estaciones. Para evitar que la diferencia de la escala de las variables afecte la clasificación, estas fueron transformadas de modo que sus rangos resultaran comparables (Gwo-Fong Lin, 2005). La Tabla No. 6 muestra la manera cómo se aplicó la transformación de tal manera que las variables tomaran valores entre 0 y 1.

Característica	Transformación
Latitud (m)	$Y=(X-X_{\min})/(X_{\max}-X_{\min})$
Longitud (m)	$Y=(X-X_{\min})/(X_{\max}-X_{\min})$
Elevación (m)	$Y=X/X_{\max}$
Precipitación Media Anual (mm)	$Y=X/X_{\max}$
Desviación de la Precipitación Media Anual (mm)	$Y=X/X_{\max}$
Precipitación media mensual	$Y=X/X_{\max}$

Tabla 6. Transformación de las variables (Gwo-Fong Lin, 2005)

3.1.3.2 Creación del modelo de Red

Para diferenciar los grupos de patrones adecuadamente, se recomienda emplear arreglos de neuronas en la capa de salida lo más grandes posible. Sin embargo, es importante tener en cuenta que si el número de neuronas en la capa de salida es muy grande, el modelo puede sobre-entrenarse y resaltar las diferencias entre cada uno de los patrones, arrojando igual número de grupos como de patrones (González, 2012). Teniendo en cuenta lo anterior, se definieron tres escenarios con distintos número de neuronas en la capa de salida. El número de iteraciones está dado, de acuerdo a recomendaciones de la literatura, por el número de neuronas de la capa de salida multiplicado por 500 (González, 2012). En la Tabla No. 7 se detallan las características de cada uno de estos escenarios. Con ayuda del Toolbox en Redes Neuronales Artificiales de Matlab, se desarrolló un aplicativo computacional para clasificar las estaciones pluviométricas de la Corporación Autónoma Regional de Cundinamarca, teniendo en cuenta las tipologías definidas.

	Tipología No. 1	Tipología No. 2	Tipología No. 3
Neuronas en la capa de salida	100	400	900
Iteraciones	50000	200000	450000

Tabla 7. Características de las tipologías

- **Tipología No. 1:** En este caso se trata de un problema de clasificación con un modelo de red de Kohonen de 100 neuronas distribuidas en un arreglo de 10 filas por 10 columnas. El número de iteraciones corresponde al número de neuronas de la capa de salida, multiplicado por 500, es decir, en este caso se harán 50000 iteraciones. Se eligió una topología hexagonal (neuronas de forma hexagonal

también), así las neuronas que no se encuentran en los bordes de la capa de Kohonen tienen 6 neuronas vecinas con las cuales se conectan virtualmente.

- **Tipología No. 2:** En este caso se trata de un problema de clasificación con un modelo de red de Kohonen de 400 neuronas distribuidas en un arreglo de 20 filas por 20 columnas. En este caso se harán 200000 iteraciones. Se eligió una topología hexagonal (neuronas de forma hexagonal también).
- **Tipología No. 3:** En este caso se trata de un problema de clasificación con un modelo de red de Kohonen de 900 neuronas distribuidas en un arreglo de 30 filas por 30 columnas. En este caso se harán 450000 iteraciones. Se tomó una topología hexagonal (neuronas de forma hexagonal también).

3.1.3.3 Visualización de resultados

El método de clasificación empleado ofrece la ventaja de mostrar los resultados sobre mapas bidimensionales sin importar el número de variables incluidas. En este sentido, para la visualización e interpretación de los resultados se destacan dos mapas particularmente: el de Hits y el de distancias entre neuronas (Mapa U-Matrix).

En el mapa de Hits se identifican las neuronas ganadoras. En éste, el número dentro de cada neurona indica la cantidad de estaciones que representa, es decir, la cantidad de victorias de cada neurona. Las neuronas con valor cero corresponden a las que no representan ningún patrón (estación). En la Figura No. 14 se muestra un ejemplo de Mapa de Hits.

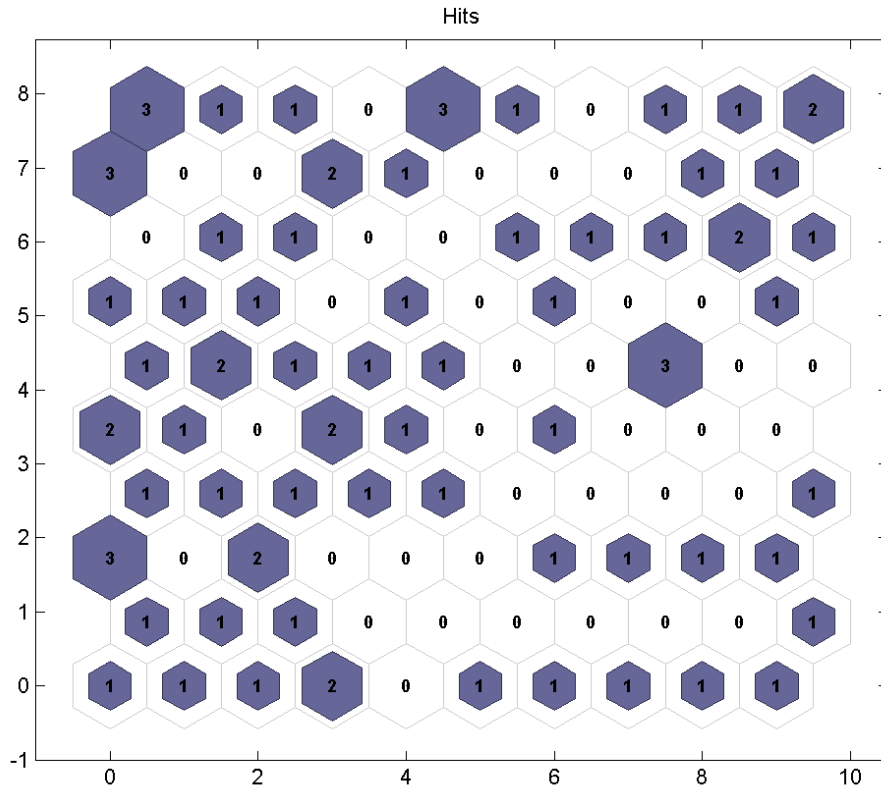


Figura 14. Ejemplo de Mapa de Hits

El Mapa de Distancia entre Neuronas, también conocido como *U-Matrix*, permite visualizar qué tan diferente es una neurona de otra. Los colores oscuros indican una gran diferencia mientras que los colores claros indican similitud entre neuronas y por tanto entre estaciones. Este gráfico permite identificar los grupos en los que se clasifica la información. En la Figura No. 15 se muestra un ejemplo de Mapa de Distancia entre Neuronas.

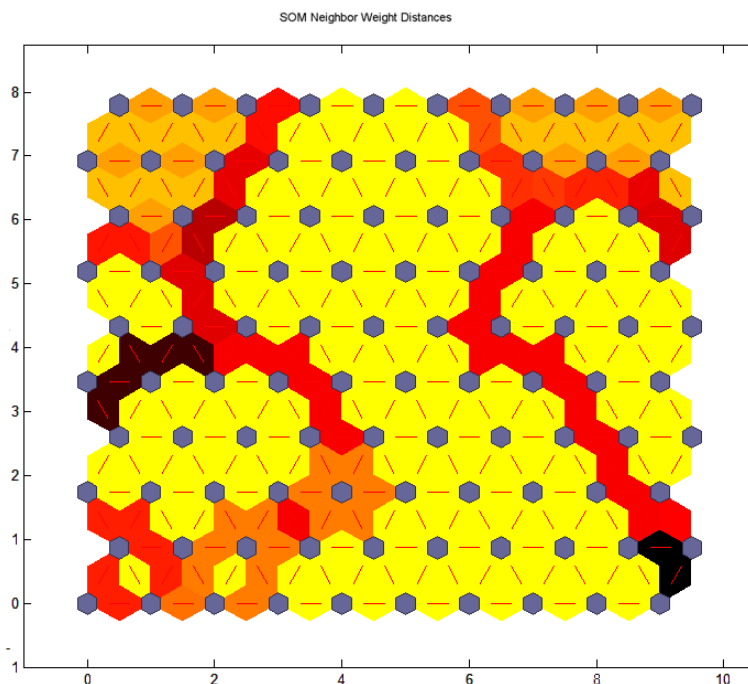


Figura 15. Ejemplo de Mapa de Distancia entre Neuronas.

Una vez identificado el número de grupos que se forman, se procede a asociar a cada neurona con la estación a la que representa, para ello se cuenta con la Matriz de Hits, en las filas se encuentran las estaciones y en las columnas las neuronas. A cada estación corresponde una neurona ganadora que se indica con el número 1. De esta forma se establecen los grupos de estaciones.

3.2 Fase 2: Evaluación del desempeño de la red pluviométrica de la región Bogotá-Cundinamarca.

3.2.1 Entropía de la información

La entropía, tal como se define en teoría de la información, es una medida de la incertidumbre de un resultado particular en un proceso aleatorio (Shanon, 1948). En esta sección se estudió la distribución de la información en cada uno de los grupos formados siguiendo el procedimiento que a continuación se describe:

1. Se estableció la longitud de la serie histórica de datos en cada caso, teniendo en cuenta la recomendación de que al menos sea un periodo no menor a 20 años y que los datos faltantes no excedieran el 20% de acuerdo a la recomendación de la

literatura (Mishra & Coulibaly, 2009). Este mismo criterio se aplicó en la fase anterior del proyecto.

2. Se calculó la entropía marginal para cada estación mediante la expresión:

$$H = - \sum_{k=1}^k p(x_k) \text{Log}[p(x_k)]$$

Donde k es un intervalo de datos discretos, x_k es un resultado correspondiente al intervalo de k, y $p(x_k)$ es la probabilidad de x_k .

3. Para cada estación, se realizó una estimación de la serie por regresión lineal múltiple a partir de los datos de las estaciones que conforman su mismo grupo.

$$Y = X\beta + \varepsilon$$

4. Se calculó la entropía marginal de la serie estimada.

$$H = - \sum_{k=1}^k p(x_k) \text{Log}[p(x_k)]$$

5. Posteriormente se calculó la entropía marginal conjunta entre los valores reales y estimados de la serie de cada estación.

$$H(X, Y) = - \sum_{k=1}^K \sum_{L=1}^L p(x_k y_L) \text{Log}[p(x_k y_L)]$$

6. Finalmente, se calcula la Transinformación de cada estación a partir de los valores encontrados previamente. La Transinformación o información mutua es una medida de la información contenida en un proceso sobre otro proceso (Mishra & Coulibaly, 2009). En este caso corresponde a la información pluviométrica de una estación contenida en las demás de su propio grupo.

$$T(X, Y) = H(X) + H(Y) - H(X, Y)$$

$H(X)$ corresponde a la entropía marginal de los datos reales, $H(Y)$ corresponde a la entropía marginal para la serie simulada y $H(X, Y)$ es la entropía conjunta.

Una vez obtenido el valor de la Transinformación se establece un ranking para identificar las estaciones con mayor y menor Transinformación, a partir de este se dan las recomendaciones para el rediseño de la red.

3.2.1 Recomendaciones para el rediseño de la red

La base fundamental en el diseño de redes de monitoreo basado el enfoque de la entropía es que las estaciones deben tener la menor transinformación como sea posible, lo que significa que las estaciones deben ser independientes entre sí. Por lo tanto los resultados de la transinformación obtenidos en el paso anterior, se interpretan así: Valores bajos de transinformación indican estaciones que comparten muy poca información, por lo tanto son más independientes. Cuando los valores de la transinformación son más grandes, las estaciones son más dependientes, por lo tanto, puede que no exista la necesidad de más estaciones en esa área. Se recomendaría instalar estaciones pluviométricas adicionales donde se obtienen valores de transinformación cercanos a 0. En la Tabla No. 8 se muestran la clasificación de la transinformación de acuerdo a su valor. Este criterio fue tenido en cuenta para dar las correspondientes recomendaciones en el rediseño de la red pluviométrica.

En el Anexo 4 se presenta el documento guía con recomendaciones generales para la CAR.

Transinformación	Concepto
0-0.5	Alto déficit
0.5-1.0	Déficit
1.0-1.5	Aceptable
1.5-2.0	Por encima del promedio
>2.0	Exceso

Tabla 8. Clasificación de la Transinformación (Mishra & Coulibaly, 2009)

Capítulo 4

Clasificación de las estaciones

4.1 Variables de entrada

Como se ha señalado en el capítulo anterior, las variables de entrada en el proceso de clasificación fueron latitud (m), longitud (m), elevación (m), precipitación media anual (mm), desviación estándar de la precipitación media anual y precipitación media mensual (mm) de cada una de las estaciones. Es decir, un total de 17 variables para cada una de las 80 estaciones. En la Tabla No 9 se observan los valores de estas variables y su correspondiente normalización para las 10 primeras estaciones. El resto de información puede ser consultada en el Anexo No. 1.

	E1	E4	E8	E9	E10	E13	E15	E16	E18	E22	
Longitud	0,6582	0,1046	0,1765	0,6240	0,3406	0,4603	0,9768	0,7852	0,6414	0,4045	
Latitud	0,4806	0,0256	0,1008	0,5132	0,3561	0,3755	0,6668	0,7215	0,3214	0,3260	
Altitud	0,7636	0,1150	0,1093	0,8806	0,7688	0,7837	0,8053	0,9002	0,7417	0,7506	
Precipitación mensual	Enero	0,3197	0,4796	0,5404	0,2984	0,4240	0,3530	0,3445	0,2358	0,5877	0,3984
	Febrero	0,3681	0,5415	0,5799	0,3155	0,4837	0,4429	0,3097	0,2851	0,5078	0,5936
	Marzo	0,3438	0,5556	0,5726	0,4048	0,4610	0,4513	0,3644	0,3543	0,4520	0,5175
	Abril	0,2817	0,4985	0,4060	0,2883	0,2691	0,3447	0,2279	0,2684	0,3012	0,3315
	Mayo	0,3257	0,5299	0,4270	0,3777	0,2608	0,3329	0,4092	0,2680	0,2975	0,3670
	Junio	0,3446	0,2578	0,2759	0,4483	0,2356	0,3174	0,5403	0,2651	0,3430	0,4590
	Julio	0,4423	0,1983	0,2441	0,5927	0,2765	0,3589	0,6043	0,3653	0,4742	0,5490
	Agosto	0,2833	0,1837	0,2197	0,5957	0,2626	0,3128	0,4148	0,2476	0,3009	0,4912
	Septiembre	0,3041	0,4688	0,4408	0,3927	0,3009	0,3613	0,3561	0,2203	0,3655	0,4860
	Octubre	0,4040	0,3501	0,3811	0,3445	0,3192	0,3827	0,3469	0,3007	0,3461	0,4283
	Noviembre	0,2871	0,3532	0,4962	0,3086	0,3615	0,3209	0,2984	0,2276	0,3648	0,3564
Diciembre	0,2593	0,3984	0,4038	0,2600	0,3633	0,4133	0,3635	0,3224	0,3810	0,4007	
Precipitación media anual	0,3286	0,4104	0,4169	0,3749	0,3246	0,3627	0,3683	0,2782	0,3720	0,4310	
Desviación de la p media anual	0,3377	0,5635	0,4935	0,3289	0,3004	0,3365	0,3209	0,2699	0,2519	0,3057	

Tabla 9. Variables de entrada normalizadas

4.2 Proceso de clasificación

Con ayuda del Toolbox de Redes Neuronales Artificiales de Matlab, se desarrolló un aplicativo computacional para clasificar las estaciones pluviométricas de la Corporación Autónoma Regional de Cundinamarca. A continuación se presentan los resultados obtenidos en los escenarios señalados.

4.2.1 Tipología No. 1

Para la visualización de los resultados, se cuenta con el mapa de Hits, donde se identifican las neuronas ganadoras (el número dentro de cada neurona indica la cantidad de estaciones que representa). El Mapa de Hits de esta tipología se muestra en la Figura No. 16

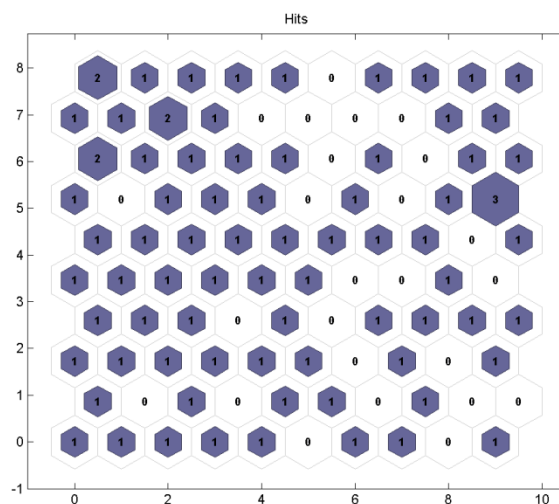


Figura 16. Mapa de Hits para la Tipología No. 1

El Mapa de Distancia entre Neuronas muestra qué tan diferente es una neurona de otra, a partir de lo cual se pueden identificar los grupos en los que se divide la información. En la Figura No. 17 se observa este mapa para la Tipología No.1, aquí se nota la presencia de zonas más oscuras que marcan la división entre los conjuntos. Se observa un total de 13 grupos con distinto número de neuronas. Para esta clasificación se han tenido en cuenta las líneas de mayor intensidad. En la Figura No. 18 se muestran de manera más clara la conformación de los grupos.

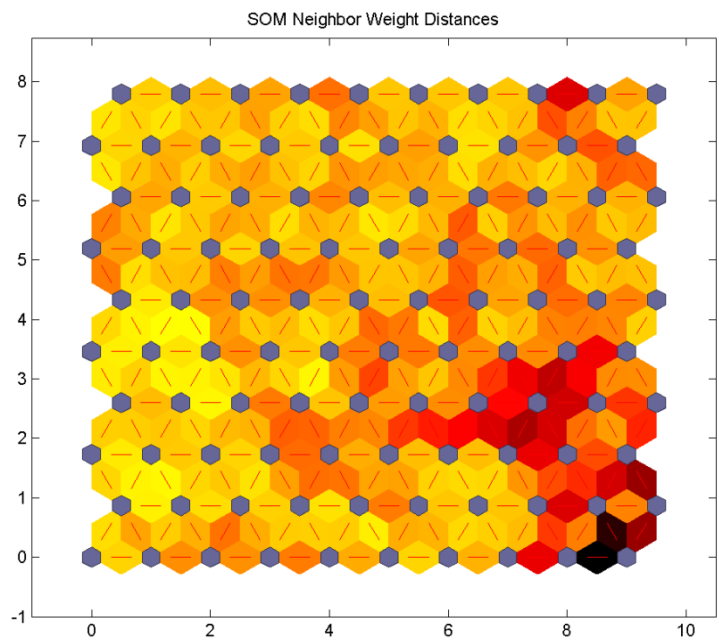


Figura 17. Mapa de Distancia entre Neuronas Tipología No. 1

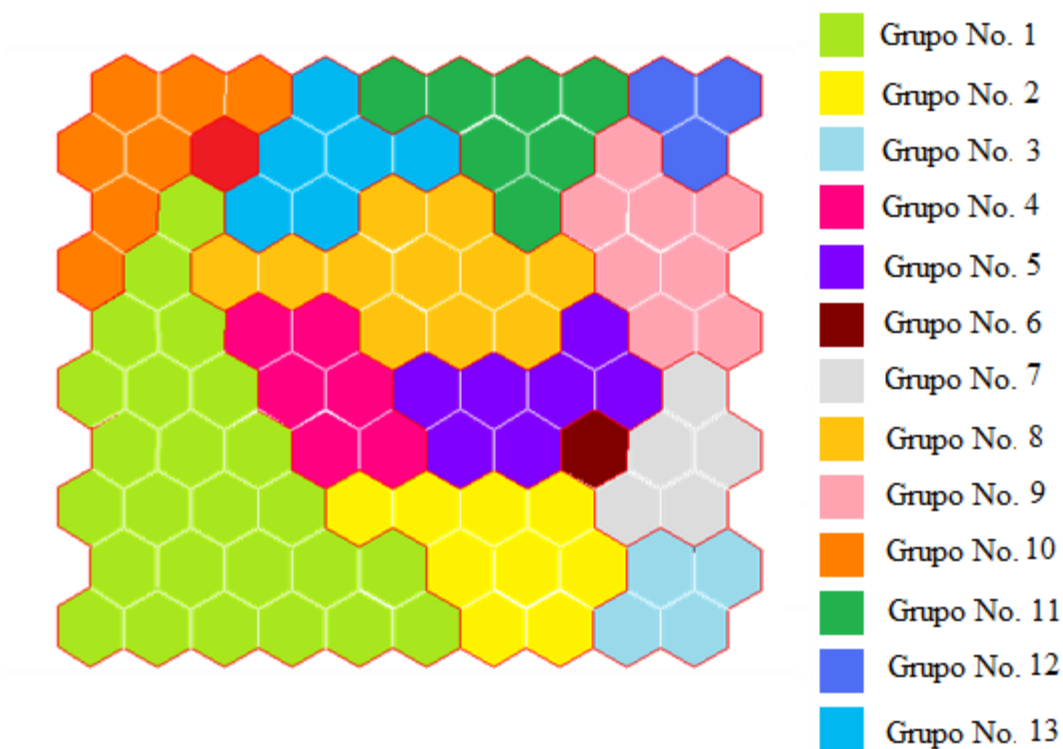


Figura 18. Formación de grupos en la tipología No. 1

El siguiente paso es asociar cada neurona con la estación a la que representa, para ello se cuenta con la Matriz de Hits, compuesta por 80 filas y 100 columnas. En las filas se encuentran las estaciones y en las columnas las neuronas. Se puede observar, a manera de ejemplo, en la Tabla No. 10, que en a la Estación E1 le corresponde la Neurona número 74, mientras que a la Estación E4 le correspondió la Neurona número 90. En el Anexo 2a se muestra la matriz de Hits completa.

	E1	E4
1	0	0
2	0	0
3	0	0
73	0	0
74	1	0
75	0	0
89	0	0
90	0	1
91	0	0
99	0	0
100	0	0

Tabla 10. Matriz de Hits

En la Tabla No. 11 se presenta un resumen de la correspondencia de neuronas y grupos para cada estación. Así mismo, en la Figura No. 19 se presenta la distribución geográfica de los grupos.

Estación	Neurona	Grupo
E1	74	13
E4	90	12
E8	99	12
E9	75	8
E10	98	11
E13	97	11
E15	35	4
E16	23	1
E18	58	8
E22	77	11
E25	72	1
E27	3	1
E28	56	8
E29	28	2
E31	7	2
E32	61	10
E33	83	13
E36	15	1
E37	1	1
E40	18	2
E41	55	8
E42	93	10
E44	79	9
E45	65	8
E49	83	13
E52	4	1
E53	94	13

Estación	Neurona	Grupo
E54	73	13
E55	100	12
E60	5	1
E61	21	1
E62	82	10
E63	39	7
E64	69	9
E65	41	1
E67	91	10
E71	11	1
E76	26	2
E77	70	9
E78	60	9
E80	38	6
E81	95	11
E85	46	5
E86	52	1
E88	30	7
E89	42	1
E91	54	4
E96	25	2
E97	37	5
E101	31	1
E102	13	1
E105	44	4
E108	63	8
E109	49	5

Estación	Neurona	Grupo
E116	81	10
E118	53	4
E122	16	2
E123	40	7
E124	80	9
E126	24	1
E129	92	10
E130	70	9
E131	45	4
E138	51	1
E140	10	3
E142	57	8
E144	43	1
E147	71	10
E149	2	1
E150	22	1
E151	67	8
E154	71	10
E156	33	1
E158	8	2
E159	89	9
E161	84	13
E162	91	10
E163	70	9
E166	32	1
E174	64	8

Tabla 11. Distribución de estaciones por grupos para la Tipología No. 1

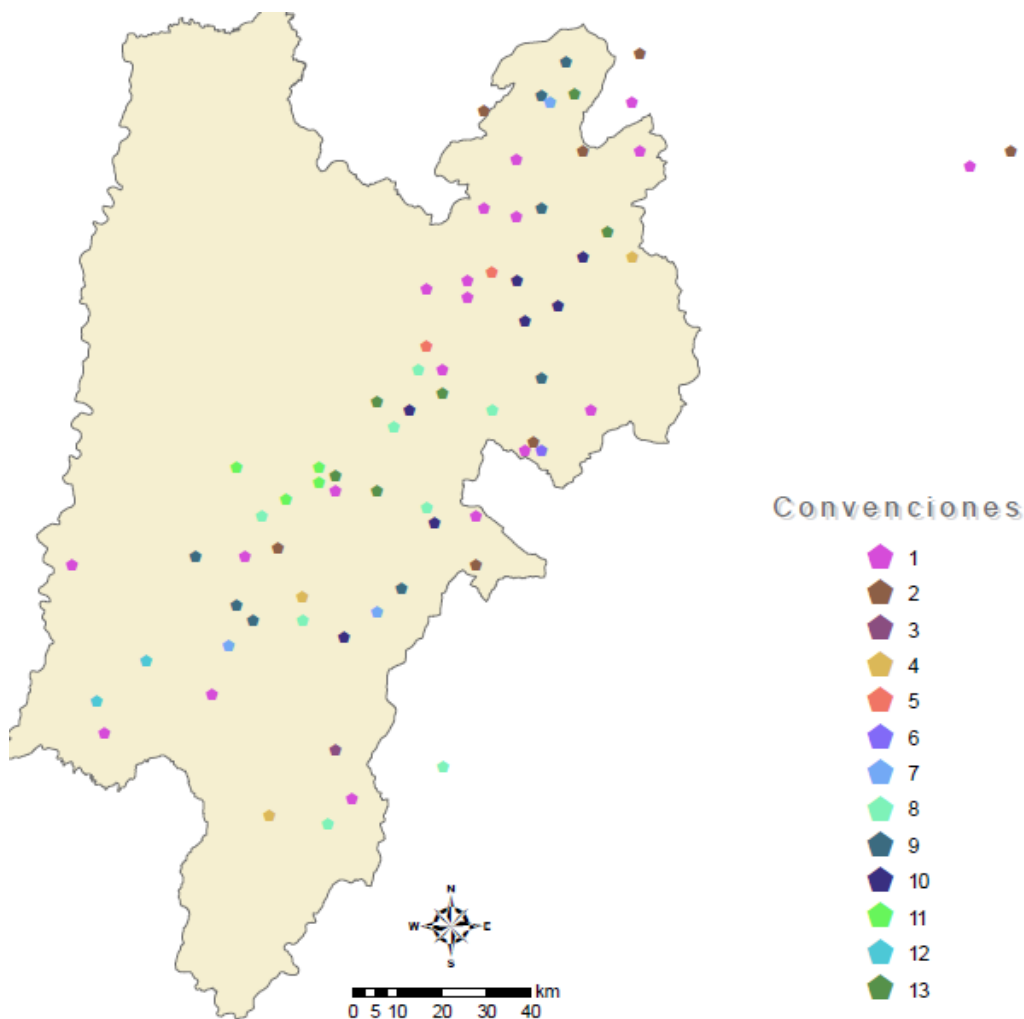


Figura 19. Distribución geográfica de los grupos para la Tipología No. 1

4.2.2 Tipología No. 2

Para la visualización de los resultados, se cuenta con el mapa de Hits, donde se identifican las neuronas ganadoras (el número dentro de cada neurona indica la cantidad de estaciones que representa). El Mapa de Hits de esta tipología se muestra en la Figura No. 20. Se observa que las neuronas ganadoras se encuentran más dispersas en relación a la tipología anterior, esto se debe al aumento del número de neuronas total de la red.

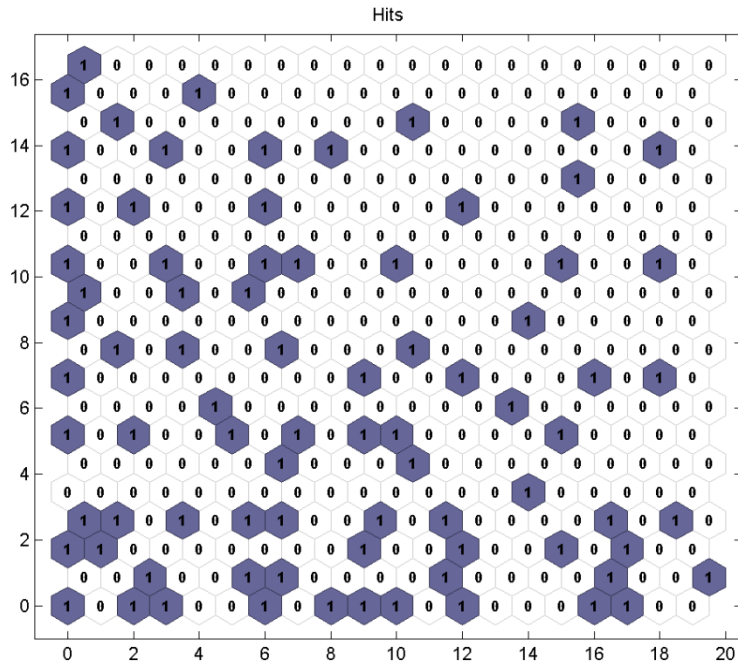


Figura 20. Mapa de Hits para la Tipología No. 2

En la Figura No. 21 se observa el Mapa de Distancia entre Distancia para la Tipología No.2, aquí se nota la presencia de zonas más oscuras que marcan la división entre los conjuntos. En este caso la demarcación es más notoria que en la Tipología No. 1. Se distinguen 50 grupos, algunos de estos quedarán conformados por neuronas no ganadoras, por lo que carecen de importancia para el estudio. En la Figura No. 22 se muestran de manera más clara la conformación de los grupos.

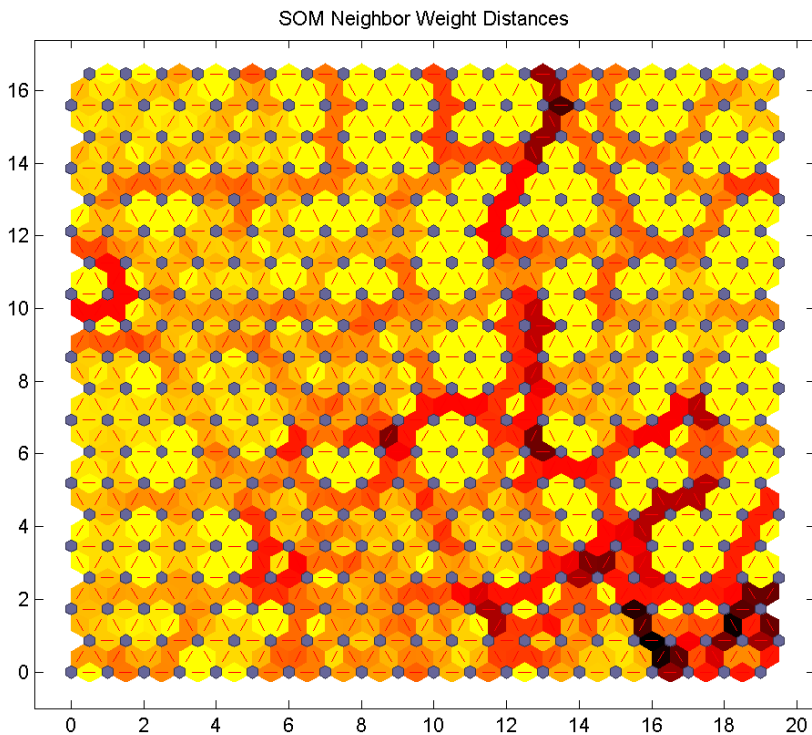


Figura 21. Mapa de Distancia entre Neuronas Tipología No. 2



Figura 22. Formación de grupos en la tipología No. 2

El siguiente paso es asociar cada neurona con la estación a la que representa, para ello se cuenta con la Matriz de Hits, compuesta por 80 filas y 400 columnas. En las filas se encuentran las estaciones y en las columnas las neuronas. Se puede observar, a manera de ejemplo, en la Tabla No. 12 que en a la Estación E1 le corresponde la Neurona número 226, mientras que a la Estación E4 le correspondió la Neurona número 339. En el Anexo 2b se muestra la matriz de Hits completa.

Neurona	E1	E4
1		
2		
225		
226	1	
227		
338		
339		1
399		
400		

Tabla 12. Matriz de Hits de la Tipología No. 2

En la Tabla No. 13 se presenta un resumen de la correspondencia de neuronas y grupos para cada estación. Así mismo, en la Figura No. 23 se presenta la distribución geográfica de los grupos.

Estación	Neurona	Grupo	Estación	Neurona	Grupo	Estación	Neurona	Grupo
E1	226	32	E53	329	46	E109	136	22
E4	339	49	E54	221	31	E116	361	44
E8	316	43	E55	356	48	E118	126	11
E9	72	14	E60	9	3	E122	32	3
E10	251	37	E61	41	11	E123	79	17
E13	191	33	E62	365	44	E124	259	39
E15	70	13	E63	58	15	E126	26	2
E16	64	11	E64	154	26	E129	324	45
E18	170	25	E65	61	11	E130	179	30
E22	95	21	E67	342	44	E131	66	12
E25	283	40	E71	23	6	E138	201	18
E27	1	1	E76	67	3	E140	40	10
E28	111	14	E77	256	38	E142	173	14
E29	56	4	E78	77	16	E144	182	18
E31	17	4	E80	18	5	E147	281	40
E32	241	34	E81	351	47	E149	27	3
E33	287	41	E85	128	19	E150	42	6
E36	10	3	E86	161	18	E151	131	20
E37	3	1	E88	37	9	E154	321	44
E40	13	4	E89	121	18	E156	184	23
E41	130	13	E91	107	12	E158	11	3
E42	327	45	E96	50	5	E159	215	27
E44	293	42	E97	53	7	E161	248	36
E45	187	24	E101	62	11	E162	381	44
E49	247	35	E102	4	2	E163	177	29
E52	7	3	E105	145	23	E166	123	18
			E108	244	35	E174	224	32

Tabla 13. Distribución de estaciones por grupos para la Tipología No. 2

Como se ha señalado algunos grupos resultaron sin ninguna estación, puesto que no poseían neuronas ganadoras, para facilitar la distinción de los grupos se reenumeran, omitiendo los que no poseen estaciones. En la tabla No. 14 se muestra la nueva distribución de las estaciones por grupos.



Estación	Grupo
E27	1
E37	1
E102	2
E126	2
E36	3
E52	3
E60	3
E76	3
E122	3
E149	3
E158	3
E29	4
E31	4
E40	4
E80	5
E96	5
E71	6
E150	6
E97	7
E88	8
E140	9
E16	10
E61	10
E65	10
E101	10
E118	10
E91	11

Estación	Grupo
E131	11
E15	12
E41	12
E9	13
E28	13
E142	13
E63	14
E78	15
E123	16
E86	17
E89	17
E138	17
E144	17
E166	17
E85	18
E151	19
E22	20
E109	21
E105	22
E156	22
E45	23
E18	24
E64	25
E159	26
E163	27
E130	28
E54	29

Estación	Grupo
E1	30
E174	30
E13	31
E32	32
E49	33
E108	33
E161	34
E10	35
E77	36
E124	37
E25	38
E147	38
E33	39
E44	40
E8	41
E62	42
E67	42
E116	42
E154	42
E162	42
E42	43
E129	43
E53	44
E81	45
E55	46
E4	47

Tabla 14. Grupos finales para la Tipología No.2

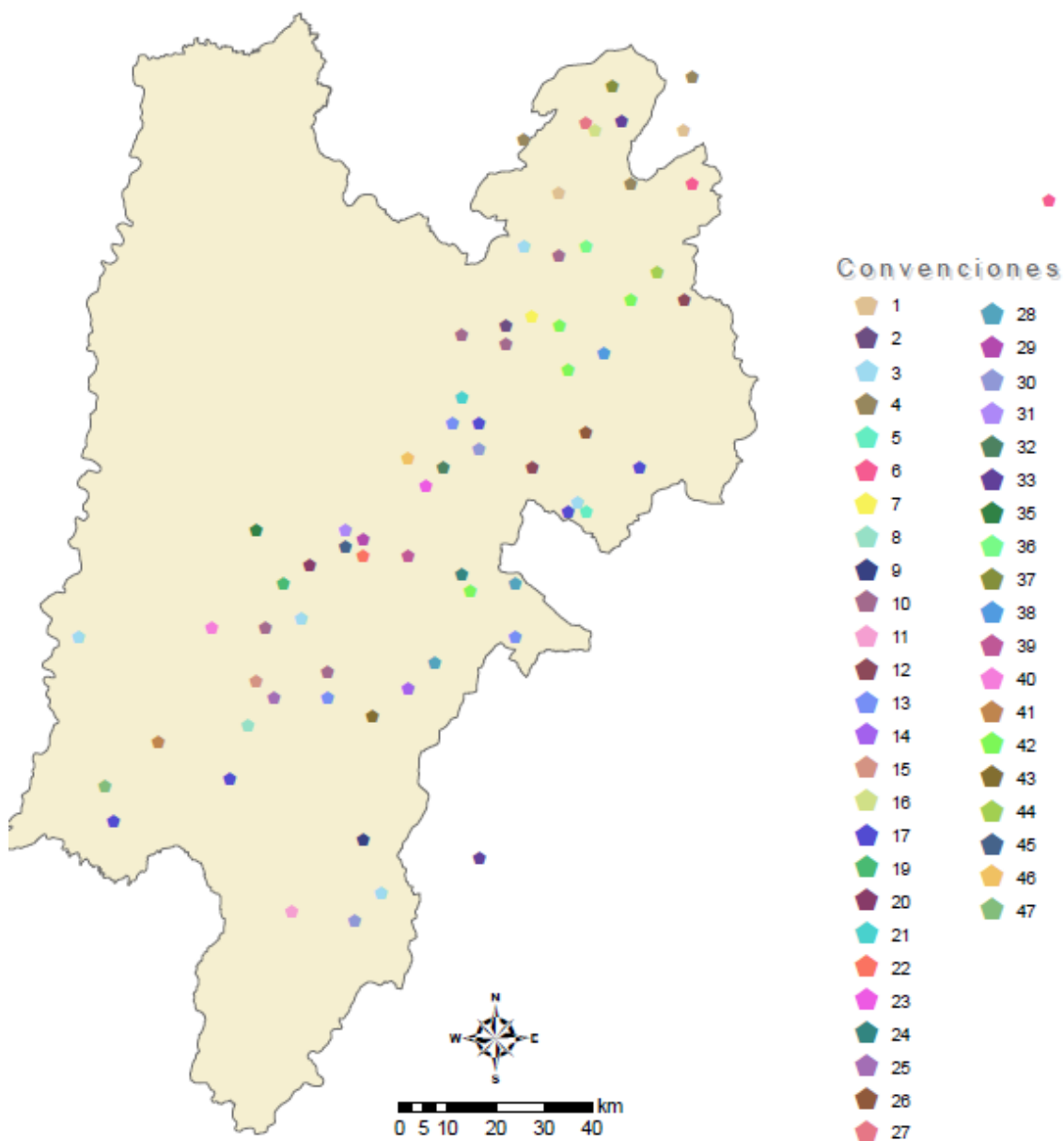


Figura 23. Distribución geográfica de los grupos para la Tipología No. 2

4.2.3 Tipología No. 3

El Mapa de Hits de esta tipología se muestra en la Figura No. 24. Se observa que las neuronas ganadoras se encuentran dispersas, como en el caso anterior, esto se debe al aumento del número de neuronas total de la red. Este hecho hace prever que los grupos estarán conformados por pocas estaciones.

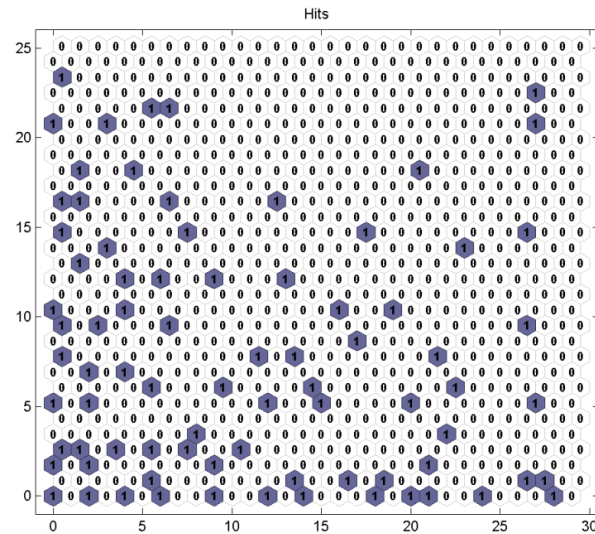


Figura 24. Mapa de Hits para la Tipología No. 3

En la Figura No. 25 se observa el Mapa de Distancia entre Distancia para la Tipología No.3, aquí se nota la presencia de zonas más oscuras que marcan la división entre los conjuntos. En este caso la demarcación es más notoria que en la Tipología No. 1. Se distinguen 65 grupos, algunos de estos quedarán conformados por neuronas no ganadoras, por lo que carecen de importancia para el estudio. En la Figura No. 26 se muestran de manera más clara la conformación de los grupos.

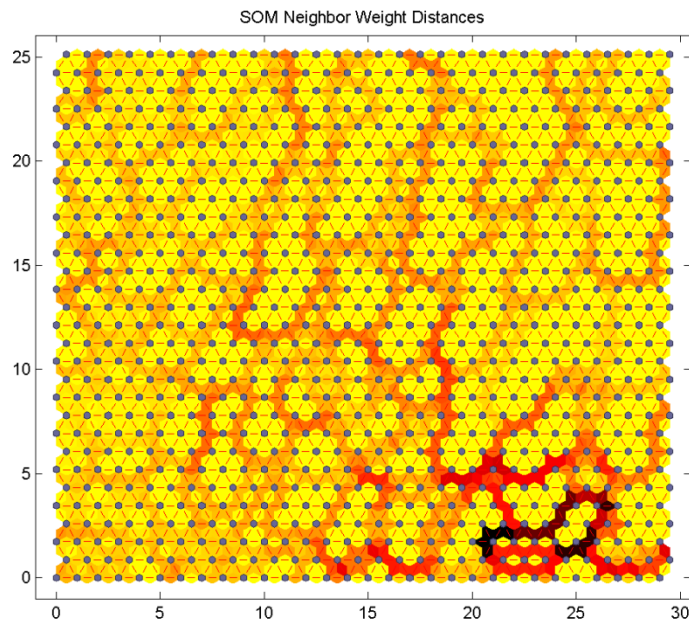


Figura 25. Mapa de Distancia entre Neuronas Tipología No. 3

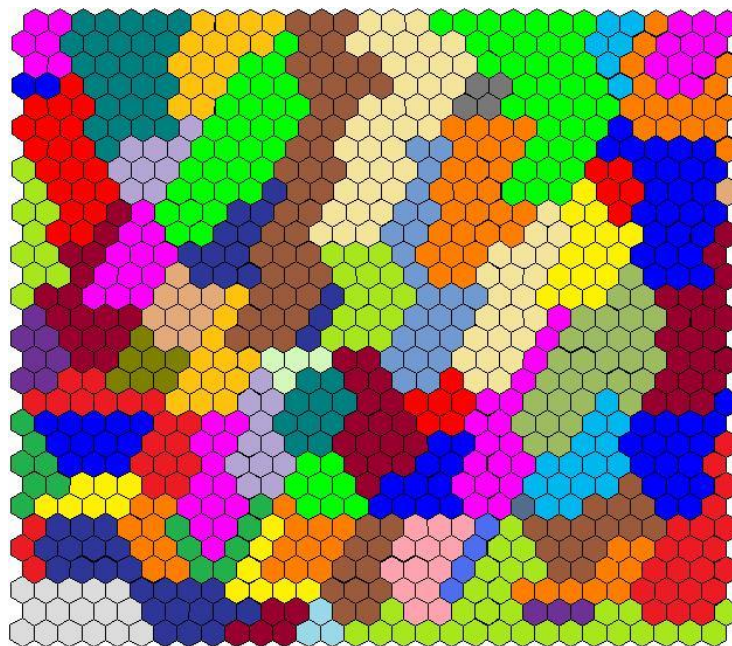


Figura 26. Formación de grupos en la tipología No. 3

Con la ayuda de la Matriz de Hits, compuesta por 80 filas y 900 columnas, se asociará a cada estación una neurona y su correspondiente grupo. En las filas se encuentran las estaciones y en las columnas las neuronas. Se puede observar, a manera de ejemplo, en la Tabla No. 15 que en a la Estación E1 le corresponde la Neurona número 427, mientras que a la Estación E4 le correspondió la Neurona número 748. En el Anexo 2c se muestra la matriz de Hits completa.

Neurona	E1	E4
1	0	0
2	0	0
427	1	0
428		
748	0	1
899	0	0
900	0	0

Tabla 15. Matriz de Hits de la Tipología No. 3

En la Tabla No. 16 se presenta un resumen de la correspondencia de neuronas y grupos para cada estación. Así mismo, en la Figura No. 27 se presenta la distribución geográfica de los grupos.

Estación	Neurona	Grupo
E1	427	38
E4	748	58
E8	537	47
E9	220	20
E10	583	52
E13	282	31
E15	70	11
E16	216	18
E18	434	42
E22	196	23
E25	425	36
E27	1	1
E28	284	22
E29	22	5
E31	21	5
E32	811	65
E33	484	50
E36	19	5
E37	63	1
E40	25	5
E41	193	21
E42	757	61
E44	380	45
E45	337	30
E49	635	56
E52	3	1
E53	518	39

Estación	Neurona	Grupo
E54	452	36
E55	808	64
E60	15	5
E61	92	2
E62	571	48
E63	58	7
E64	528	53
E65	181	28
E67	724	60
E71	5	1
E76	10	3
E77	292	33
E78	208	26
E80	143	16
E81	430	40
E85	44	13
E86	243	29
E88	57	7
E89	271	28
E91	129	19
E96	13	4
E97	47	14
E101	91	8
E102	36	2
E105	98	10
E108	361	35
E109	377	43

Estación	Neurona	Grupo
E116	721	49
E118	7	2
E122	49	5
E123	233	25
E124	504	46
E126	96	9
E129	511	48
E130	357	34
E131	101	12
E138	331	18
E140	82	66
E142	318	32
E144	183	17
E147	572	49
E149	61	1
E150	94	2
E151	225	22
E154	632	49
E156	333	18
E158	29	5
E159	651	57
E161	577	51
E162	756	60
E163	201	24
E166	245	29
E174	365	37

Tabla 16. Distribución de estaciones por grupos para la Tipología No. 3

Para facilitar la distinción de los grupos se reenumeran, omitiendo los que no poseen estaciones. En la tabla No. 17 se muestra la nueva distribución de las estaciones por grupos.

Estación	Grupo
E27	1
E37	1
E52	1
E71	1
E149	1
E61	2
E102	2
E118	2
E150	2
E76	3
E96	4
E29	5
E31	5
E36	5
E40	5
E60	5
E122	5
E158	5
E63	6
E88	6
E101	7
E126	8
E105	9
E15	10
E131	11
E85	12
E97	13

Estación	Grupo
E80	14
E144	15
E16	16
E138	16
E156	16
E91	17
E9	18
E41	19
E28	20
E151	20
E22	21
E163	22
E123	23
E78	24
E65	25
E89	25
E86	26
E166	26
E45	27
E13	28
E142	29
E77	30
E130	31
E108	32
E25	33
E54	33
E174	34

Estación	Grupo
E1	35
E53	36
E81	37
E18	38
E109	39
E44	40
E124	41
E8	42
E62	43
E129	43
E116	44
E147	44
E154	44
E33	45
E161	46
E10	47
E64	48
E49	49
E159	50
E4	51
E67	52
E162	52
E42	53
E55	54
E32	55
E140	56

Tabla 17. Grupos finales para la Tipología No.3

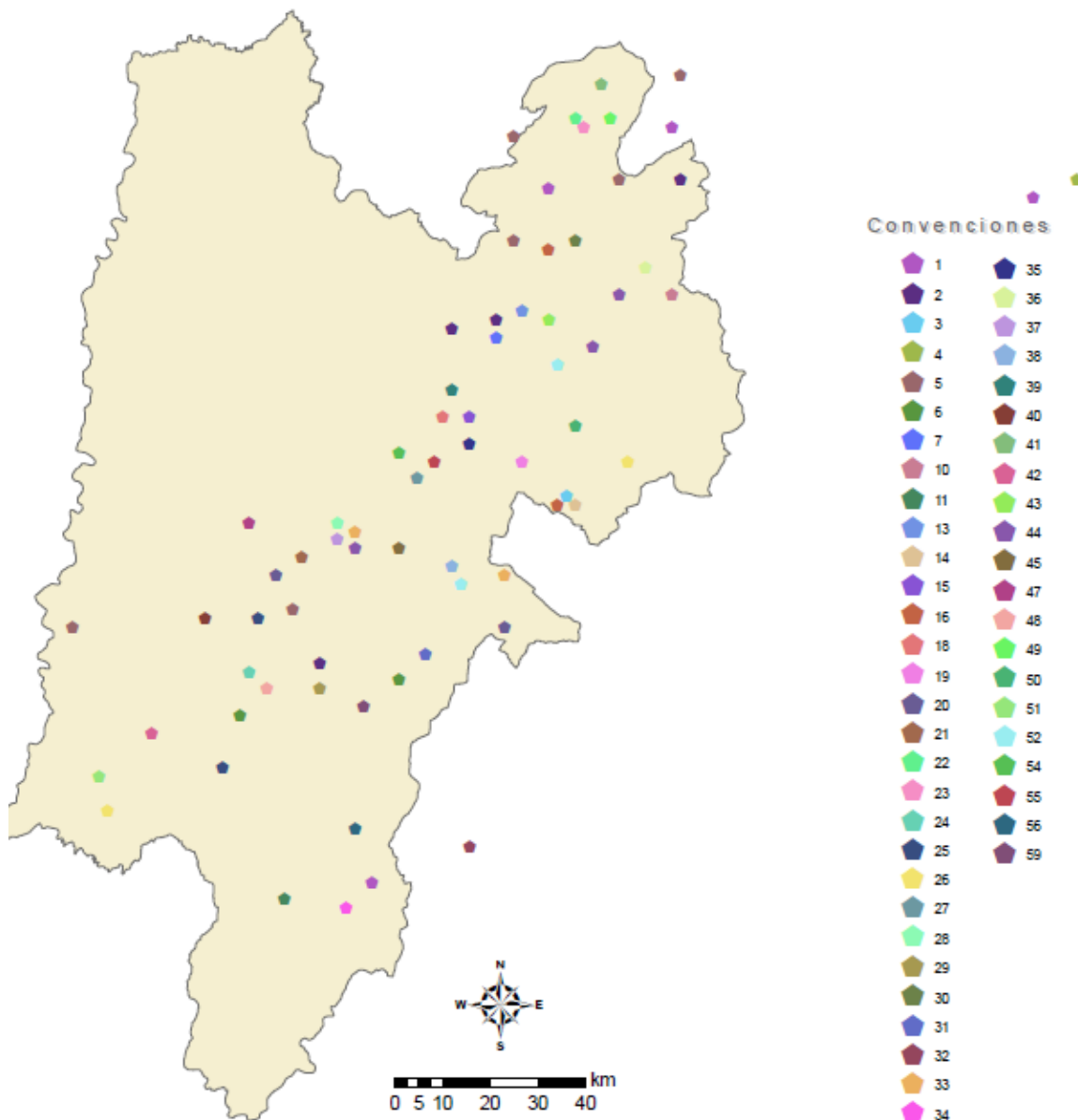


Figura 27. Distribución geográfica de los grupos para la Tipología No. 3

4.2.3 Análisis de la clasificación

En la Tabla No. 18 se presenta un resumen de los resultados obtenidos para las distintas tipologías. Se observa que al aumentar el número de neuronas, aumenta el número de grupos formados, esto se entiende teniendo en cuenta en un proceso de clustering un número alto de neuronas en la capa de salida puede ocasionar que el modelo se sobre-entrene, resaltando diferencias entre cada uno de los patrones, es por esto que en la

Tipología No. 3, con 900 neuronas, cerca del 55% de las estaciones fueron clasificadas de manera individual.

TIPOLOGÍA No.	NÚMERO DE NEURONAS	NÚMERO DE GRUPOS VÁLIDOS*
1	100	13
2	400	47
3	900	56

* Grupos en los que al menos una neurona es ganadora

Tabla 18. Resumen de la clasificación por tipología

Es significativo observar que al realizar un análisis comparativo entre las distintas tipologías se encuentra la permanencia de ciertos rasgos en la clasificación. Un caso particular lo constituye la estación E140 que siempre fue clasificada de manera individual, así mismo la estación E80, en las tipologías No. 1 y No. 3. En la Tabla No. 19 se muestran las combinaciones de estaciones más frecuentes en las tres clasificaciones.

No.	Estaciones
1	E101-E61-E16-E65
2	E116-E154-E162-E163- E62-E67
3	E122-E29-E31-E36-E40-E60- E158
4	E142-E28
5	E144-E166-E86-E89-E138
6	E149-E27-E37-E52-E71
7	E150-E61-E102
8	E16-E101
9	E28-E142-E9
10	E42-E129
11	E76-E122

Tabla 19. Combinación de estaciones comunes en las tipologías

Capítulo 5

Cálculo de la transformación

Se procede ahora a evaluar la entropía en cada uno de los grupos conformados en la fase anterior. Para esta fase se empleará el código de Matlab mostrado en el Anexo No. 3a. A manera de ejemplo se presenta el procedimiento aplicado para el grupo No. 9 compuesto por las estaciones mostradas en la Tabla No. 20 Al final del capítulo se presentará un resumen de resultados de todos los grupos.

Estación	Longitud	Latitud	Altitud
E124	948800	988200	550
E130	960600	1000400	790
E159	951300	1017600	1150
E163	963800	1000500	1110
E44	955600	1011130	1050
E64	965470	1009920	2050
E77	959500	997800	1100
E78	957190	1003360	1300

Tabla 20. Estaciones pluviométricas del Grupo No. 9

5.1 Datos de entrada

La variable de entrada para medir la entropía, en este caso, será la precipitación mensual. Como se ha señalado la red pluviométrica de la Corporación Autónoma Regional de Cundinamarca posee registros históricos de precipitación mensual desde 1931. Sin embargo la longitud de las series para cada una de las estaciones varía de acuerdo a su fecha de instalación, por ello la cantidad de datos de entrada en cada grupo puede variar conforme a la información disponible, se ha establecido un mínimo de 120 datos en común para realizar el procedimiento.

Para las estaciones del Grupo No. 9, se registra un primer dato en 1986 y culmina en 2011. En la Tabla No. 21 se muestra parte de los datos de precipitación mensual para estas estaciones.

Hito		Precipitación mensual (mm)							
Año	Mes	E124	E130	E159	E163	E44	E64	E77	E78
1995	1	15	2	18	9,9	7,2	52	4,1	3,6
	2	83,4	63	171,6	113,7	29,8	47	71,5	73,7
	3	97,8	191,8	113,6	179,8	134,5	92	45,7	136,9
	4	88,2	91	68,9	117,8	242	0	15,6	0,7
	5	92,5	137	119,7	160,6	94,2	82	95,7	109,4
	6	123,6	129	58,1	143,6	76,1	121,5	89,3	84,4
	7	79	66	101,8	60,5	57,4	64	73,3	64,4
	8	106,9	110	78,7	72,9	135,4	171,3	91,9	178,7
	9	55	15,9	185,2	36,3	140,4	53,5	81,1	88,6
	10	136,6		200,7	144	138,1	164	102,3	106
	11	152,2	307	83,6	252,5	82,7	134	140,6	182,2
	12	108,2	86	52,4	106,4	88,5	168	62,6	154,1
1996	1	158,5	52,3	176,8	125,7	83	53,8	186,3	150,9
	2	136,9	60	52,4	100,5	178,6	46	63,8	80,6
	3	315	150	101,4	187,7	153,4	196	134	192,1
	4	122,5	82	84,4	62,1	62,5	85	79,7	77,5
	5	165,6	157	130	202,7	116,6	162	155,1	187,7
	6	30,5	119	81,8	128,3	120,5	55	113,2	142,2
	7	72,7	36,6	36	65,9	83,4		57	83
	8	26,6	34,2	95,7	29,9	34,9	139	27,7	77,6
	9	68,4	48	80,1	68,7	25,1	14,5	68,5	64,8
	10	211,8	168	150,1	158,1	219,4	64,5	180,3	130,2
	11	130	103,5	106,1	110,2	108	76	109,1	128,9
	12	110,8	62,4	71,8	65,9	79,6	33,5	87,2	141,3

Tabla 21. Precipitación mensual – Grupo No. 9

5.2 Cálculo de la Entropía y de la Transinformación

A manera de ejemplo, se muestra, a continuación, el cálculo de estos valores para dos estaciones.

5.2.1 Estación E163 con E77

Se denomina, en este caso, la estación E163 como “Serie 1” y la estación E77 como “Serie 2”. En la Figura No. 28 se puede observar el pluviograma de cada una de estas estaciones.

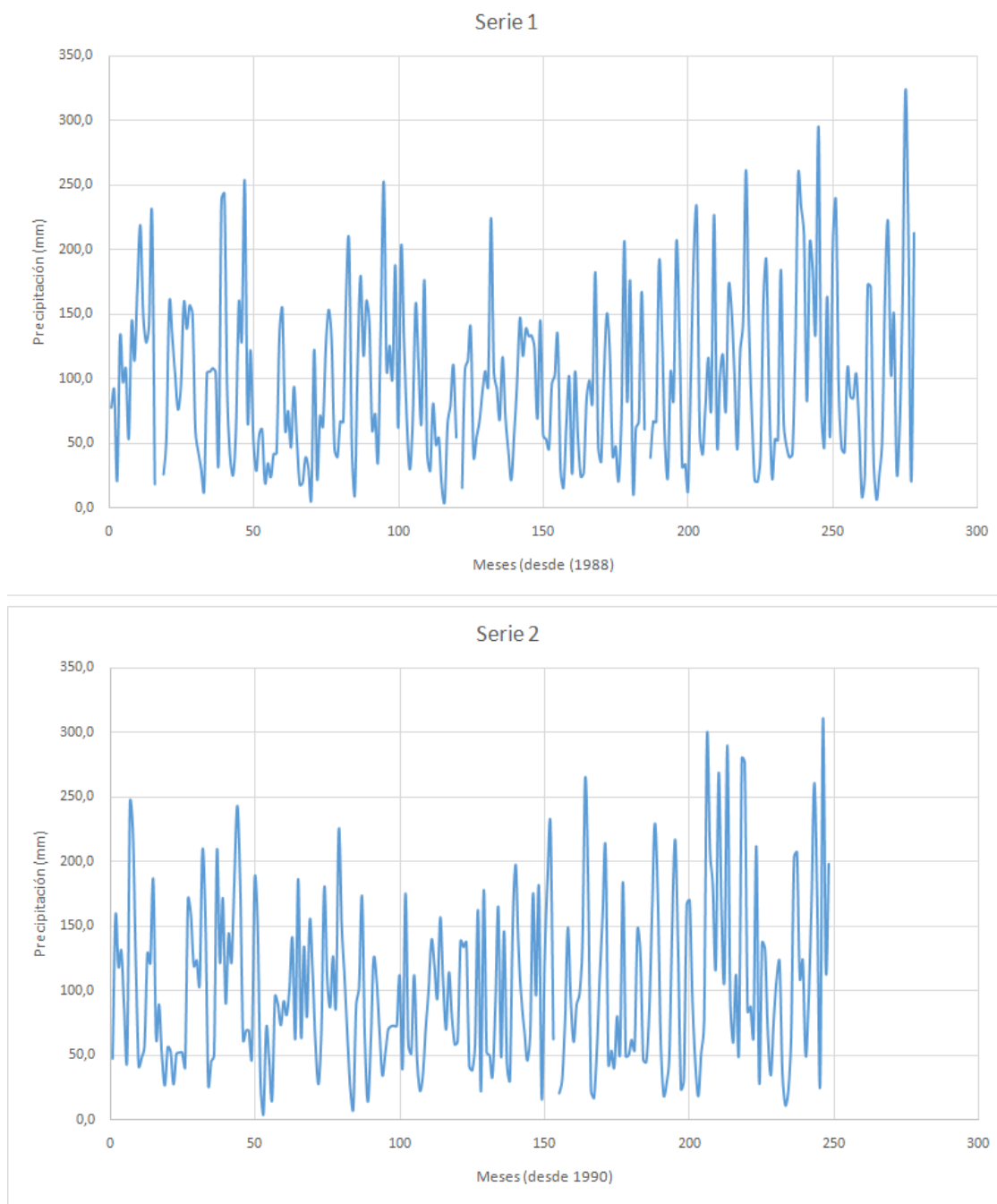


Figura 28. Pluviograma de las Estaciones E163 y E77

Se procede a continuación a elaborar el histograma de cada serie. Para determinar el número de clases de tuvo en cuenta la Regla de Sturges, según la cual éste viene dado por la expresión:

$$c = 1 + 3.3322 * \text{Log}N$$

Donde N es el tamaño de la muestra.

En la Figura No. 29 se puede apreciar el histograma de las series 1 y 2.

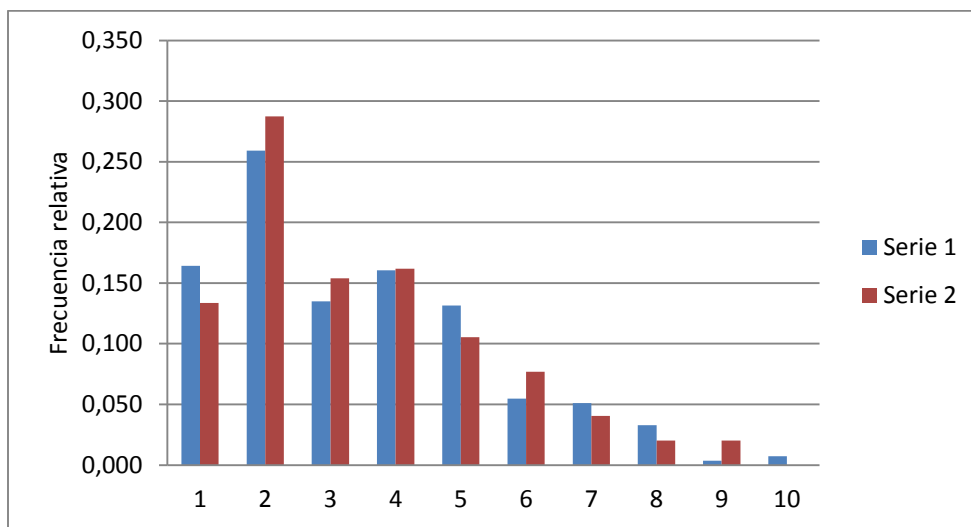


Figura 29. Histograma Serie 1 y 2

A partir de la Serie 1 se genera la Serie 2 mediante una regresión lineal múltiple. Los resultados de esta regresión se muestran en la Figura No. 30.

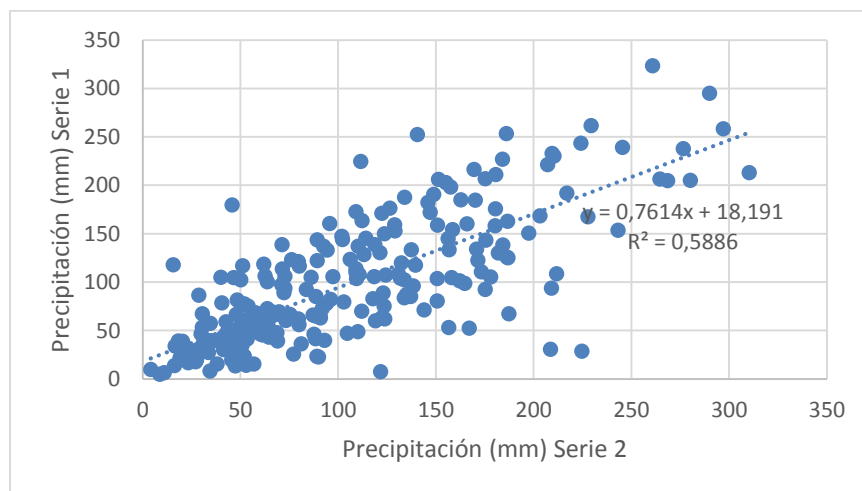


Figura 30. Serie 1 a partir de la Serie 2

Se construye un histograma de la Serie 1 generada a partir de la Serie 2. Éste puede ser observado en la Figura No. 31.

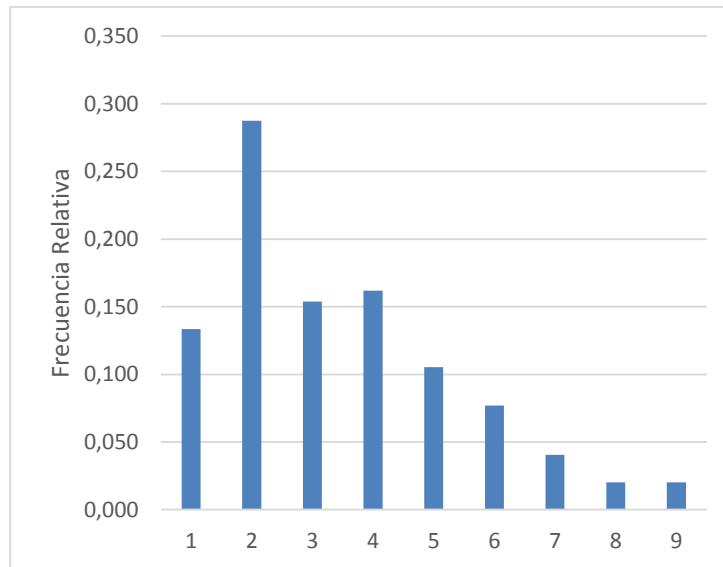


Figura 31. Histograma Serie 1 Generada a partir de Serie 2

En la Figura No. 32 se puede apreciar una comparación entre la Serie 1 observada (datos reales) y la Serie 1 generada a partir de la regresión.

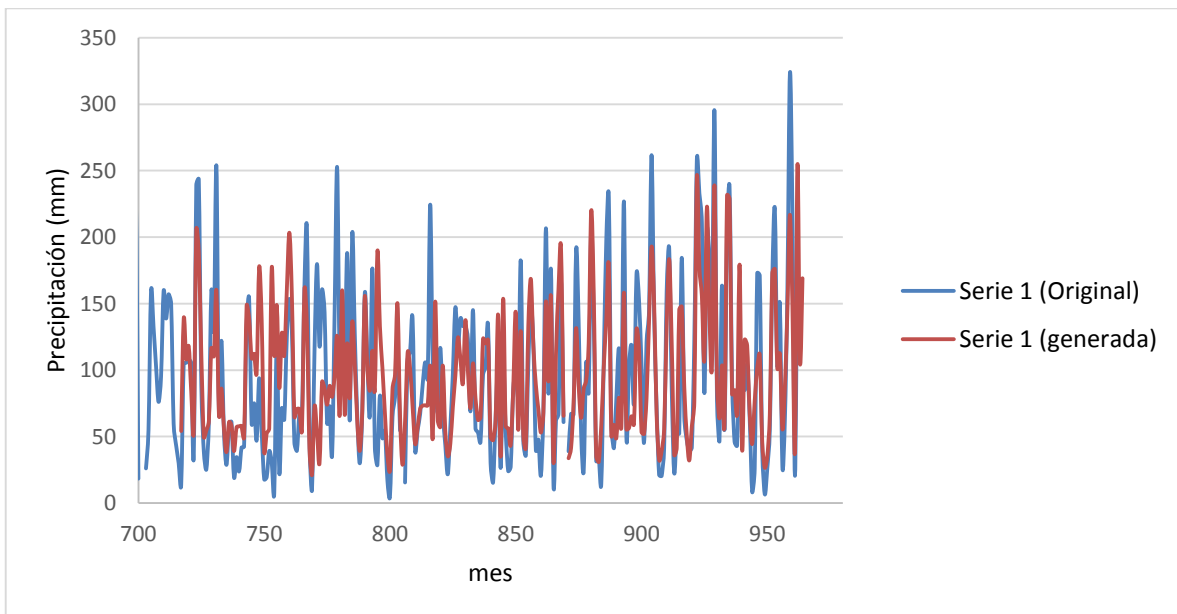


Figura 32. Comparación entre la Serie 1 Original y la Serie 1 generada

De manera similar, se procede ahora a generar la Serie 2 a partir de la Serie 1 mediante una regresión lineal múltiple. Los resultados de esta regresión y el correspondiente histograma pueden ser observados en las Figuras No. 33 y 36 respectivamente.

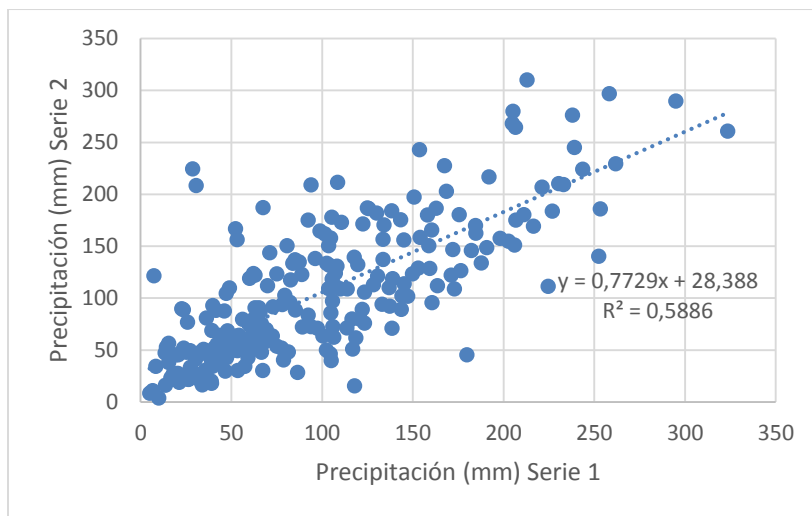


Figura 33. Serie 2 a partir de la Serie 1

En la Figura No. 34 se puede apreciar una comparación entre la Serie 2 observada (datos reales) y la Serie 2 generada a partir de la regresión.

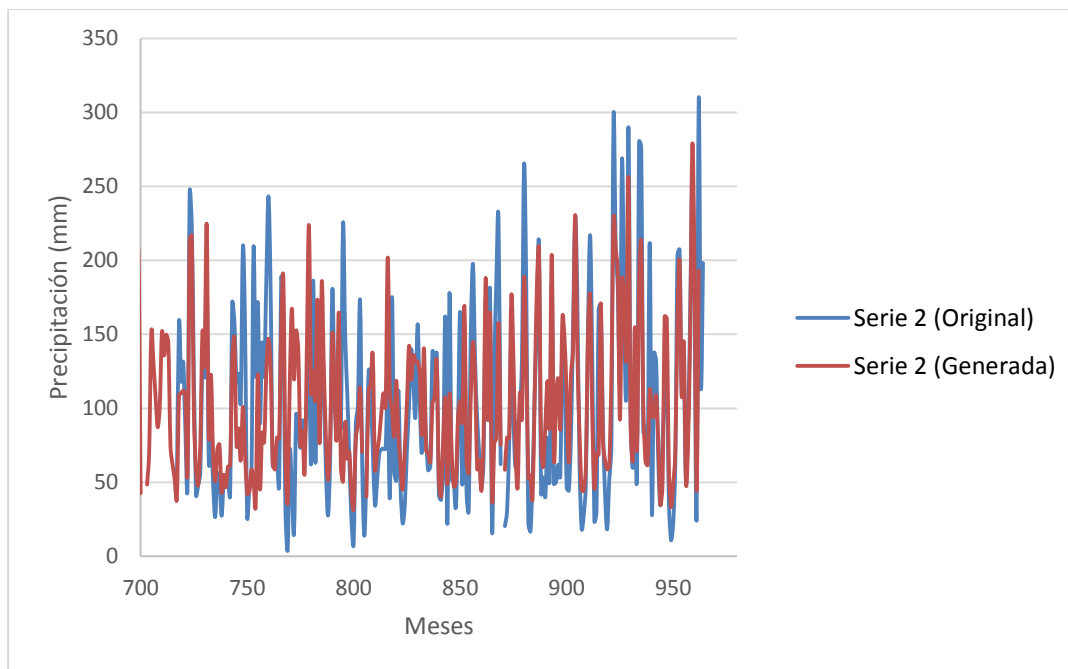


Figura 34. Comparación entre la Serie 2 Original y la Serie 2 generada

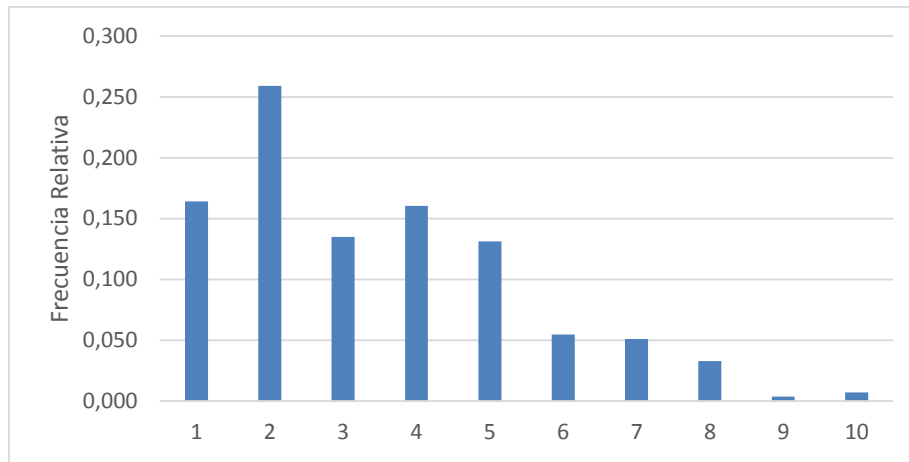


Figura 35. Histograma Serie 2 Generada a partir de Serie 1

En la Figura No. 36 se muestran los histogramas bivariados para las series originales y las series generadas mediante la regresión.

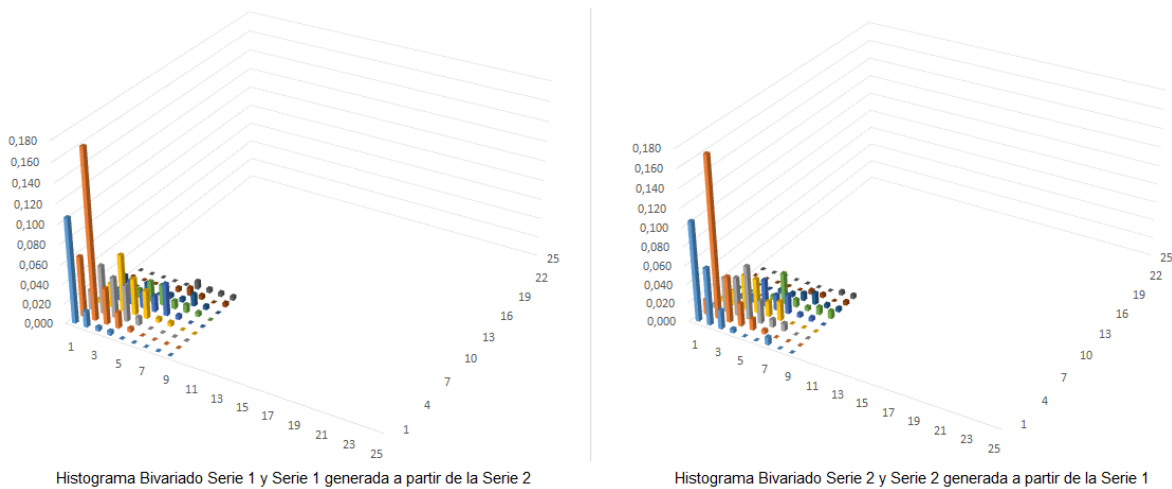


Figura 36. Histogramas bivariados

Se procede ahora al cálculo de la entropía marginal para las Serie 1 y 2. Así mismo se calcula la entropía marginal para las series generadas, empleando la expresión:

$$H = - \sum_{k=1}^k p(x_k) \text{Log}[p(x_k)]$$

Donde k es un intervalo de datos discretos, x_k es un resultado correspondiente al intervalo de k, y $p(x_k)$ es la probabilidad de x_k .

Luego se calcula la entropía conjunta entre la serie original y la serie generada, mediante la expresión:

$$H(X, Y) = - \sum_{k=1}^K \sum_{L=1}^L p(x_k y_L) \text{Log}[p(x_k y_L)]$$

Finalmente se calcula la entropía para cada una de las estaciones, mediante la expresión:

$$T(X, Y) = H(X) + H(Y) - H(X, Y)$$

H(X) corresponde a la entropía marginal de los datos reales, H(Y) corresponde a la entropía marginal para la serie simulada y H(X,Y) es la entropía conjunta.

Para la Serie 1, correspondiente a la estación E163, se encontraron los valores registrados en la Tabla No. 22.

Entropía Marginal de la Serie 1	2,823
Entropía Marginal Serie 1 generada con la Serie 2	2,787
Entropía Conjunta	4,695
Transinformación	0,915

Tabla 22. Transinformación Serie 1 (E163)

Para la Serie 2, correspondiente a la estación E77, se encontraron los valores registrados en la Tabla No. 23.

Entropía Marginal de la Serie 2	2,787
Entropía Marginal Serie 2 generada con la Serie 1	2,823
Entropía Conjunta	4,695
Transinformación	0,915

Tabla 23. Transinformación Serie 2 (E77)

Este mismo procedimiento se repitió con cada una de las 8 estaciones que conforman el Grupo No. 9, para obtener finalmente el valor de la transinformación de cada combinación, en el siguiente numeral se presenta un resumen de estos resultados.

5.3 Resultados por grupos

Se presentan los resultados de la entropía para cada uno de los grupos de la Tipología No. 1, en el Anexo No. 3b se aprecian en detalle los pasos previos para obtener tales resultados.

5.3.1 Grupo No. 1

	E101	E102	E126	E138	E144	E149	E150	E156	E16	E166	E25	E27	E36	E37	E52	E60	E61	E65	E71	E86	E89
E101	0	0,794	0,569	0,516	0,297	0,482	0,846	0,183	0,517	0,552	0,414	0,618	0,480	-1,560	0,374	0,515	0,608	0,465	0,650	0,142	0,605
E102	0,792	0	0,585	0,571	0,350	0,535	0,962	0,197	0,530	0,495	0,367	0,641	0,612	-1,499	0,340	0,592	0,591	0,469	0,660	0,191	0,814
E126	0,569	0,585	0	0,479	0,258	0,382	0,624	0,338	0,518	0,646	0,403	0,440	0,483	0,282	0,493	0,452	0,598	0,451	0,577	0,175	0,859
E138	0,516	0,571	0,479	0	0,534	0,353	0,439	0,776	0,448	0,664	0,526	0,298	0,326	0,326	0,380	0,392	0,380	0,517	0,682	0,873	0,941
E144	0,297	0,350	0,258	0,534	0	0,197	0,216	0,698	0,274	0,645	0,523	0,135	0,198	0,250	0,259	0,208	0,227	0,236	0,571	0,818	0,650
E149	0,482	0,535	0,384	0,353	0,197	0	0,728	0,100	0,546	0,520	0,397	0,612	0,774	-1,234	0,279	0,394	0,584	0,400	0,553	-0,024	0,724
E150	0,846	0,962	0,624	0,439	0,216	0,728	0	0,386	0,438	0,707	0,567	0,629	0,780	0,398	0,674	0,671	0,852	0,655	0,966	0,342	0,816
E156	0,183	0,196	0,346	0,776	0,698	0,100	0,386	0	0,180	0,323	0,358	0,281	0,260	-1,704	0,121	0,104	0,282	0,170	0,511	0,959	0,418
E16	0,517	0,530	0,518	0,448	0,274	0,546	0,438	0,180	0	0,172	0,394	0,595	0,646	-1,427	0,375	0,421	0,557	0,410	0,644	-0,293	0,157
E166	0,552	0,495	0,646	0,664	0,645	0,527	0,707	0,323	0,172	0	0,009	0,505	0,255	-1,934	0,637	0,448	0,634	0,312	0,543	0,498	0,683
E25	0,414	0,367	0,403	0,526	0,523	0,390	0,567	0,358	0,394	0,009	0	0,421	0,514	-1,983	0,277	0,390	0,473	0,288	0,398	0,214	0,299
E27	0,618	0,641	0,440	0,298	0,135	0,612	0,629	0,281	0,595	0,505	0,421	0	0,847	0,347	0,103	0,562	0,542	0,495	0,635	0,161	0,788
E36	0,480	0,612	0,483	0,326	0,198	0,774	0,780	0,258	0,646	0,255	0,512	0,847	0	0,375	0,585	0,658	0,634	0,633	0,787	0,091	0,200
E37	-1,560	-1,499	0,282	0,326	0,250	-1,234	0,398	-1,704	-1,431	-1,935	-1,983	0,347	0,375	0	0,145	-1,667	0,367	-1,620	-1,481	-1,951	-1,701
E52	0,374	0,340	0,493	0,380	0,259	0,279	0,674	0,121	0,375	0,637	0,277	0,103	0,585	0,145	0	0,347	0,627	0,291	0,434	-0,121	0,524
E60	0,515	0,592	0,452	0,392	0,208	0,394	0,671	0,104	0,421	0,448	0,390	0,562	0,658	-1,667	0,347	0	0,545	0,319	0,472	0,145	0,711
E61	0,608	0,591	0,598	0,380	0,227	0,584	0,852	0,282	0,557	0,634	0,473	0,542	0,634	0,367	0,627	0,545	0	0,486	0,699	0,152	0,771
E65	0,465	0,470	0,451	0,517	0,236	0,400	0,655	0,164	0,410	0,312	0,290	0,495	0,633	-1,620	0,291	0,319	0,486	0	0,455	-0,072	0,299
E71	0,650	0,660	0,577	0,682	0,571	0,553	0,966	0,511	0,640	0,543	0,398	0,635	0,786	-1,481	0,434	0,472	0,699	0,453	0	0,532	0,759
E86	0,142	0,191	0,175	0,873	0,818	-0,024	0,342	0,959	-0,293	0,498	0,214	0,161	0,091	-1,951	-0,121	0,145	0,152	-0,072	0,532	0	0,390
E89	0,605	0,814	0,859	0,941	0,650	0,724	0,816	0,418	0,157	0,683	0,299	0,788	0,200	-1,701	0,524	0,711	0,771	0,299	0,759	0,390	0

Tabla 24. Transinformación para las estaciones del Grupo No. 1

5.3.2 Grupo No. 2

	E122	E158	E29	E31	E40	E76	E96
E122	0	0,6547	0,8253	0,9212	0,6368	0,6088	0,5839
E158	0,6547	0	0,7449	0,8644	0,8050	0,6871	0,5827
E29	0,8253	0,7449	0	0,8261	0,7534	0,6428	0,4192
E31	0,9212	0,8656	0,8244	0	0,8402	0,7143	0,5483
E40	0,6368	0,8050	0,7534	0,8402	0	0,7327	0,6477
E76	0,6088	0,6871	0,6428	0,7143	0,7327	0	0,6181
E96	0,5839	0,5827	0,4192	0,5483	0,6477	0,6181	0

Tabla 25. Transinformación para las estaciones del Grupo No. 2

5.3.2 Grupo No. 3

Para el Grupo No. 3 no se realizó este procedimiento porque está conformado por una sola estación, la E140, lo que indica que no presenta similitud en su información con las demás estaciones.

5.3.4 Grupo No. 4

	E105	E118	E131	E15	E91
E105	0	0,6029	0,6958	0,3943	0,6756
E118	0,6029	0	0,6125	0,9319	0,5792
E131	0,6958	0,6125	0	0,5893	0,7039
E15	0,3934	0,9319	0,5933	0	0,7641
E91	0,6756	0,5792	0,7039	0,7680	0

Tabla 26. Transinformación para las estaciones del Grupo No. 4

5.3.5 Grupo No. 5

	E109	E85	E97
E109	0	0,3690	0,3377
E85	0,3690	0	0,2948
E97	0,3229	0,2948	0

Tabla 27. Transinformación para las estaciones del Grupo No. 5

5.3.6 Grupo No. 6

Para el Grupo No. 6 no se realizó este procedimiento porque está conformado por una sola estación, la E80, lo que indica que no presenta similitud en su información con las demás estaciones.

5.3.7 Grupo No. 7

	E123	E63	E88
E123	0	0,5608	0,5083
E63	0,5608	0	0,6599
E88	0,5127	0,6599	0

Tabla 28. Transinformación para las estaciones del Grupo No. 7

5.3.8 Grupo No. 8

	E108	E142	E151	E174	E18	E28	E41	E45	E9
E108	0	0,4931	0,5249	0,5771	0,4807	0,4157	0,4884	0,4507	0,3872
E142	0,4931	0	0,3996	0,1386	0,4134	0,2845	0,2682	0,5212	0,4674
E151	0,5249	0,3996	0	0,5405	0,3682	0,3182	0,1189	0,4386	0,5355
E174	0,5771	0,1386	0,5405	0	0,5005	0,3137	0,3199	0,6148	0,4923
E18	0,4807	0,4134	0,3682	0,5005	0	0,1117	0,2634	0,3957	0,4336
E28	0,4157	0,2845	0,3182	0,3137	0,1117	0	0,2024	0,4412	0,4039
E41	0,4884	0,2682	0,1189	0,3199	0,2634	0,2024	0	0,2211	0,3776
E45	0,4507	0,5212	0,4386	0,6148	0,3957	0,4412	0,2211	0	0,6454
E9	0,3872	0,4756	0,5355	0,4923	0,4336	0,4024	0,3776	0,6454	0

Tabla 29. Transinformación para las estaciones del Grupo No. 8

5.3.9 Grupo No. 9

	E124	E130	E159	E163	E44	E64	E77	E78
E124	0	0,5108	0,5690	0,7270	0,7105	0,4220	0,6901	0,8367
E130	0,5108	0	0,3564	0,8055	0,4929	0,2769	0,7308	0,7796
E159	0,5690	0,3564	0	0,6271	0,5853	0,1131	0,5178	0,5322
E163	0,7270	0,8055	0,6271	0	0,6432	0,4923	0,9153	0,6488
E44	0,7105	0,4929	0,5853	0,6432	0	0,3584	0,6533	0,7412
E64	0,4220	0,2769	0,1131	0,4923	0,3584	0	0,3900	0,5573
E77	0,6901	0,7308	0,5178	0,9153	0,6533	0,3900	0	0,9157
E78	0,8367	0,7796	0,5322	0,6488	0,7412	0,5573	0,9157	0

Tabla 30. Transinformación para las estaciones del Grupo No. 9

5.3.10 Grupo No. 10

	E116	E129	E147	E154	E162	E32	E42	E62	E67
E116	0	0,7432	0,6308	0,5208	0,8745	0,3183	0,3145	0,4843	0,8510
E129	0,7432	0	0,5181	0,2187	0,8213	0,9637	0,3522	0,7769	0,5612
E147	0,6308	0,5181	0	0,3376	0,7269	1,0082	0,2647	0,3937	0,6326
E154	0,5208	0,2187	0,3376	0	0,7098	0,0547	0,3213	0,1779	0,2523
E162	0,8745	0,8213	0,7269	0,7098	0	0,3740	0,6358	0,8661	0,8596

E32	0,3183	0,9637	1,0082	0,0547	0,3740	0	0,2871	0,7460	0,9725
E42	0,3145	0,3522	0,2647	0,3213	0,6358	0,2871	0	0,2921	0,3269
E62	0,4843	0,7769	0,3937	0,1779	0,8661	0,7460	0,2921	0	0,6113
E67	0,8510	0,5612	0,6326	0,2523	0,8596	0,9725	0,3269	0,6113	0

Tabla 31. Transinformación para las estaciones del Grupo No. 10

5.3.12 Grupo No. 12

	E4	E55	E8
E4	0	0,4404	0,6151
E55	0,4404	0	0,5259
E8	0,6151	0,5259	0

Tabla 32. Transinformación para las estaciones del Grupo No. 12

5.3.13 Grupo No. 13

	E1	E161	E33	E49	E53	E54
E1	0	0,6053	0,5746	0,4360	0,6952	0,6099
E161	0,6076	0	0,4980	0,6812	0,1719	0,5101
E33	0,5802	0,4980	0	0,4322	0,8659	0,9018
E49	0,4360	0,6812	0,4322	0	0,7924	0,5087
E53	0,6952	0,1719	0,8659	0,7924	0	0,7299
E54	0,6099	0,5101	0,9018	0,5087	0,7299	0

Tabla 33. Transinformación para las estaciones del Grupo No. 13

De acuerdo a la Tabla No. 8, las estaciones, en promedio, presentan déficit de información, lo que indica que cada una de ellas es independiente, siendo inviable la reconstrucción de la serie histórica de precipitación de una estación a partir de las otras estaciones con la que comparte grupo.

Conclusiones y Recomendaciones

Mediante la unión de la técnica de Mapas Auto-organizados o Redes de Kohonen y conceptos de entropía de la información, se logró evaluar la distribución de las estaciones de la red pluviométrica de la región Bogotá-Cundinamarca. Esta unión de conceptos supone novedad, siendo un aporte para el estudio de redes pluviométricas.

La aplicación de redes neuronales artificiales para la clasificación de las estaciones de la red pluviométrica del área de estudio, mostró que tales estaciones pueden ser agrupadas en 13, 47 y 56 grupos, dependiendo del número de neuronas en la capa de salida. En la Tipología No. 1, con 100 neuronas en la capa de salida, se obtuvo un total de 13 grupos, en la Tipología No. 2, con 400 en la capa de salida, se obtuvo un total de 47 grupos y, finalmente, en la Tipología No. 3, con 900 neuronas en la capa de salida, se obtuvo un total de 56 grupos. Se observa que al aumentar el número de estas neuronas, aumenta el número de grupos formados, esto viene a estar justificado en que, en un proceso de clustering un número alto de neuronas en dicha capa, puede ocasionar que el modelo se sobre-entrene, resaltando diferencias entre cada uno de los patrones. Se ajustan así los resultados obtenidos a lo planteado por la literatura.

A pesar del cambio del número de neuronas en la capa de salida, conservando las variables de entrada, se encontró la permanencia de ciertos rasgos en la clasificación, lo que asegura la existencia de patrones similares en los grupos obtenidos. Tal es el caso de las estaciones E142 (La Casita) y E28 (Santa Teresa), ubicadas en cercanías al municipio de La Calera, que en las distintas tipologías fueron clasificadas en un mismo grupo. Este hecho hace notar cierta correspondencia entre ellas, por lo que se recomienda, para estudios futuros o para la reingeniería de la red, considerar la posible reconstrucción de la serie histórica de una de éstas a partir de la otra, lo que permitiría la reubicación de una de ellas en sitios donde hay menor número de estaciones. Igual situación ocurre con las estaciones E42 (San Jorge) - E129 (Doña Juana), E76 (Monserrate) - E122 (Esclusa) y E16 (Tres Esquinas) - E101 (El Hato No.2), que fueron clasificadas en grupos similares en las distintas tipologías.

Un caso contrario lo constituyen las estaciones E80 (Las Margaritas) y E140 (Central No. 2), correspondientes a las oficinas provinciales de Sabana Occidente y Bogotá – La Calera, respectivamente. Estas estaciones fueron clasificadas de manera individual, este hecho hace ver que la información que poseen no se podría reconstruir a partir de las otras estaciones cercanas, por ello se recomienda a la Corporación Autónoma Regional de Cundinamarca, en caso de una posible reingeniería de la red, no reubicar estas estaciones puesto que su información es única, esto sumado a que presentan un valioso registro histórico desde 1959.

Se encontró que la parte nororiental de la zona de estudio (correspondiente a los municipios de Yacopí y Puerto Salgar) y la parte sur (correspondiente al municipio de Cabrera), presentan baja cobertura de estaciones pluviométricas, por lo que se recomienda aumentar el número de estaciones en esta zona. Para ello se puede evaluar el traslado de algunas estaciones que fueron clasificadas en un mismo grupo en las distintas tipologías, como se señaló previamente (E142, E28, E42, E129, E76, E122, E16 y E101). Por otra parte, a pesar de la concentración de estaciones en la zona central del área estudiada, no se recomienda el traslado de otras estaciones distintas a las ya señaladas, puesto que, con este método, no es posible reconstruir la información que cada una aporta a partir de las estaciones vecinas.

Al evaluar la distribución de las estaciones pluviométricas en cada una de los grupos obtenidos mediante un análisis de su registro histórico, se encontró que los valores de Transinformación son bajos, lo que permite concluir que la red pluviométrica estudiada presenta deficiencia de información, debido a que sólo un par de estaciones arrojaron valores aceptables de transinformación, estas corresponden las estaciones E32 (Santa Isabel) y E154 (Campobello), sin embargo, dado que se encuentran en zonas distantes, municipio de Tabio y Madrid, respectivamente, no es recomendable la reubicación de una de estas estaciones bajo este análisis.

Era de esperarse que por la alta concentración de estaciones en la zona central del área de estudio, se presentaran en las estaciones allí ubicadas, mejores valores de transinformación, sin embargo los resultados obtenidos muestran la misma tendencia que en el resto del área de estudio, valores deficitarios. Se puede concluir así, que el criterio de cercanía entre estaciones no garantiza la homogeneidad de la información que

aportan, por lo que sería poco recomendable el traslado de alguna de estas estaciones bajo este criterio.

Debido a los bajos valores de transinformación encontrados, se recomienda una integración de la información con las redes pluviométricas de otras entidades (como el IDEAM) que tienen estaciones en esta misma área, para ampliar la cobertura de la red. Se recomienda realizar una nueva evaluación de la red pluviométrica de la zona, incluyendo, además de las estaciones de la Corporación Autónoma Regional de Cundinamarca, las estaciones de otras entidades que hacen presencia en la zona.

En suma, se recomienda ampliar la cobertura de la red pluviométrica en las zonas nororiental y sur del área de estudio. Para ello, se podría disponer de un grupo de estaciones que, de acuerdo con la clasificación obtenida mediante redes neuronales, presentan características similares. No obstante, esta reubicación debe evaluarse en detalle, porque los valores de transinformación obtenidos para las mismas estaciones son deficitarios. Se recomienda evaluar, en un futuro estudio, la transinformación de estas estaciones con sus vecinas más próximas. Si en tal caso, presentan valores aceptables de transinformación, el traslado sería viable. Por otra parte, como se ha señalado, las estaciones E80 y E140 constituyen un caso particular, puesto que en las tres tipologías de clasificación, mostraron tener valores atípicos respecto a las demás estaciones. Se recomienda no trasladar estas estaciones, valorando la información única que aportan.

Trabajos citados

A. Kheirkhah, A. A. (2013). *Improved estimation of electricity demand function by using of artificial neural network, principal component analysis and data envelopment analysis*. Computers & Industrial Engineering , 425–441.

A.K. Lohani, R. K. (2012). *Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques*. Journal of Hydrology, 23–35.

A.K. Mishra, P. C. (2010). *Hydrometric network evaluation for Canadian watersheds*. Journal of Hydrology, 420–437.

Ainslie, B., C. Reuten, D. S., Le, N. D., & Zidek, J. V. (2009). *Application of an entropy-based Bayesian optimization technique to the redesign of an existing monitoring network for single air pollutants*. Journal of Environmental Management, 2715-2729.

Aliguliyev, R. M. (2009). *Performance evaluation of density-based clustering methods*. Information Sciences .

Ani E. Asimakopoulou, G. J. (2013). *Estimation of seasonal variation of ground resistance using Artificial Neural Networks*. Electric Power Systems Research , 113–121.

Anish C. Turlapaty, V. G. (2010). *Precipitation data fusion using vector space transformation and artificial neural networks*. Pattern Recognition Letters .

Aurbacher, J., & Dabbert, S. (2010). *Generating crop sequences in land-use models using maximum entropy and Markov chains*. Agricultural Systems , 470-479.

Bernal, G., & Gutiérrez, J. (1999). *Aplicación de Redes Neuronales en la Programación y Control de la Producción*. Medellín: Universidad EAFIT.

Calisto, O. E. (2002). *El aforo de ríos con un punto de velocidad basado en el principio de la máxima entropía*. Ingeniería Hidráulica en México , 5-19.

CAR. (2009). *Caracterización de usuarios CAR*. Bogotá: Subdirección de Gestión Social - CAR.

CAR. (2009). *Estudio de diagnóstico, prospectiva y formulación para la cuenca hidrográfica del río Minero*. Bogotá: Corporación Autónoma Regional de Cundinamarca.

CAR. (2011). *Plan de acción 2007- junio 2012*. Bogotá: Corporación Autónoma Regional de Cundinamarca.

CAR. (2012). *Plan de acción 2012-2015* CAR. Bogotá: Corporación Autónoma Regional de Cundinamarca.

CAR. (2006). *Plan de ordenación y manejo de la cuenca hidrográfica del río Bogotá*. Bogotá: Corporación Autónoma Regional de Cundinamarca.

Carvalho, F. M. (2008). *Clustering constrained symbolic data*. Pattern Recognition Letters.

Ching-Hsue Cheng, J.-W. W.-C. (2010). *OWA-weighted based clustering method for classification problem*. Expert Systems with Applications.

Daleziosa, N. R., & Tyraskis, P. A. (2003). *Maximum entropy spectra for regional precipitation analysis and forecasting*. Journal of Hydrology, 25-42.

Dursun, M., & Özden, S. (2014). *An efficient improved photovoltaic irrigation system with artificial neural network based modeling of soil moisture distribution – A case study in Turkey*. Computers and Electronics in Agriculture .

Fenga, L., & Hongb, W. (2007). *On hydrologic calculation using artificial neural networks*. Applied Mathematics Letters , 453–458.

Fernández, Ó. (1991). *El análisis de clúster: aplicación, interpretación y validación*. Papers: Revista de sociología , 65-76.

Fukushima, K. (2013). *Artificial vision by multi-layered neural networks: Neocognitron and its advances*. Neural Networks , 103–119.

Gontijo, W. C. (2007). *Avaliação e redimensionamento de redes para o monitoramento pluviométrico utilizando o método sharp e o conceito de entropia*.

González, F. (2012). *Agrupación ecohidrológica de corrientes en la cuenca Magdalena-Cauca dentro del marco de referencia ELOHA, empleando mapas autorganizados de Kohonen*. Bogotá: Pontificia Universidad Javeriana.

Gwo-Fong Lin, L.-H. C. (2005). *Identification of homogeneous regions for regional frequency analysis using the self-organizing map* . Journal of Hydrology .

Hooman Adib, R. H. (2013). *Modeling and optimization of FischereTropsch synthesis in the presence of Co (III)/Al₂O₃ catalyst using artificial neural networks and genetic algorithm*. Journal of Natural Gas Science and Engineering , 14-24.

IDEAM. (2007). *Diseño de la ampliación de la red hidrometeorológica de Bogotá y cuenca alta del río Tunjuelo*. Santafé de Bogotá: IDEAM.

IDEAM. (2002). *Sistema de información ambiental de Colombia* . Bogotá: Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia.

- J.P. Noonan, P. B. (2007). *On estimation error using maximum entropy density estimates*. *Kybernetes* , 52 - 64.
- Jain, A., & Kumar, A. M. (2007). *Hybrid neural network models for hydrologic time series forecasting*. *Applied Soft Computing* , 585-592.
- Jung-Woo Kim, Y. A. (2010). *Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT stream flow simulation*. *Journal of Hydrology*.
- K. Srinivasa Raju, D. N. (2011). *Classification of micro watersheds based on morphological characteristics*. *Journal of Hydro-environment Research*, 101-109.
- Kalyani, & Swarup. (2011). *Particle swarm optimization based K-means clustering approach for security assessment in power systems*. *Expert Systems with Applications* , 10839–10846.
- Karabacaka, K., & Cetin, N. (2013). *Artificial neural networks for controlling wind–PV power systems: A review*. *Renewable and Sustainable Energy Reviews* .
- Kohler Max, L. R. (1977). *Hidrología para ingenieros*. Bogotá: McGraw-Hill.
- Martínez Alfaro, P., Martínez Santos, P., & Castaño, S. (2005). *Fundamentos de Hidrogeología*. Madrid: Ediciones Mundi-Prensa.
- Ministerio de Ambiente y Desarrollo Sostenible. (2005). *Plan de ordenación y manejo de la cuenca del río Blanco – Negro – Guayuriba*. Bogotá.
- Mishra, A. K., Özger, M., & Singh, V. P. (2009). *An entropy-based investigation into the variability of precipitation*. *Journal of Hydrology* , 139-154.
- Mishra, A., & Coulibaly, P. (2009). *Hydrometric network evaluation for Canadian watersheds*. *Journal of Hydrology*.
- Monsalve Sáenz, G. (1995). *Hidrología en la ingeniería*. Santafé de Bogotá: Escuela Colombiana de Ingeniería.
- Muhammed Ernur Akiner, A. A. (2012). *Modeling and forecasting river flow rate from the Melen Watershed, Turkey*. *Journal of Hydrology* , 121–129.
- Niño, L., & De Moya, M. (2005). *Representación y clasificación de datos geoespaciales usando Redes Neuronales*. XXXI Conferencia Latinoamericana de Informática.
- OEA. (2010). *Manual para el diseño, instalación, operación y mantenimiento de sistemas comunitarios de alerta temprana ante inundaciones*. Departamento de Desarrollo Sostenible - Secretaría General de la Organización de los Estados Americanos.

Payandeh Najafabadi, A., Hatami, H., & Najafabadi, M. O. (2011). *A maximum-entropy approach to the linear credibility formula*. Insurance: Mathematics and Economics.

Puangthongthub, S., Wangwongwatana, S., Kamens, R. M., & Serre, M. L. (2007). *Modeling the space/time distribution of particulate matter in Thailand and optimizing its monitoring network*. Atmospheric Environment, 7788-7805.

Qian-Jin Liu, Z.-H. S.-F.-D. (2013). *Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet-ANN approach on the Loess Plateau, China, using the Wavelet-ANN approach*. Geomorphology, 181-190.

Ribeiro, F., & Bezerra, M. C. (2014). *Evaluation of chemical composition of waters associated with petroleum production using Kohonen neural networks*. Fuel.

Riccardo Taormina, K.-w. C. (2012). *Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon*. Engineering Applications of Artificial Intelligence, 1670-1676.

Sambu Seo, K. O. (2004). *Self-organizing maps and clustering methods for matrix data*. Neural Networks.

Shanon. (1948). *A Mathematical Theory of Communication*. The Bell System Technical Journal.

Shengwei Wang, J. F. (2011). *Application of seasonal time series model in the precipitation forecast*. Mathematical and Computer Modelling .

Srinivas, S. T. (2008). *Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering*. Journal of Hydrology.

Tomi Kinnunen, I. S. (2011). *A distance based clustering method for arbitrary shaped clusters in large datasets*. Pattern Recognition Letters.

Vidal Díaz, d. R. (1998). *Diseño de tipologías de consumidores mediante la utilización conjunta del Análisis Cluster y otras técnicas multivariantes*. Revista española de economía agraria, 75-104.

Vinod, H. D. (2006). *Maximum entropy ensembles for time series inference in economics*. Journal of Asian Economics.

Weber, T. (2010). *Maximum entropy modeling of mature hardwood forest distribution in four U.S. states*. Forest Ecology and Management, 779-788.

Yegnanarayana. (2006). *Artificial neural networks*. New Delhi: Prentice-Hall of India.

Zhao, Z., & Zhang, Y. (2011). *Design of ensemble neural network using entropy theory*. Advances in Engineering Software, 838-845.