

UNIVERSIDAD PARAGUAYO ALEMANA



OPTIMIZACIÓN DE RED DE
PLUVIÓMETROS EN EL ÁREA DE GRAN
ASUNCIÓN MEDIANTE MÉTODO ANÁLISIS
DE COMPONENTES PRINCIPALES

TESIS

PARA OBTENER EL TÍTULO DE

INGENIERO INDUSTRIAL

PRESENTADA POR

EDUARDO JOSÉ RODRÍGUEZ AGUILAR

TUTOR: PROF. JUAN PABLO NOGUES, PhD

EXAMINADOR: ING. HORACIO CANIZA, PhD

Yo, **Eduardo José Rodríguez Aguilar** declaro que esta tesis titulada “*Optimización de red de pluviómetros en el área de Gran Asunción mediante método análisis de componentes principales*” y el trabajo presentado en ella es de mi autoría.

AUTOR

FECHA

FIRMA

Resumen

La verificación de la eficiencia de una red de pluviómetros es fundamental a la hora de toma de decisiones en cuanto a la administración de recursos destinados a este. Este estudio busca el aumento de la eficiencia de una red de pluviómetros ubicada en la zona de Gran Asunción reduciendo la cantidad de pluviómetros mediante la utilización de análisis de componentes principales (PCA). Se analizaron los registros diarios de precipitaciones en el periodo de 2016 a 2017. Se utilizó el método de interpolación IDW para el rellenado de información faltante, obteniéndose un método de representar los patrones de lluvia para un mismo día. Se desarrollaron dos enfoques de selección de pluviómetros basados en PCA. El primer enfoque analizó la red de pluviómetros identificando los componentes que explican la mayor parte de la varianza del sistema completo. El primer enfoque dio como resultado un método capaz de reducir la red de 27 pluviómetros a una selección de 5. El segundo enfoque se basó en el análisis del primer componente principal y al uso de PCA reiteradamente en diversos subconjuntos creados por radios geográficos. El segundo enfoque dio como resultado un método capaz de clasificar a todos los pluviómetros con un sistema de puntajes y poseer una flexibilidad mayor al realizar la selección de pluviómetros. En ambos casos se realizaron evaluaciones de las selecciones comparándolos con el conjunto de pluviómetros totales y selecciones realizadas de forma aleatoria. El estudio presenta una metodología que resulta en información valiosa sobre los costos y beneficios de retener pluviómetros.

Palabras Clave: *Análisis de componentes principales PCA, Inverse Distance Weighting (IDW), Red de pluviómetros, Distribución de lluvia, Asunción, Paraguay.*

Zusammenfassung

Die Überprüfung der Effizienz eines Regenmessernetzes ist von grundlegender Bedeutung, wenn Entscheidungen über die Verwaltung der dafür bestimmten Ressourcen getroffen werden. Diese Studie zielt darauf ab, die Effizienz eines Pluviometer-Netzwerks im Gran Asunción zu erhöhen, indem die Anzahl der Niederschlagsmesser mithilfe der Hauptkomponentenanalyse (PCA) reduziert wird. Die täglichen Niederschlagsaufzeichnungen wurden im Zeitraum von 2016 bis 2017 analysiert. Die IDW-Interpolationsmethode wurde verwendet, um fehlende Informationen einzugeben und eine Methode zur Darstellung von Regenmustern für denselben Tag zu erhalten. Es wurden zwei Ansätze zur Auswahl von Regenmessern auf der Basis von PCA entwickelt. Der erste Ansatz analysierte das Regenmessernetzwerk und identifizierte die Komponenten, die den größten Teil der Varianz des Gesamtsystems erklären. Der erste Ansatz führte zu einer Methode, mit der das Netzwerk von 27 Regenmessgeräten auf eine Auswahl von 5 reduziert werden konnte. Der zweite Ansatz basierte auf der Analyse der ersten Hauptkomponente und der wiederholten Verwendung von PCA in verschiedenen durch geografische Radios erstellten Teilmengen. Der zweite Ansatz führte zu einer Methode, mit der alle Regenmesser mit einem Bewertungssystem klassifiziert werden können und die Flexibilität bei der Auswahl von Mehrfachmessgeräten bietet. In beiden Fällen wurden Auswertungen der Stichproben vorgenommen, wobei sie mit den Gesamtmengen der Regenmenge und den zufällig getroffenen Stichproben verglichen wurden. Die Studie präsentiert eine Methodik, die wertvolle Informationen über die Kosten und den Nutzen von Messgeräten liefert.

Schlüsselwörter: *Hauptkomponentenanalyse (PCA), Inverse Distanzwichtung (IDW), Regenmessnetz, Regenverteilung, Asunción, Paraguay.*

Índice

	Página
Resumen	III
Zusammenfassung	IV
Índice	V
Índice de figuras	VII
Índice de tablas	IX
Capítulo 1. Introducción	1
1.1 Antecedentes	6
1.1.1 Objetivo general	7
1.1.2 Objetivos específicos	7
1.2 Alcance	7
Capítulo 2. Metodología	9
2.1 Datos utilizados	10
2.2 Métodos de interpolación	11
2.2.1 Inverse distance weighting	11
2.3 Análisis de componente principal	12
2.3.1 Centrado y estandarizado	16
2.3.2 Matriz de covarianza	18
2.3.3 Autovalores y autovectores	18
2.3.4 Rotación	19
2.4 Selección de pluviómetros	20
2.4.1 Método de análisis completo	21
2.4.2 Método de análisis con radios	21
2.5 Evaluación de resultados	25
Capítulo 3. Resultados y análisis	26
3.1 Preparación de datos	26
3.1.1 Anderson–Darling test	26
3.2 IDW	27
3.3 PCA	29
3.3.1 Método de análisis completo	29
3.3.2 Método de análisis con radios.	34

3.4	Evaluación de métodos y comparación con selección aleatoria	37
3.4.1	Mapas de calor	37
3.4.2	Selección aleatoria de pluviómetros	41
	Capítulo 4. Conclusión	43
	Apéndices	45
A.	Distancias entre pluviómetros	45
	Referencias bibliográficas	46

Índice de figuras

	Página
1 Ubicación de pluviómetros en mapa de Gran Asunción.	2
2 Distribución de falta de datos de mediciones por lluvia en datos analizados	4
3 Distribución de valores de milímetros de lluvia registrados por todos los pluviómetros	5
4 Grafico de datos centrados de base de datos de dos dimensiones con sus respectivos componentes principales. Tomado de Smith (2002)	13
5 Gráfico de datos reconstruidos utilizando solo un componente principal. Tomado de Smith (2002)	14
6 Comparación de dispersión de datos de pares de pluviómetros más lejanos y más cercanos y con su reconstrucción de componentes principales.	15
7 Distribución de datos de par de pluviómetros en cada paso de la estandarización.	17
8 Circunferencias en donde fueron realizados PCA con radios para el par de pluviómetros más cercanos entre sí (7 y 26).	24
9 Circunferencias en donde fueron realizados PCA con radios para el par de pluviómetros más lejanos entre sí (9 y 25).	24
10 Distribución de valores de milímetros de lluvia registrados por todos los pluviómetros.	27
11 Distribución de las medias diarias de lluvia para los 350 días originales. . .	28
12 Distribución de las medias diarias de lluvia para los 42 días resultantes del recorte.	28
13 Mapa de calor realizado con resultados de la interpretación IDW utilizando todos los pluviómetros para un día analizado en el PCA.	29
14 Gráfico del Scree-Test para los 27 componentes principales.	31
15 Cargas de los primeros 2 componentes principales.	32
16 Mapa de calor creado utilizando las mediciones de 27 pluviómetros de un solo día excluido del análisis en PCA.	38
17 Mapa de calor creado utilizando las mediciones de 5 pluviómetros en un solo día excluido del análisis de PCA.	38
18 Mapa de calor creado utilizando las mediciones reales de 5 pluviómetros más las mediciones interpoladas en los otros 22 pluviómetros de un solo día excluido del análisis de PCA.	39
19 Mapa de calor creado utilizando las diferencias absolutas entre las interpolaciones con datos originales de la Fig. 16 y las interpolaciones de la Fig. 17.	40

20	Mapa de calor creado utilizando las diferencias absolutas entre las interpolaciones con datos originales de la Fig. 16 y las interpolaciones de la Fig. 18.	40
21	Distribución de medias de errores al regresar 1000 combinaciones aleatorias de 5 pluviómetros.	41
22	Distribución de medias de errores al regresar 1000 combinaciones aleatorias de 20 pluviómetros.	42

Índice de tablas

	Página
1 Componentes y sus respectivos autovalores con porcentaje de varianza explicado.	30
2 Matriz de carga de factores después de rotación Varimax	33
3 Cálculo de estadísticas de precipitación para 6 días para el PCA completo .	34
4 Rangos de pluviómetros	35
5 Cálculo de estadísticas de precipitación para 6 días para el PCA con Radios	36
6 Distancias en metros entre pluviómetros	45

Capítulo 1

Introducción

La utilización de instrumentos para registrar mediciones de lluvia ha servido, desde la antigüedad, para lograr un mejor entendimiento de las variaciones en el comportamiento de la lluvia. Un registro de estas variaciones es utilizado en actividades que dependen en gran medida de las cantidades de lluvia y sus variaciones como lo es por ejemplo la agricultura ([Zhang et al., 2012](#)), el estudio del cambio climático ([Trenberth, 2011](#)) y el estudio de potenciales desastres naturales como las inundaciones ([Christensen & Christensen, 2003](#)). En Asunción la combinación de factores como la falta de un desagüe pluvial y la presencia de basura en las calles han causado que el evento de una lluvia dé como resultado raudales que imposibilitan el tráfico vial y causen daños materiales a la propiedad pública ([ABC Color, 2018](#)).

Un análisis del comportamiento y variabilidad de las lluvias en Asunción se ve imposibilitado por la ausencia de mediciones detalladas y a largo plazo de la lluvia. Un registro de lluvias completo puede arrojar datos valiosos para potenciales proyectos de modelado, de mejoras en alertas meteorológicas o de optimización de manejo de suelo.

El proyecto “Variabilidad Espacial y Temporal de la Lluvia sobre el gran Asunción y sus Implicaciones (14-INV-189)” propuesta por la Universidad Paraguayo Alemana responde a la necesidad de datos suficientes para un análisis previsión del problema de raudales. El área de estudio del proyecto esta conformado por Gran Asunción, área metropolitana ubicada al oeste de Paraguay conformada por las ciudades de Asunción, Areguá, Capiatá, Fernando de la Mora, Itaguá, Lambaré, Limpio, Luque, Mariano Roque Alonso, Ñemby, San Antonio, San Lorenzo y Villa Elisa. Tiene una superficie de 929 km². Los datos analizados provienen de 27 pluviómetros ubicados dentro del área de Gran Asunción como puede verse en la Fig. 1. 25 de dichos pluviómetros corresponden a la investigación realizada por la Universidad Paraguayo Alemana mientras que los 2 restantes a la Dirección Nacional de Aeronáutica Civil (DINAC). Esto da como resultado una red de pluviómetros de densidad igual a un pluviómetro por cada 34,4 km²

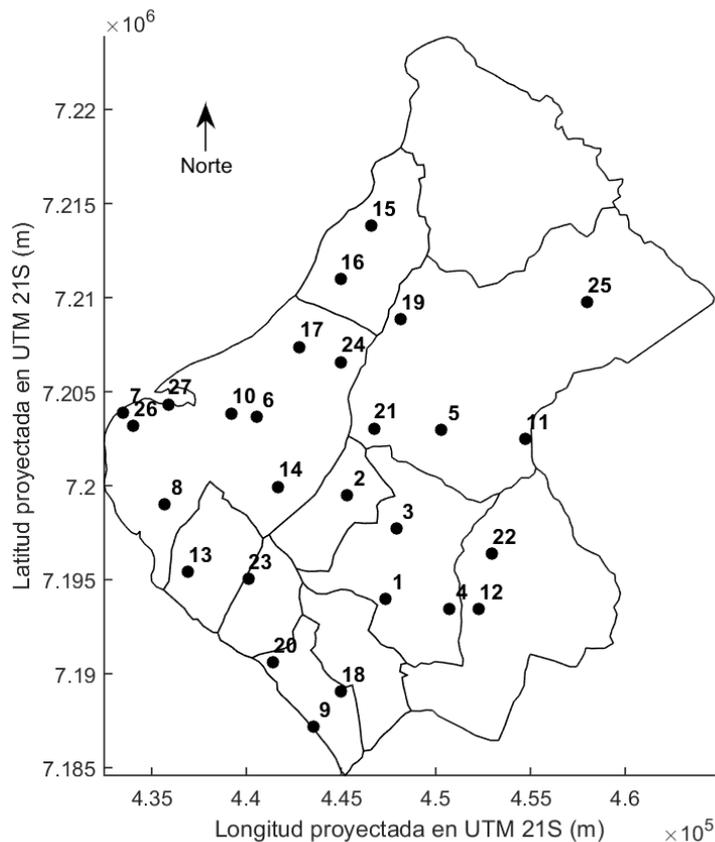


Figura 1: Ubicación de pluviómetros en mapa de Gran Asunción.

Una de las mayores dificultades del proyecto, es encontrar lugares adecuados para la instalación de los pluviómetros, ya que estos necesitan espacios abiertos libres de árboles o edificios, y la obtención de permisos para instalarlos y acceder a ellos a largo plazo, para la calibración, el mantenimiento de los medidores y la recolección de datos. Estas razones causan que la inversión individual escape el valor de compra y signifique también una inversión en recursos humanos. En general proyectos de investigación como este tienen a su disposición presupuestos limitados y por esto es necesario un análisis que garantice la eficacia y efectividad del proyecto en función del número de pluviómetros y su localización.

Una opción para optimizar la red es el empleo del análisis de componente principal (PCA) como un método estadístico con el fin de justificar la potencial eliminación de pluviómetros que aporten en menor medida información a la red de pluviómetros. El PCA es una herramienta estadística que se basa en la reducción de dimensiones de un conjunto de datos mientras se retiene la mayor parte de la variación original (Jolliffe, 2011). La idea es básicamente poder demostrar que ciertas dimensiones aportan más que otras a la variabilidad total del conjunto de datos y poder reducir la cantidad de datos extrayendo las dimensiones menos influyentes.

Los resultados del PCA son altamente dependientes de las decisiones tomadas en cada uno de sus pasos, especialmente en los pasos de preparación de los datos y en la selección de componentes principales. Esto posibilita la creación de más de un solo método de optimización de los pluviómetros y obtener métodos que reducen las variables con enfoques diferentes. Estos métodos deben ser validados y comparados entre sí para una mejor decisión en su selección.

El PCA necesita una base de datos completa para su correcta utilización. El periodo de tiempo de los datos corresponde a la de junio de 2016 a diciembre de 2017. En junio de 2016 se contaban con 15 pluviómetros, para septiembre de 2016 este número aumento a 23 y para marzo de 2017 los 27 pluviómetros ya habrían realizado alguna medición. Por esta razón la base de datos posee falta de datos que pueden verse en la Fig. 2. Es posible afrontar este problema mediante métodos de rellenado de información que utilizan herramientas como interpolación, extrapolación o regresión. Para este estudio se da énfasis en los métodos de interpolación.

La base de datos presenta un volumen de lluvia por día representada en la Fig. 3, se puede ver como una gran mayoría de días contiene valores de lluvia pequeños. El estudio se centra en el análisis de los días con volúmenes de lluvia suficientes para tener un impacto en el problema de lluvia.

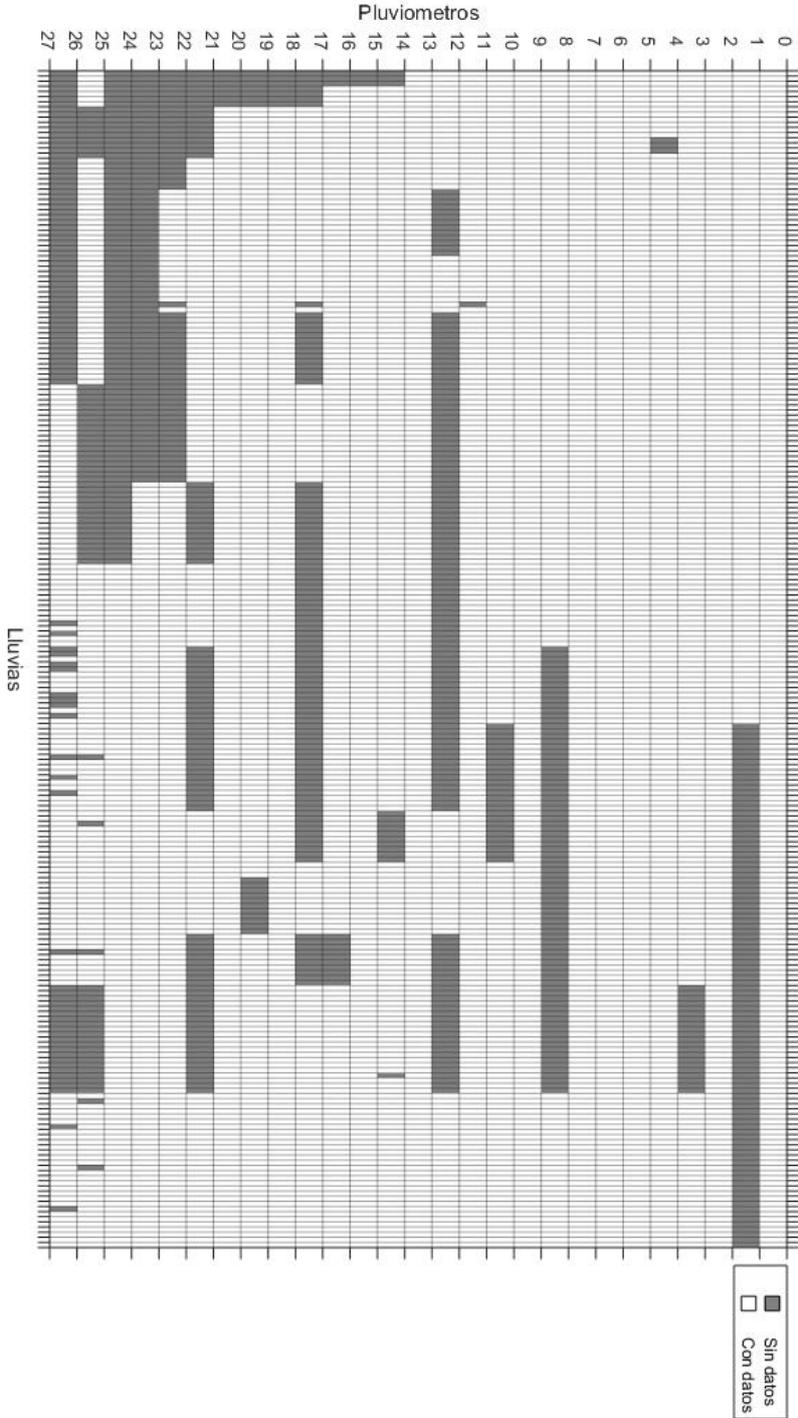


Figura 2: Distribución de falta de datos de mediciones por lluvia en datos analizados

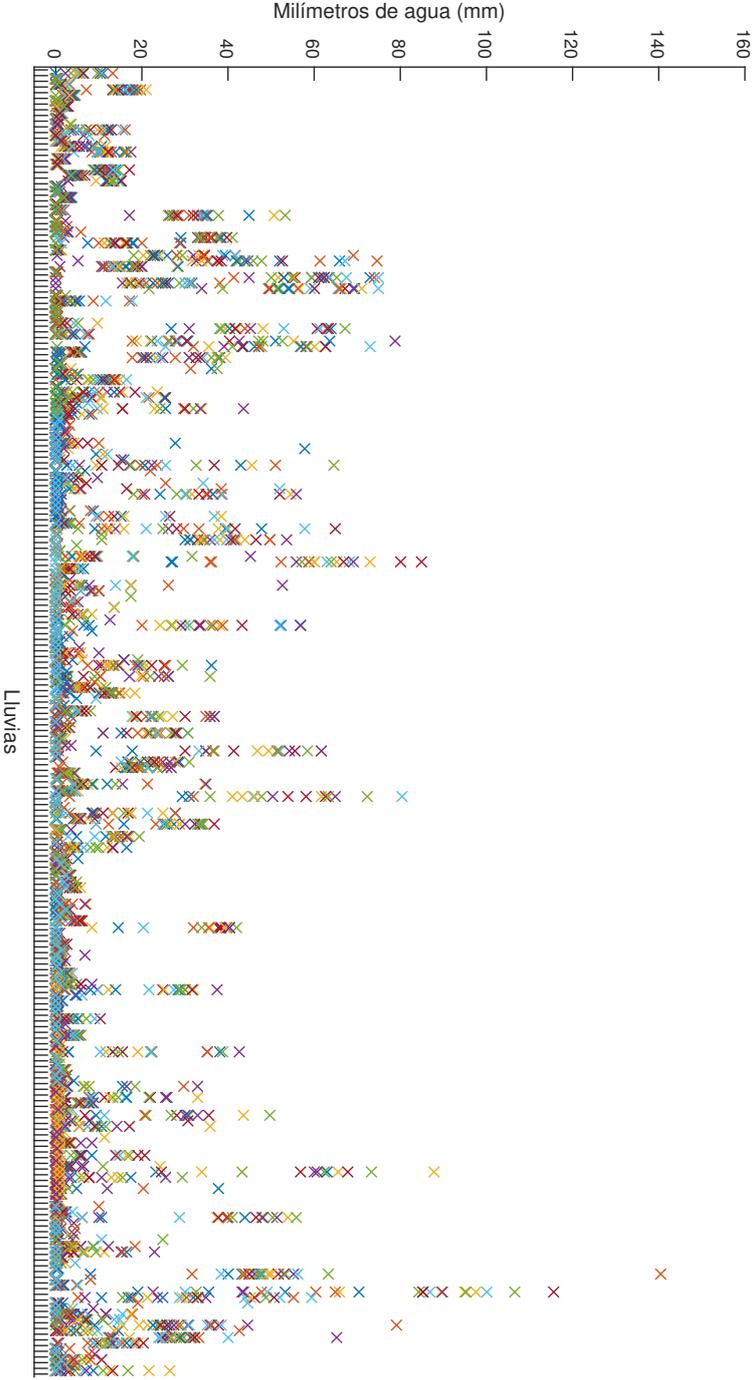


Figura 3: Distribución de valores de milímetros de lluvia registrados por todos los pluviómetros

1.1. Antecedentes

[Gangopadhyay et al. \(2001\)](#) aplicaron el PCA para analizar las variaciones de nivel de agua en el acuífero de Bangkok a través de mediciones tomadas de pozos que forman parte de la red de monitoreo de agua de la ciudad. Los autores mencionan la dificultad de mantener dicha red con recursos públicos y la posibilidad de la aparición de una restricción monetaria en el futuro que fuerce la eliminación de ciertos medidores. Con el fin de clasificar a los pozos en función de su importancia en la variabilidad total, se realizó el PCA en grupos diferentes de pozos basados en su disposición espacial (En base a radios de 10 km). Los pozos que aparecían más veces entre los responsables de la mayor variabilidad fueron llamados pozos principales. Se concluyó que la eliminación de ciertos pozos que no son principales no resulta en una pérdida grande de variabilidad en el sistema. Los resultados fueron que el sistema es capaz de utilizar solo estos pozos principales y mantener una medición adecuada de la variación de los niveles de agua. Retener 56 de los 86 pozos significaría una pérdida de solo 13 % de la variación del Sistema (35 % menos pozos). Los datos utilizados en el análisis fueron medidos en un lapso de 13 años.

[Ouyang \(2005\)](#) utiliza el PCA para evaluar las estaciones de monitoreo de calidad de agua sobre el río St. Johns en Florida. Las estaciones recolectan información sobre muchas características diferentes del agua del río como la temperatura, el color, la salinidad, la acidez, los niveles de minerales, etc. Los datos analizados fueron de un lapso de 3 años. Se utiliza un PCA seguido de una rotación varimax para la interpretación de los resultados. Se pudo concluir que existe potencial de mejora de eficiencia eliminando 3 de las 22 estaciones. Se demostró que la pérdida de información es mínima dentro de estos 3 años, mientras que se sugiere una validación con un periodo de tiempo más extenso.

En [Gangopadhyay et al. \(2001\)](#) se menciona un método que utiliza programas de simulación geoestadísticos para el relleno de información faltante en mediciones de nivel de agua dentro del acuífero de Bangkok a través de los años. Esto fue utilizado para su posterior utilización en un PCA. Para datos de lluvias ([Dirks et al., 1998](#)) realizaron una comparación utilizando validación cruzada para la predicción de mediciones de lluvia en una red de pluviómetros de alta densidad. Concluyen que la aplicación de IDW es la mejor opción en comparación con métodos como el de Thiessen y métodos más complejos como el Kriging. De igual manera [Keblouti et al. \(2012\)](#) realizaron el mismo análisis para una red de baja densidad teniendo conclusiones similares, aunque presentando niveles de error mayores. Comparaciones de métodos de interpolación con un nivel de profundidad más alto fueron revisadas en [Li & Heap \(2014\)](#). El funcionamiento general del método para su

aplicación fue consultado en [Shepard \(1968\)](#).

1.1.1. Objetivo general

Realizar una selección de pluviómetros más importantes a retener para el proyecto 14-INV-189, basada en la utilización de análisis de componentes principales.

1.1.2. Objetivos específicos

- ◇ Creación de un sistema de clasificación de localizaciones de pluviómetros en base a resultados del PCA.
- ◇ Estimación de error en la variabilidad producida por la reducción de pluviómetros.
- ◇ Validación y comparación de resultados con selección de pluviómetros de forma aleatoria.

1.2. Alcance

Este estudio busca proponer una metodología adecuada utilizando el PCA para ayudar a la toma de decisiones en la instalación de pluviómetros y su eficiencia en el registro de datos sin tener en cuenta las posibles conclusiones posteriores al análisis de dichos datos.

Se busca proponer una metodología adecuada utilizando el PCA para ayudar a la toma de decisiones en la instalación de pluviómetros y su eficiencia en el registro de datos sin tener en cuenta las posibles conclusiones posteriores al análisis de dichos datos.

El análisis se desarrolla en un periodo de tiempo de 1 año y medio en comparación a estudios similares que utilizan datos recolectados en periodos de tiempo que abarcan varios años e incluso décadas. Al realizar la selección de días a ser analizados la cantidad de mediciones es todavía menor. Se debe tener en cuenta que los resultados serán más confiables al utilizarse una muestra mayor de mediciones.

La utilización del PCA también implica un análisis sobre las varianzas de cada variable y los resultados son altamente dependientes del procesamiento previo de los datos. Las valoraciones de los pluviómetros resultantes de los métodos aplicados están sujetos a interpretaciones de la importancia de la varianza para el sistema. Se espera valorar a los pluviómetros de una forma que el sistema mantenga la mayor cantidad de variabilidad residual.

Capítulo 2

Metodología

Para analizar la red de pluviómetros se tiene como información el registro de mediciones de precipitaciones de los 27 pluviómetros de la zona de Gran Asunción durante el periodo de tiempo que va desde mitades de 2016 a finales de 2017. Este registro cuenta con días de muy baja intensidad de lluvias que no serán incluidos en el análisis. El registro presenta también faltas de mediciones en ciertos pluviómetros. Para la adecuada aplicación del PCA la información faltante es rellenada con el método de interpolación de datos de Inverse Distance Weighting (IDW).

En este estudio se propone adaptar el PCA y utilizar la interpretación de sus resultados como base para la toma de decisiones sobre la cantidad de pluviómetros a utilizar en recolección de datos de lluvia y así poder garantizar que el número de los pluviómetros sea el adecuado para explicar el comportamiento de la lluvia en la zona de Gran Asunción. En el estudio se proponen dos enfoques distintos para la utilización del PCA. El primero toma el conjunto de datos como un todo y realiza un PCA para mantener la mayor cantidad de variabilidad del mismo. La elección de que pluviómetros mantener se basa en los resultados de este único PCA. El segundo método toma en cuenta la separación espacial de los pluviómetros y los divide en muchos subgrupos y realiza un PCA para cada subgrupo y con los resultados se crea un sistema de puntuación que es la base para la elección de que pluviómetros mantener.

Las combinaciones de pluviómetros seleccionados para su retención son analizadas comparando sus valores estadísticos con los valores estadísticos originales de la red de pluviómetros completas. Para estas comparaciones se utilizan datos omitidos en los PCA desarrollados. Las comparaciones sirven como visualización de pérdida de información en el sistema original. Para finalizar el estudio se desarrollan comparaciones de estas combinaciones con combinaciones aleatorias. Estas comparaciones se realizan en base al potencial para predecir valores de lluvias en puntos donde se eliminaron pluviómetros y para esto se utilizan regresiones e interpolaciones.

2.1. Datos utilizados

Los pluviómetros registran las lluvias en milímetros de lluvia (mm) lo cual equivale a un mm de altura en un espacio de un m^2 a un litro de agua por metro cuadrado ($l/m^2 = dm^3/m^2 = 0,001m^3/m^2 = 0,001m$). Estos pluviómetros pueden registrar un valor de $0 mm$, que equivale a ninguna lluvia, un valor mínimo de lluvia de $0,2386 mm$ y valores continuos mayores al mínimo. La base de datos contiene un registro diario de los milímetros de lluvia medidos por los 27 pluviómetros en donde no son guardados los días en que no se han registrado valores de lluvia (días en donde todos los pluviómetros registran $0 mm$). Inicialmente se poseen datos para 350 días. Los volúmenes de lluvia registrados en estos días están representados gráficamente en la Fig. 3. Se puede observar como una gran porción de las medidas están en el rango de $0 mm$ a $20 mm$, el promedio de lluvia para todas las mediciones es de los $4 mm$.

Ya que existen en la base de datos días en donde existen mediciones de valores muy bajos que afectan la variabilidad sobre las pequeñas lloviznas mientras que el estudio busca analizar las lluvias grandes, se definió como día de lluvia a “Los días en los cuales la media de lluvia sea mayor a $10 mm$ ”. Así se pueden eliminar días en los que se registran lluvias muy débiles que añadirían ruido a los datos debido a la precisión de los medidores. También se utilizó el *Grubbs’ test for outliers* Grubbset al. (1950), asumiendo la normalidad de la muestra, para poder desechar valores de mediciones que estén muy por encima de los valores medidos en el mismo día por otros pluviómetros. Los outliers pueden deberse a errores en los medidores y causar distorsiones en cálculos de media y de desviación estándar. Estos outliers se tomaron como errores y en su lugar se registraron como datos faltantes.

En la Fig. 2 se puede observar distribución de datos faltantes en cada pluviómetro a través del tiempo en la base de datos. Estas celdas vacías pueden deberse a errores técnicos, problemas físicos con los pluviómetros (invasión de abejas, basuras, caída de los equipos, etc) o simplemente a que todavía no estaban instalados estos equipos. Para su análisis estas observaciones faltantes fueron rellenadas utilizando Inverse distance weighting (IDW).

Para los análisis se toman en cuenta el 85% de los días de lluvia resultantes. Esto se debe a que el 15% restante se utilizó como una muestra externa a los análisis en donde se evaluaron los resultados de estos.

2.2. Métodos de interpolación

Existen muchos métodos de interpolación espacial que pueden ser clasificados en los geoestadísticos univariados (Simple kriging, Ordinary kriging, etc.), los geoestadísticos multivariados (Universal kriging, Cokriging, etc.), los no geoestadísticos (Nearest neighbours, Inverse distance weighting, etc.) y los métodos combinados. Para la selección de un método de interpolación es importante observar la naturaleza de los datos a analizar y sus características.

Al analizar y trabajar con datos de lluvia, [Dirks et al. \(1998\)](#) realizaron una comparación de distintos métodos de interpolación y comparando la tasa de error de cada uno. Concluyeron que para redes de alta densidad (un pluviómetro por cada $0,37 \text{ km}^2$), debido a su alta complejidad, los métodos de kriging no son mejores que métodos simples como Nearest neighbours y Inverse distance weighting. Cambien concluyen que el IDW produce campos de lluvia relativamente realistas y permite análisis de variabilidad posteriores.

Una comparación similar entre IDW y Ordinary kriging demostró resultados similares para una red de pluviómetros de baja densidad en la ciudad de Annaba, Nigeria por [Keblouti et al. \(2012\)](#). Se analizó una red de pluviómetros de baja densidad de un pluviómetro por cada $142,9 \text{ km}^2$ con datos de un periodo de tiempo de 29 años. Se concluyó que aunque los errores de interpolación son más altos que una red de alta densidad el método de IDW es el mejor para representar la distribución de lluvia.

En comparación la red de pluviómetro estudiada es de un pluviómetro por cada $34,4 \text{ km}^2$

2.2.1. Inverse distance weighting

El Inverse distance weighting (IDW) es un método de interpolación de datos espaciados irregularmente que utiliza valores de puntos muestreados para estimar los valores en puntos no muestreados. El IDW se caracteriza por realizar un promedio ponderado de los valores muestreados y utilizar la inversa de las distancias de los puntos muestreados al punto a estimar como ponderaciones. En la Ec. (1) se representa de forma general el método de IDW como una función de interpolación para un espacio de dos dimensiones, donde z_p es el valor a estimar en el punto p , (x, y) son las coordenadas del punto p , n es

el número de puntos muestreados disponibles, z_i es el valor de un punto muestreado, d_i es la distancia z_i a z_p y u es el parámetro de potencia.

$$z_p(x, y) = \begin{cases} \frac{\sum_{i=1}^n z_i * d_i^{-u}}{\sum_{i=1}^n d_i^{-u}} & \text{si } d_i \neq 0, \\ z_i & \text{si } d_i = 0, \end{cases} \quad (1)$$

De esta forma se obtuvieron valores para un campo de lluvias interpolado. Este campo tiene un tipo de superficie que dependerá de la elección del parámetro de potencia u . Según datos experimentales valores mayores a 2 producen superficies planas y saltos de valores muy precipitados pudiéndose llegar a un caso similar al de los polígonos de Thiessen, mientras valores menores a 2 dan más peso a puntos más lejanos. Con un valor de $u = 2$ se obtiene resultados satisfactorios para la descripción de una superficie de valores interpolados (Shepard, 1968).

2.3. Análisis de componente principal

Al realizar mediciones de lluvia, los datos conseguidos por los pluviómetros en un mismo día son generalmente de lluvias de una misma nube. Se espera que parte de los datos estén correlacionadas entre sí. Esto significaría que parte de la información medida por un pluviómetro este también almacenada dentro de los pluviómetros restantes. El objetivo del análisis es el estudio de la correlación de las variables independientes (en este caso pluviómetros) para medir el potencial de reducción del número de estas variables sin perder gran parte de la información del sistema.

Los pasos del PCA que son importantes para este estudio son los siguientes:

1. Centrado y estandarizado
2. Cálculo de matriz de covarianza
3. Cálculo de autovalores y autovectores
4. Rotación de componentes.

El análisis termina con una matriz de componentes principales, estos serán vectores de tamaño igual a la cantidad de variables de la matriz de datos originales. El primer componente es un vector unitario cuya dirección es la que explica la mayor cantidad de varianza posible del sistema. El segundo componente será un vector unitario ortogonal al primer componente, o sea no correlacionado a este, que explique la mayor cantidad de varianza posible y se sigue así hasta tener un componente por cada variable, todos ortogonales entre sí. Se busca expresar al sistema de esta forma con la esperanza de que las varianzas de una gran parte de los componentes principales sean tan minúsculas que puedan ser obviados (Manly, 1986), mientras que el sistema pueda ser explicado satisfactoriamente con una pequeña porción de los componentes principales. Si los datos no están correlacionados entre sí, el análisis no funciona ya que este busca explicar cómo las variables aumentan o disminuyen en conjunto.

En la Fig. 4 se graficaron, para una base de datos de dos variables con datos centrados, los dos componentes principales. Se puede ver como el primer componente principal describe la dirección hacia donde los datos crecen en conjunto mientras que el segundo componente abarca la variación en sentido ortogonal al primer componente.

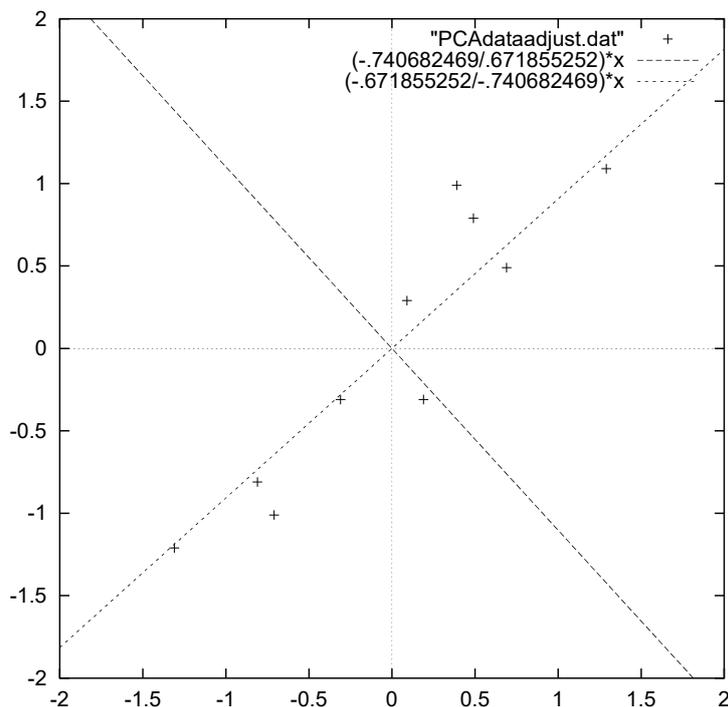
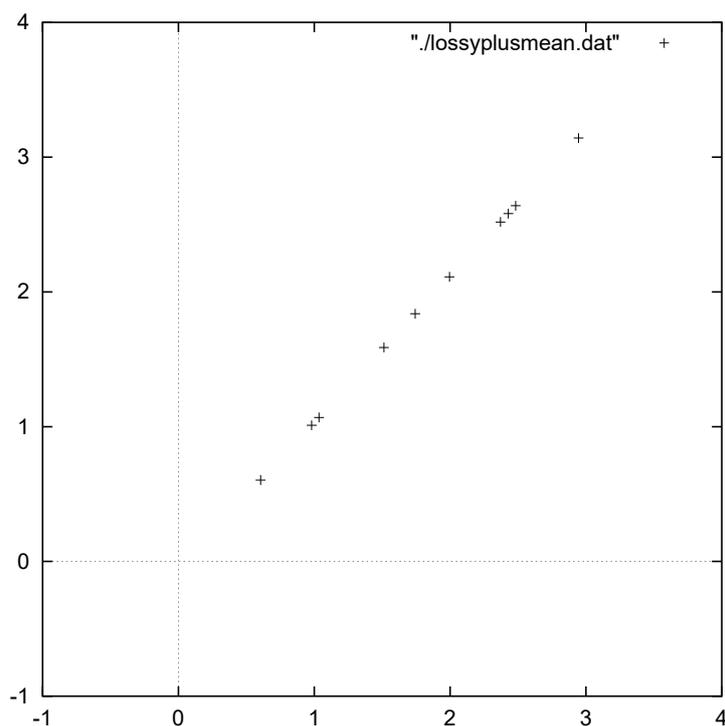


Figura 4: Grafico de datos centrados de base de datos de dos dimensiones con sus respectivos componentes principales. Tomado de Smith (2002)

Cada componente principal está conformado por sus coeficientes en cada dimensión del espacio del sistema de datos originales. Dicho de otra forma, cada componente principal está formado por sus proyecciones en cada variable. De esta forma es posible saber cuánto

una variable influye dentro de cada componente principal. Si una variable es dominante (con un valor absoluto alto) dentro de un componente principal se puede concluir que esta es la que mejor explica la varianza que a su vez es explicada por ese componente principal. Así se busca definir cuáles son las variables más importantes para la base de datos.

Para entender el funcionamiento del PCA se muestran en la Fig. 6 los efectos de una comprensión de datos utilizando PCA. En la figura fueron comparados las mediciones de dos pares de pluviómetros, los más alejados entre sí a 27 km de distancia (véase Fig. 6a) y los más cercanos entre sí a 900 m de distancia (véase Fig. 6c). Con el PCA es posible analizar los componentes principales y retener una cantidad menor de ellos. Por ejemplo se puede representar todos los datos en función de su proyección en un solo componente, obteniéndose una base de datos en una sola dimensión. De esto se puede volver a una base de datos de mas dimensiones perdiendo gran parte de la información original. En la Fig. 5 se observa un ejemplo en el que se realiza una reconstrucción de los datos no centrados de la Fig. 4 utilizando solo un componente. En la figura se realizó



pluviómetros más lejanos (véase Fig. 6b) que en los más cercanos (véase Fig. 6d). Estos resultados son explicados por la razón de que el PCA trabaja mejor en conjuntos de datos más correlacionados entre sí. Notese que no se realizó una extracción de variables (pluviómetros) sino de componentes principales.

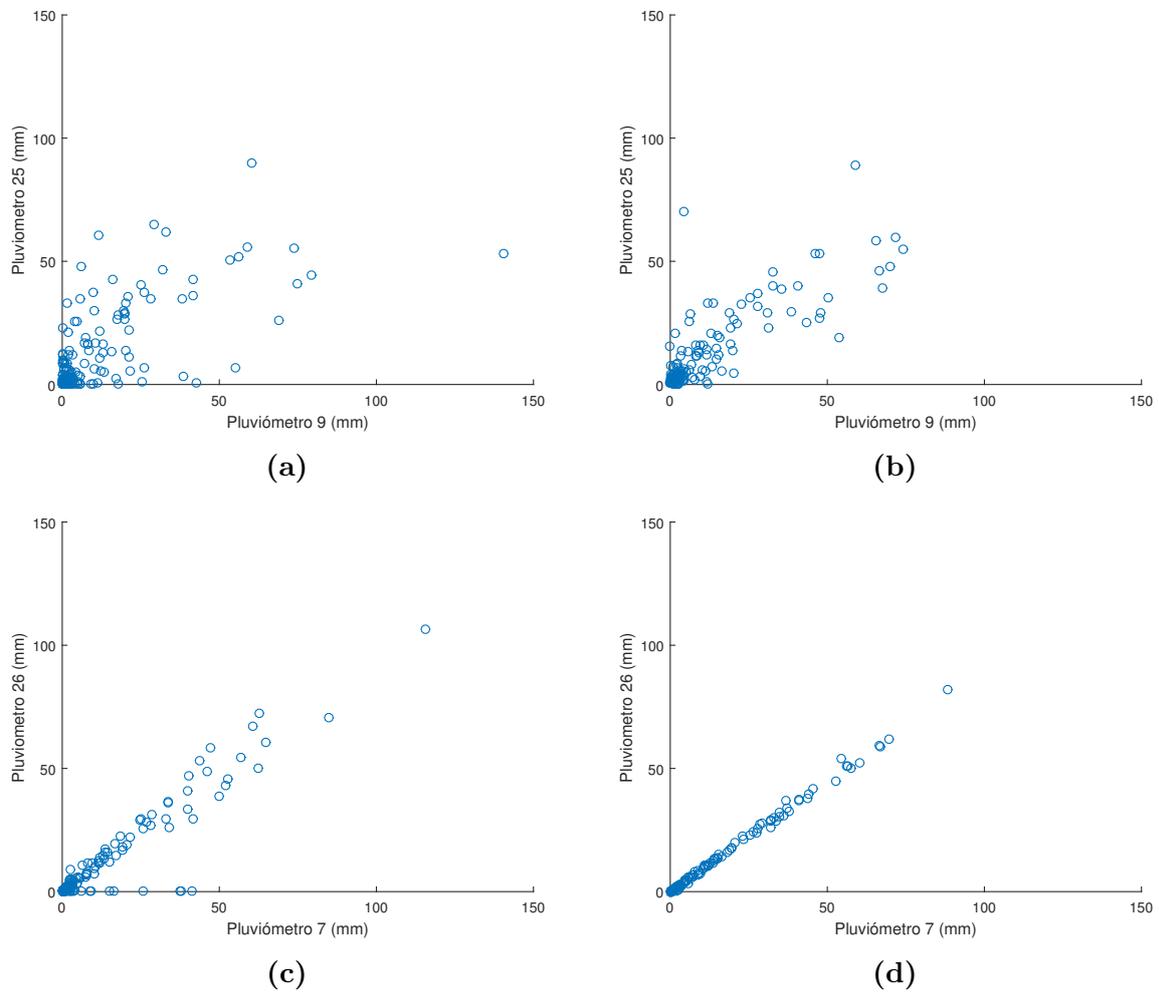


Figura 6: Comparación de dispersión de datos de pares de pluviómetros más lejanos y más cercanos y con su reconstrucción de componentes principales. **a)** Pluviómetros más lejanos (27 km). **b)** Reconstrucción de pluviómetros más lejanos. **c)** Pluviómetros más cercanos (27 km). **d)** Reconstrucción de pluviómetros más cercanos.

2.3.1. Centrado y estandarizado

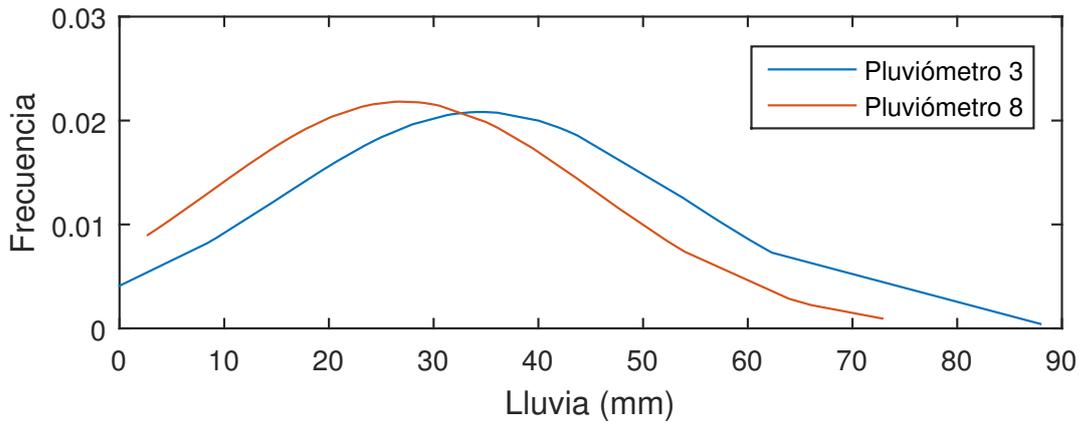
Si X es una matriz de tamaño $n \times p$ de vectores P de mediciones de donde n es la cantidad de días en donde existen registros de lluvia y p la cantidad de variables independientes (pluviómetros). Generalmente, se expresa esta matriz como la desviación de cada medición con respecto al promedio de las mediciones en cada pluviómetro. Esto se consiguió realizando la diferencia entre cada medición y el promedio \bar{p} de las mediciones de su pluviómetro dejando una matriz como en la Ec. (2).

$$\tilde{X} = \begin{pmatrix} x_{1,1} - \bar{p}_1 & x_{1,2} - \bar{p}_2 & \cdots & x_{1,p} - \bar{p}_p \\ x_{2,1} - \bar{p}_1 & x_{2,2} - \bar{p}_2 & \cdots & x_{2,p} - \bar{p}_p \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} - \bar{p}_1 & x_{n,2} - \bar{p}_2 & \cdots & x_{n,p} - \bar{p}_p \end{pmatrix} \quad (2)$$

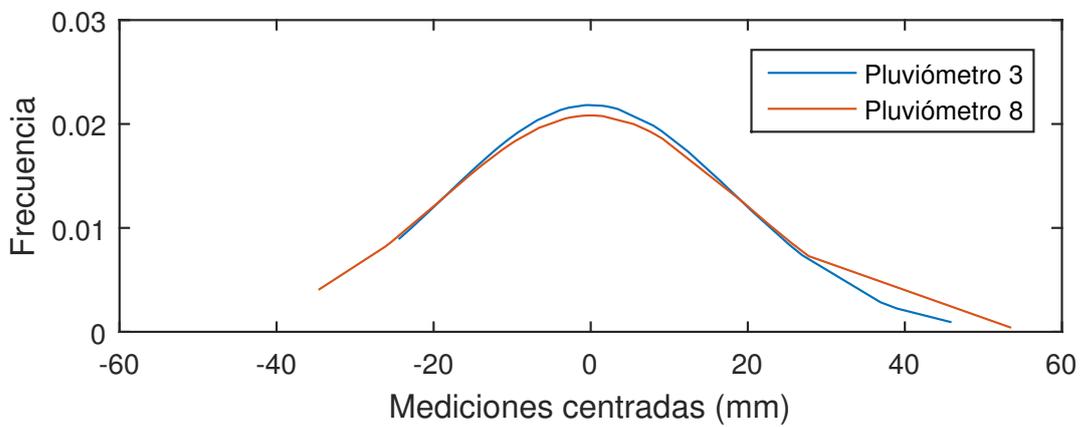
Al utilizar la matriz de mediciones centradas en el PCA, se priorizan las variables con mayor varianza y estas tendrán mayor influencia sobre los componentes principales. Para evitar que solo las variables con mayor varianza sean representadas se procedió a extraer la desviación estándar de cada variable. Esto se consigue dividiendo cada medición por la desviación estándar de su variable σ_p dejando una matriz como en la Ec. (3). El resultado fueron variables con varianza y desviación estándar igual a 1.

$$\tilde{X}_\sigma = \begin{pmatrix} \tilde{x}_{1,1}/\sigma_{p_1} & \tilde{x}_{1,2}/\sigma_{p_2} & \cdots & \tilde{x}_{1,p}/\sigma_{p_p} \\ \tilde{x}_{2,1}/\sigma_{p_1} & \tilde{x}_{2,2}/\sigma_{p_2} & \cdots & \tilde{x}_{2,p}/\sigma_{p_p} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_{n,1}/\sigma_{p_1} & \tilde{x}_{n,2}/\sigma_{p_2} & \cdots & \tilde{x}_{n,p}/\sigma_{p_p} \end{pmatrix} \quad (3)$$

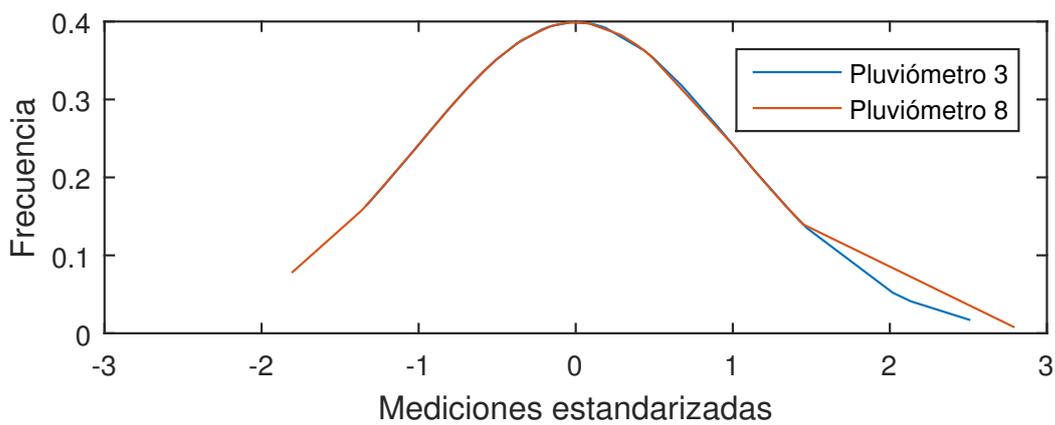
Las diferencias de estas matrices y el proceso de estandarización se puede ver gráficamente en la Fig. 7 para los pluviómetros 3 y 8. Se eligió este par aleatoriamente. Con estos pasos la matriz quedó estandarizada.



(a)



(b)



(c)

Figura 7: Distribución de datos de par de pluviómetros en cada paso de la estandarización de datos. **a)** Distribución de datos sin estandarización. **b)** Distribución de datos luego del centrado **c)** Distribución de datos final después del proceso de estandarización.

2.3.2. Matriz de covarianza

Después de la estandarización se obtuvo una matriz que contiene datos sobre como varían las mediciones con respecto a las medias para cada variable independientemente.

Para medir la variabilidad de las mediciones entre las variables, se calculó la covarianza entre dos variables. Se creó una matriz C en donde se computan todas las combinaciones de variables como se ve en la Ec. (4). Nótese que la diagonal de C esta formada por las covarianzas de variables consigo mismas lo que es lo mismo que la varianza de estas variables, y para la matriz estandarizada iguales a 1. Entonces se obtuvo una matriz en donde el primer vector columna indica la cantidad en que la primera variable está relacionada con todas las otras variables incluyéndose a sí misma, mientras que la segunda indica las relaciones de la segunda y así hasta la última variable.

$$C = \begin{pmatrix} cov(P_1, P_1) & cov(P_1, P_2) & \cdots & cov(P_1, P_p) \\ cov(P_2, P_1) & cov(P_2, P_2) & \cdots & cov(P_2, P_p) \\ \vdots & \vdots & \ddots & \vdots \\ cov(P_p, P_1) & cov(P_p, P_2) & \cdots & cov(P_p, P_p) \end{pmatrix} \quad (4)$$

2.3.3. Autovalores y autovectores

Una vez calculada la matriz de covarianzas se procedió a calcular la matriz de autovalores y autovectores de esta. Para cada dimensión del sistema se obtuvieron la misma cantidad de autovectores y cada autovector posee su único autovalor o viceversa. Los autovalores y autovectores están asociados con matrices cuadradas y son el producto de una auto-descomposición. Un auto vector de la matriz A es un vector u que cumpla con la Ec. (5)

$$Au = \lambda u \quad (5)$$

En donde λ es un escalar que representa al autovalor del autovector. Para calcular todos los autovectores y autovalores de una matriz cuadrada se define a U como la matriz que tiene por columnas a los autovectores de A , y a Λ como la matriz diagonal que tiene como diagonal los autovalores de A . Al resolver la (6) se obtiene la auto-descomposición.

$$AU = U\Lambda \quad (6)$$

El resultado de estos cálculos dan a conocer información importante del sistema ya que cada autovector es un componente principal. Para obtener la matriz de componentes principales final se ordenó la matriz de autovectores en orden descendente de acuerdo al valor del autovalor de cada autovector. De esta manera el primer vector columna de esta matriz es el primer componente principal y el segundo vector columna el segundo componente principal y así sucesivamente. Es posible calcular también la proporción de varianza explicada por cada componente principal si se divide al autovalor del componente sobre la cantidad de variables del sistema.

Esta matriz de autovectores ordenada y sus correspondientes autovectores es el resultado final del PCA. Las conclusiones posteriores a la utilización del PCA son el resultado de las interpretaciones sobre los valores de los coeficientes de los componentes principales. Sin embargo, existen pasos posteriores que tradicionalmente siguen al PCA y ayudan a la interpretación de los resultados.

Ya que el PCA es utilizado como un método de comprimir información uno de sus fines es determinar cuáles componentes principales son irrelevantes para explicar la varianza del sistema. [Kaiser \(1960\)](#) sugiere que, para un análisis sobre variables estandarizadas, los componentes principales cuyos autovalores sean menores a 1 pueden considerarse como no importantes ya que explican menos varianza que la que posee cada variable. O “Conservar solamente aquellos factores cuyos autovalores son mayores a la unidad” a la cual se refiere como la regla de Kaiser. Otro método es la utilización de un scree-test para representar gráficamente a los autovalores y elegir los más significativos de una manera tal que se retienen los componentes principales anteriores a que haya una disminución de repentina de la pendiente formada por los autovalores, dicho de otra forma, se forme un ‘codo’ en la gráfica. Otra forma de interpretar el test es trazando una recta formada por los dos últimos autovalores y retener los componentes que queden arriba de esta ([Cattell, 1966](#)).

2.3.4. Rotación

Aunque ya no forma parte del PCA generalmente se realizó una rotación sobre el espacio de componentes principales elegidos a retener. Esto se realizó como un método de facilitar las interpretaciones del PCA, ya que cuando se retienen más de un componente la interpretación de las variables dominantes es relativamente difícil. El tipo de rotación más utilizada es la rotación Varimax. La rotación de tipo Varimax tiene las ventajas de que mantiene la ortogonalidad de los componentes principales y mantiene la proporción

de varianza explicada del sistema original en el nuevo espacio de dimensiones reducidas. También tiende a llevar a los coeficientes de cada componente a su valor máximo posible o a cero lo que facilita la decisión de cual variable es ‘importante’ (Jolliffe, 2011). La rotación busca una combinación lineal de los componentes en donde se maximice la varianza de las cargas, esto puede expresarse como en la Ec. (7).

$$V = \sum (q_{p,l}^2 - \overline{q_{p,l}^2})^2 \quad (7)$$

En donde se busca maximizar V , $q_{p,l}^2$ es el cuadrado de la carga de la variable p del componente l y $\overline{q_{p,l}^2}$ es el promedio de los cuadrados de las variables.

Se refiere a los nuevos componentes rotados como ‘factores’. La varianza explicada por estos factores está más equilibradamente distribuida que en los componentes principales. Este tipo de rotación está relacionado con el análisis de factores y por ello se confunde al análisis de factores con el PCA pero estos dos análisis son distintos en concepto y no deberían ser confundidos y se podría llamar a esta técnica como un PCA seguido por una rotación Varimax.

2.4. Selección de pluviómetros

Para la selección de pluviómetros y la decisión de mantener cada uno de ellos se utilizó al PCA de dos formas diferentes. El primer método consistió en un PCA que analice todos los pluviómetros como un conjunto analizando los componentes principales, aplicando la regla de Kaiser y una rotación varimax para decidir cuales pluviómetros explican satisfactoriamente la varianza del conjunto de pluviómetros completo. El segundo método consistió en un sistema de puntuación de pluviómetros realizando PCA con pluviómetros dentro de distintos radios y clarificarlos dependiendo de la cantidad de veces que son analizados y la cantidad de veces que el pluviómetro es el dominante dentro del radio.

2.4.1. Método de análisis completo

Se realizaron todos los pasos del PCA para las mediciones de lluvia después de su selección y rellenado por interpolación. Para este método es útil pensar a toda el área de gran asunción como un conjunto de pluviómetros y el objetivo del análisis sería definir los pluviómetros que mejor explican la variación de este conjunto con cada lluvia. No se tuvieron en cuenta las separaciones geográficas que existen entre pluviómetros y se supuso que cada día de lluvia afecta a todos estos. Se ignoró el radio de influencia de las lluvias.

Primero se realizó la centralización y estandarización de las mediciones. Sobre estos se calculó la matriz de covarianza para luego realizar el cálculo de la matriz de autovalores y la de autovectores. Después del ordenamiento según autovalor y autovector se analizaron los componentes principales. Se aplicó la regla de káiser y el scree-test para tomar la decisión de cuantos componentes principales retener. Se realizó una rotación varimax sobre el espacio de componentes principales retenidos y se analizaron los resultados.

Aunque dependiendo de los resultados es posible elegir los pluviómetros antes de la rotación se prefiere realizar la rotación para mejorar la distribución de varianza explicada de los componentes principales y obtener una interpretación más fácil. En la matriz de coeficientes de los factores luego de la rotación se eligieron las variables con el coeficiente absoluto mayor (para este análisis no es importante el signo de este ya que solo indica la dirección del vector unitario). Es posible elegir a más de una variable por factor dependiendo de las interpretaciones que estos resultados pueden dar. Se espera que este método arroje una cantidad de variables mucho menor que la cantidad de variables original.

2.4.2. Método de análisis con radios

En este método se realizó el PCA reiteradamente para distintas selecciones de pluviómetros y se impusieron algunas condiciones para estas selecciones. Este método se basó en la idea de que pluviómetros muy lejanos no poseen correlación suficiente para que merezcan ser analizados dentro de un mismo PCA (Esto se puede ver gráficamente en la Fig. 6) y que pluviómetros con menos correlación con el resto son más importantes para una red de pluviómetros ya que aportan información única al sistema. En base a estas suposiciones se creó un método de puntuación de pluviómetros en base a circunferencias con centro en pluviómetros en donde se implementó un PCA a los pluviómetros internos y se

registró el pluviómetro más influyente para luego repetir el proceso en un radio mayor. El pluviómetro más influyente fue aquel con el mayor coeficiente absoluto sobre el primer componente principal que generalmente explica de 50 a 90 % de la varianza del sistema. Se debe tener en cuenta que este método solo tomó en cuenta al primer componente principal para su valoración por la razón de que este se considera que es el componente que explica el comportamiento general de la lluvia. Sin embargo, el método puede ser alterado para incluir más componentes y una rotación varimax si es necesaria. Este método está inspirado en y mantiene similitud con el método utilizado por [Gangopadhyay et al. \(2001\)](#) para la valoración de mediciones de nivel de agua en pozos.

Antes de realizar los cálculos fue necesario definir la cantidad de radios a utilizar por cada centro y la graduación en la que estos aumenten. Para esto se estudió la matriz de distancias de los pluviómetros. Los pluviómetros más cercanos se encuentran a un poco menos que 900 metros y el segundo par más cercano a un poco más que 1300 metros. En base a esto se eligió como primer radio el valor de 1000 metros y se aumentó el tamaño de este gradualmente de a 1000 metros. El mayor valor fue en principio de 27 kilómetros ya que los pluviómetros que están más alejados unos del otro se encuentran a un poco menos que 27 kilómetros. Pero en base a resultados experimentales y a la presuposición sobre la correlación entre pluviómetros muy lejanos se decidió elegir como máximo radio a 10 kilómetros. Los resultados sugirieron que después de este radio los pluviómetros con mayor puntaje mantienen dominancia y las puntuaciones más bajas se vuelven más pequeñas lo que dificulta la posterior separación en niveles de los pluviómetros.

Primero se eligió un pluviómetro como centro de la circunferencia de aplicación del PCA, se realizó el PCA con los pluviómetros dentro de este radio y se registró cuales pluviómetros fueron analizados y cuál de ellos es el dominante dentro del primer componente principal. Es importante considerar que en los casos en que en los radios solo exista un pluviómetro este será el dominante, lo que aumenta su puntaje final, que significa que hace que pluviómetros apartados tengan un mejor puntaje acorde a una de las presuposiciones. Se volvió a repetir el mismo proceso para un radio mayor hasta cubrir todos los rangos de radios definidos. Esto se repitió nuevamente cambiando el pluviómetro utilizado como centro, llegando a utilizar a todos los pluviómetros como centro de las circunferencias. Al terminar estos pasos se obtuvo una lista de puntuaciones para cada pluviómetro que ofrece información sobre la cantidad de veces que un pluviómetro ha sido analizado y las veces que cada uno ha sido el dominante.

Debido a la manera en que es estandarizada la información antes del PCA, en los casos en que se analicen un par de pluviómetros ambos pluviómetros tendrán las mismas cargas en los componentes principales y no es posible realizar una decisión. Para estos casos se definió la dominancia en función al pluviómetro con mayor varianza en sus mediciones.

En la Fig. 8 puede verse una descripción gráfica del método de PCA con radios en los pluviómetros más cercanos entre sí. En cada circunferencia en la figura fue realizado un PCA con los pluviómetros que se encuentran dentro de cada circunferencia. Se observa como este par de pluviómetros comparten los mismos otros pluviómetros en muchas de sus circunferencias. En la Fig.9 puede verse el caso análogo para el par de pluviómetros más lejanos entre sí. En este caso estos pluviómetros no comparten ningún otro pluviómetro en ninguno de los PCA en los que estos son analizados. Con la información resultante se procedió a dividir la cantidad que un pluviómetro es el dominante sobre la cantidad de veces que un pluviómetro es analizado, esto dio como resultado un valor único para cada pluviómetro al que se refiere como *rango* de dicho pluviómetro. El rango tiene un valor que se encuentra entre 0 y 1. Si el rango de un pluviómetro es igual a 0 significa que en ningún PCA es considerado como dominante, y si este tiene un valor igual a 1 significa que este ha sido dominante en todos los radios considerados. De este modo se obtuvo una forma de diferenciar a todos los pluviómetros entre si y poder definir cuáles son más cruciales que otros en base a cuantos pluviómetros se desea extraer.

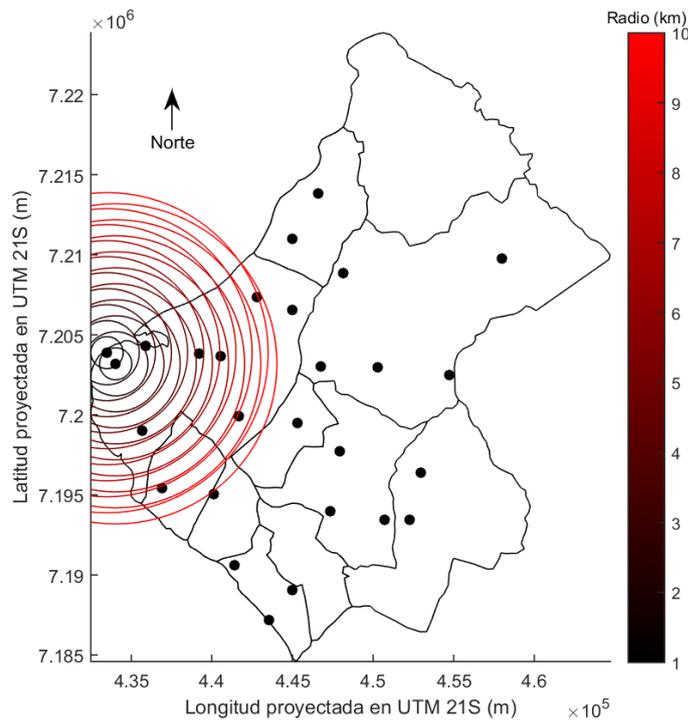


Figura 8: Circunferencias en donde fueron realizados PCA con radios para el par de pluviómetros más cercanos entre sí (7 y 26).

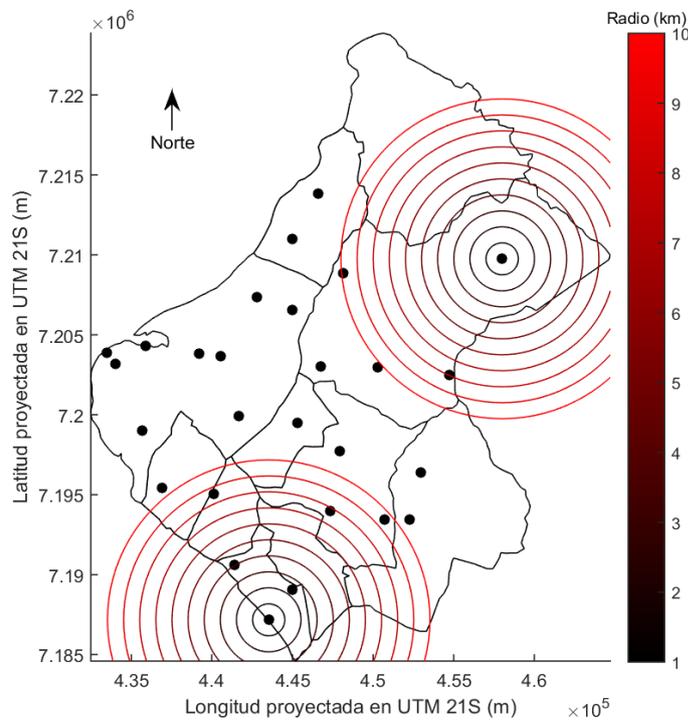


Figura 9: Circunferencias en donde fueron realizados PCA con radios para el par de pluviómetros más lejanos entre sí (9 y 25).

2.5. Evaluación de resultados

Para ambos métodos se realizó una comparación del sistema intacto con casos en los que se retuvieron una cantidad menor de pluviómetros. Los días en los cuales se realizaron estas comparaciones son el 15% del conjunto de mediciones original no analizado en los ninguno de los métodos anteriores. Las características de las mediciones utilizadas son la media de lluvia registrada en el día, la desviación estándar y el coeficiente de variación que es la desviación estándar dividida por la media. Para cada día y caso se calculó un porcentaje de error en base al coeficiente de variación del caso en el que se retienen todos los pluviómetros. Así se tuvo un error de cero para las mediciones completos y un nivel de error para cada caso de extracción de pluviómetros que puede indicar lo que se está perdiendo en información del sistema.

Se realizaron comparaciones del sistema original con las mediciones aisladas de los pluviómetros a retener resultantes de los métodos de PCA. Estas comparaciones utilizaron los patrones de lluvia generados con interpolación IDW y mapas de calor como referencia. También se compararon los resultados de PCA en su potencial para predecir mediciones de lluvia en otros pluviómetros utilizando regresión lineal múltiple. Se realizaron regresiones para cada pluviómetro a desechar utilizando como predictores las mediciones de los pluviómetros seleccionados en los días utilizados en el PCA y como respuestas las mediciones de cada pluviómetro a desechar. Los días a ser predecidos son los días excluidos de los PCA y se calculó el promedio de error absoluto en este conjunto de predicciones. Luego se compararon este promedio con un número grande de combinaciones aleatorias de pluviómetros.

Capítulo 3

Resultados y análisis

3.1. Preparación de datos

La base de datos resultante cuenta con 42 días luego del recorte de días que no cumplían con la nueva condición de día de lluvia. La media de lluvia por pluviómetro paso de ser 4 *mm* a 25 *mm*. La nueva distribución de datos puede observarse en la Fig. 10, en la cual ya no existe una aglomeración entre los 1 a 20 *mm*. Se realizó también la eliminación de los outliers según el test de Grubbs para identificar outliers con un nivel significativo del 0.01. Con este test fueron identificados 4 outliers que fueron tomados como errores y en consecuencia una falta de medición.

3.1.1. Anderson–Darling test

Para comprobar la normalidad en la distribución de la base de datos antes y después de la selección de los días se realizó la prueba Anderson-Darling de normalidad. Esta prueba asume la hipótesis nula de que los datos a ser probados vienen de una población normalmente distribuida (Anderson & Darling, 1952). Se realizó la prueba con un nivel significativo del 0.05, un valor generalmente aceptado y utilizado como el estandar, para las mediciones los 350 días del primer pluviómetro, el resultado fue un rechazo a la hipótesis nula. Esto significa que la prueba rechaza la hipótesis de que los datos provengan de una población normalmente distribuida. La distribución de estos datos pueden observarse en la Fig. 11, donde la curva no sigue la forma de una distribución normal.

Para la base de datos recortada podemos ver gráficamente en la Fig. 12 como al parecer la curva sí sigue la forma de una distribución normal. Al realizar la prueba de Anderson-

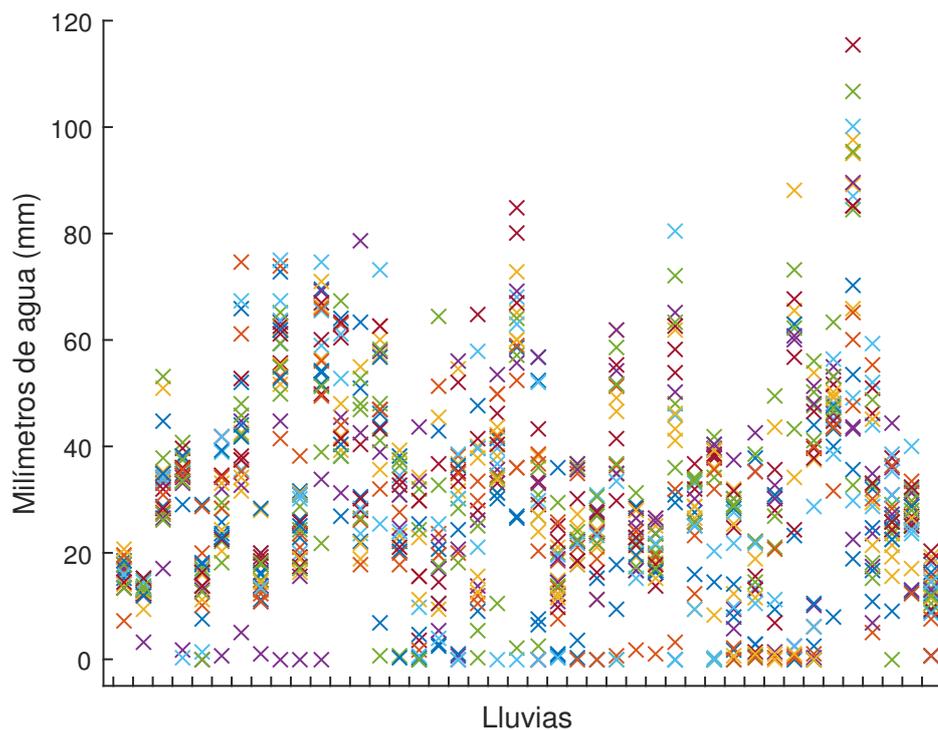


Figura 10: Distribución de valores de milímetros de lluvia registrados por todos los pluviómetros.

Darling, el resultado fue que falla en rechazar la hipótesis nula. Esto significa que la prueba falla en rechazar la hipótesis nula, o bien no rechaza la hipótesis de que los datos provengan de una población normalmente distribuida. Aunque la prueba no demuestra de que se trate de una distribución normal, esta apunta a que este normalmente distribuida después de realizarse el corte de datos. Esto es conveniente para el posterior PCA ya que este asume que las observaciones están normalmente distribuidas para cada variable (Rummel, 1988).

3.2. IDW

Para la interpolación de datos faltantes fue seleccionado como parámetro de potencia $u = 2$ basado en los resultados de IDW para datos de lluvia diarios de Dirks et al. (1998) y los análisis de Shepard (1968). Este parámetro es de hecho el parámetro estándar para una interpolación IDW.

La interpolaciones se realizaron para días en donde los datos faltantes para cada día fueron

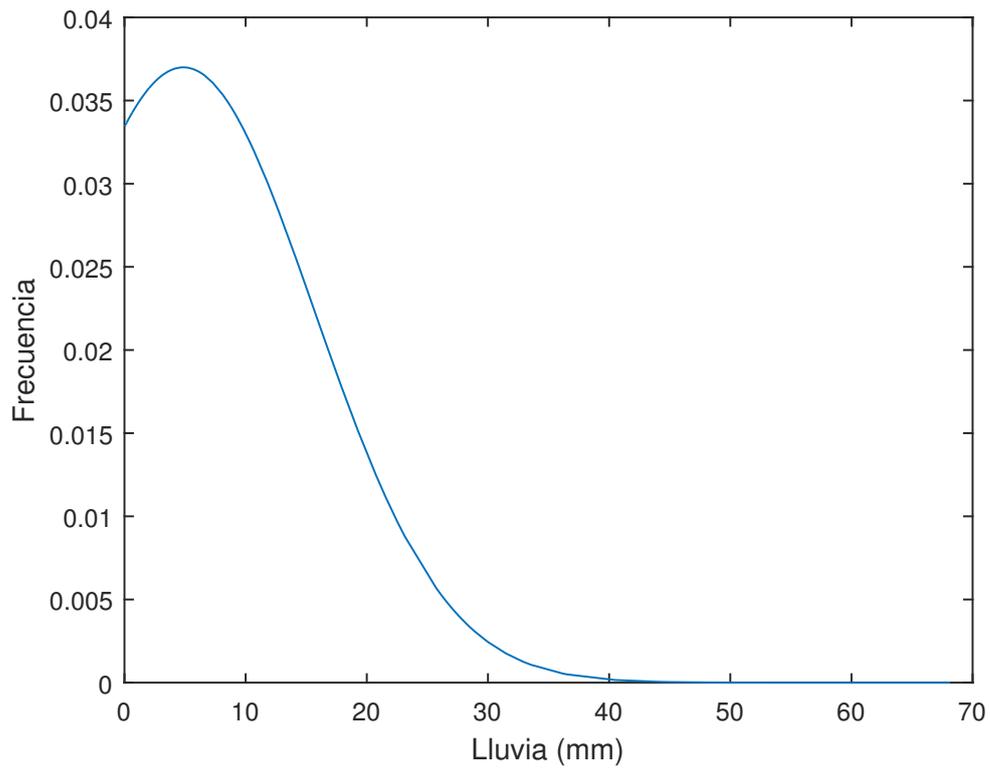


Figura 11: Distribución de las medias diarias de lluvia para los 350 días originales.

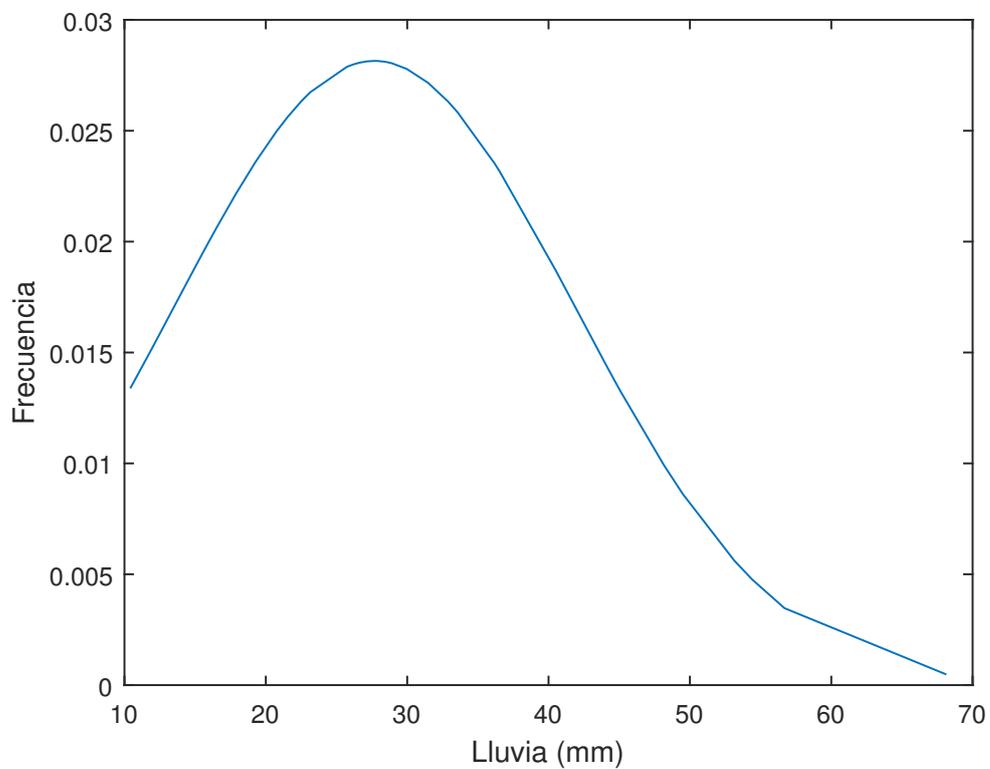


Figura 12: Distribución de las medias diarias de lluvia para los 42 días resultantes del recorte.

interpolados utilizando los datos medidos del mismo día. Con esta interpolación se pudo hacer un mapa de calor de Gran Asunción. El mapa de calor en la Fig. 13 representa las variaciones de mediciones en todos los puntos de la ciudad para un día específico utilizado en los PCA. El mapa es una proyección de las latitudes y longitudes en metros en la zona UTM 21S.

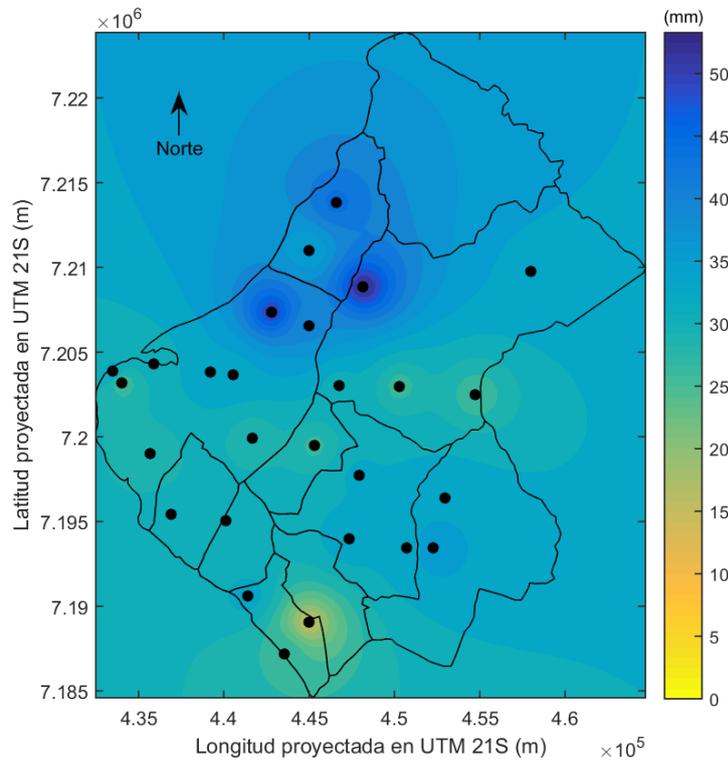


Figura 13: Mapa de calor realizado con resultados de la interpretación IDW utilizando todos los pluviómetros para un día analizado en el PCA.

3.3. PCA

3.3.1. Método de análisis completo

Se realizó el primero método para los 36 días (85% del total) consiguiéndose las matrices de los 27 componentes principales y sus respectivos autovalores. Los autovalores y la proporción de varianza explicada por cada componente se encuentran en la Tabla 1. Se observa como el primer componente sí explica, como se esperaba, la mayor parte de la varianza de las mediciones llevándose un poco menos que 52% del total mientras que a partir del componente número 13 el porcentaje de varianza explicado es menor al 1%.

Tabla 1: Componentes y sus respectivos autovalores con porcentaje de varianza explicado.

Componente	Autovalor		
	Total	Porcentaje de varianza	Cumulativo
1	13.823	51.195	51.195
2	3.066	11.356	62.551
3	2.449	9.071	71.622
4	1.493	5.528	77.150
5	1.340	4.965	82.115
6	0.736	2.726	84.841
7	0.673	2.494	87.334
8	0.559	2.071	89.405
9	0.506	1.874	91.279
10	0.485	1.795	93.074
11	0.379	1.403	94.477
12	0.335	1.240	95.717
13	0.202	0.748	96.465
14	0.183	0.677	97.143
15	0.158	0.584	97.727
16	0.135	0.502	98.229
17	0.113	0.418	98.647
18	0.104	0.384	99.031
19	0.090	0.333	99.364
20	0.052	0.191	99.555
21	0.042	0.157	99.712
22	0.032	0.117	99.829
23	0.018	0.067	99.896
24	0.012	0.045	99.940
25	0.008	0.031	99.972
26	0.005	0.018	99.989
27	0.003	0.011	100.000

Al aplicar la regla de Kaiser se puede concluir que se deberían retener solo 5 componentes ya que desde el componente número 6 los autovalores son inferiores a la unidad. Mientras que si se lleva a cabo el scree-test podríamos retener solo 3 componentes ya que estos forman un codo como puede verse en la Fig. 14. Se decidió por retener 5 componentes ya que la diferencia de porcentaje de varianza retenida entre retener 3 componentes y retener 5 componentes es de casi el 10%. Los 5 componentes retenidos explican en su totalidad 82% de la varianza de los datos originales.

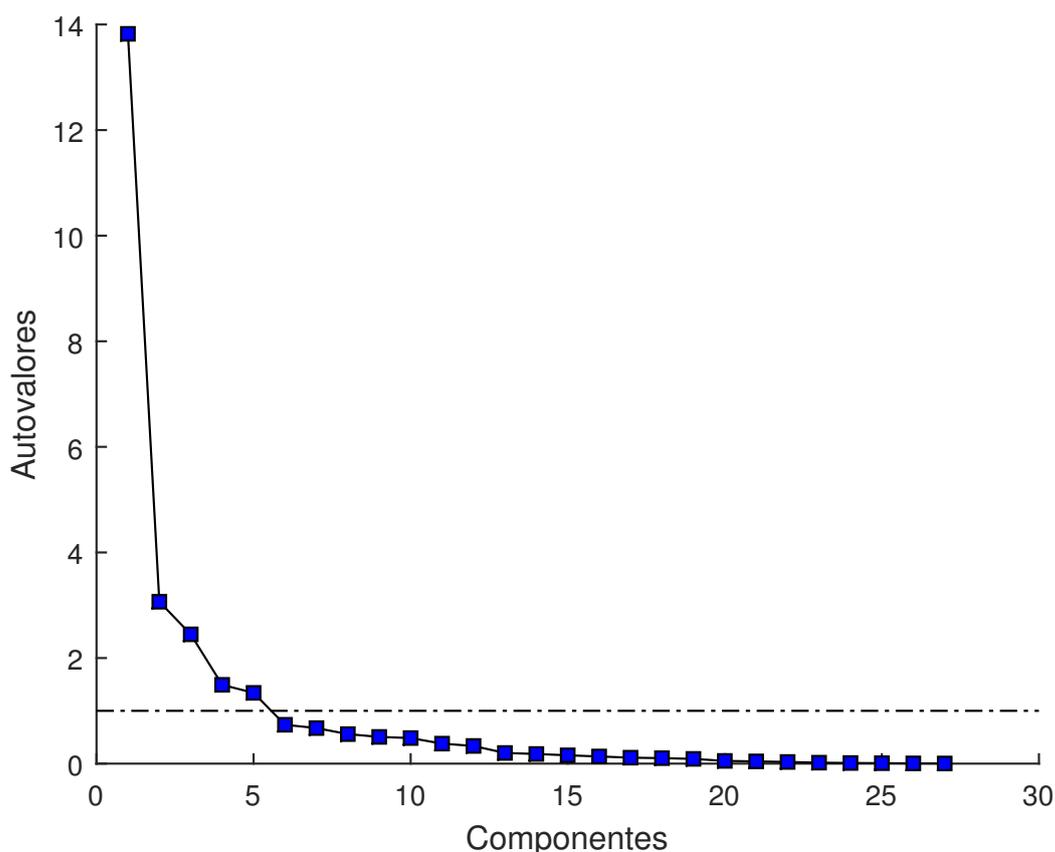
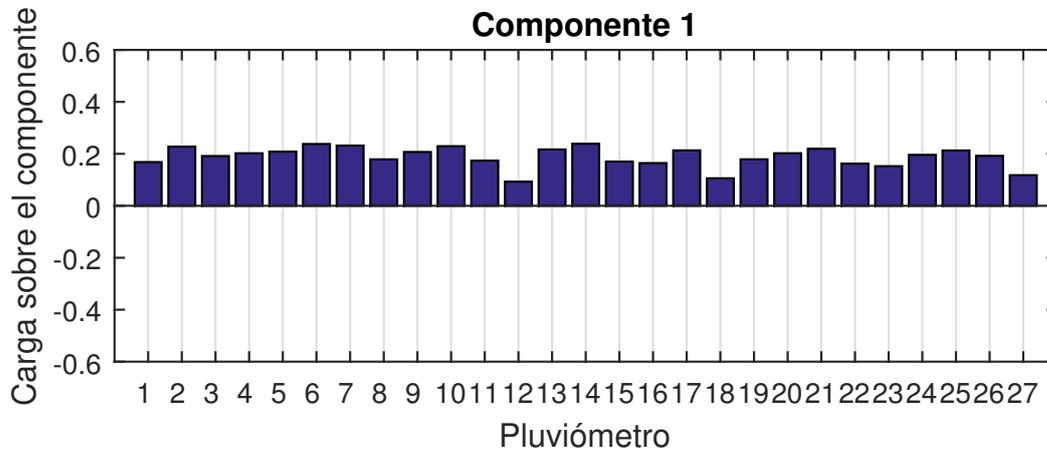


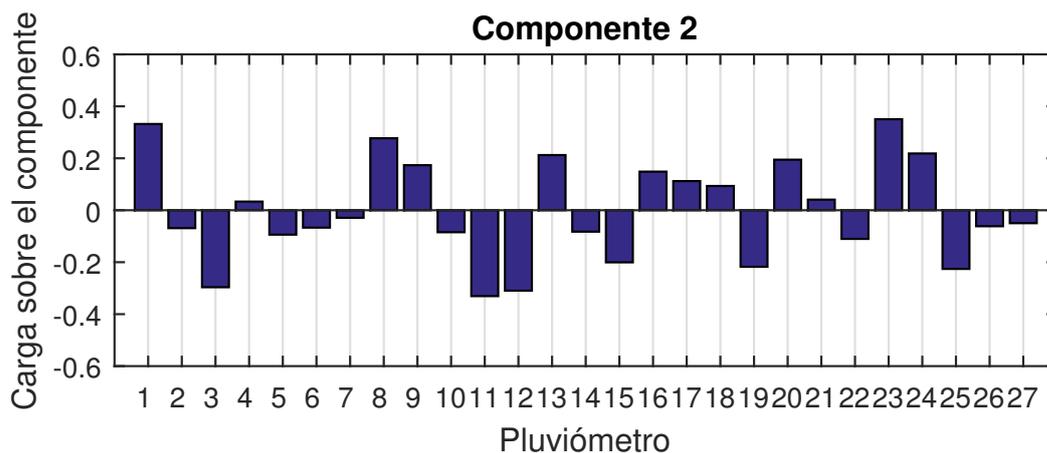
Figura 14: Gráfico del Scree-Test para los 27 componentes principales.

En la Fig. 15 se observan los coeficientes del primer (Véase Fig. 15a) y segundo componente (Véase Fig. 15b) para cada pluviómetro (51 y 11% respectivamente). En el primer componente se puede ver como ningún pluviómetro resulta particularmente dominante, la mayoría se acerca al valor de 0,2 y todos son positivos. Esto muestra como la mayor parte de la varianza describe simplemente el comportamiento general de la lluvia en donde cuando existe una lluvia todos los pluviómetros varían de la misma forma; a mayor lluvia mayores valores en las mediciones. El segundo componente ofrece menos homogeneidad en sus coeficientes, en este caso la interpretación es menos obvia. Se podría decirse que estos explican como pluviómetros más cercanos comparten coeficientes similares mientras que los lejanos no, pero esta afirmación es hecha con mucha menos seguridad. La dificultad de

interpretación empeora al tomar en cuenta más componentes ya que estos explican menores cantidades de variación. En base a esto la decisión de utilizar una rotación Varimax dentro del espacio de los 5 primeros componentes parece justificarse.



(a)



(b)

Figura 15: Cargas de los primeros 2 componentes principales. a) Primer componente. b) Segundo componente.

Los coeficientes resultantes de la rotación pueden observarse en la Tabla 2. Después de la rotación el porcentaje de varianza explicada por cada factor está más equilibradamente distribuida pero el total de varianza explicada sigue siendo del 82%. En los coeficientes de cada factor podemos observar una más clara dominancia por ciertos pluviómetros. Aunque en cada factor existen 2 o 3 pluviómetros con valores absolutos altos se seleccionó a solo un pluviómetro por factor obteniendo como resultado a los pluviómetros 16, 18, 22, 23, y 26 como los pluviómetros más importantes del sistema.

Para evaluar el efecto de mantener solo estos 5 pluviómetros se calcularon las característi-

Tabla 2: Matriz de carga de factores después de rotación Varimax

Pluviometro	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
1	0.013	0.348	-0.112	0.082	-0.029
2	0.213	0.070	-0.051	-0.193	-0.018
3	0.298	-0.133	0.092	-0.137	-0.148
4	0.090	0.081	-0.036	0.139	-0.310
5	0.145	-0.072	-0.180	-0.007	-0.089
6	0.294	0.016	-0.104	-0.081	0.084
7	0.344	0.068	0.034	0.074	-0.030
8	0.085	0.417	0.061	-0.073	-0.056
9	-0.077	0.220	-0.166	-0.024	-0.260
10	0.348	-0.016	-0.084	-0.025	0.122
11	0.245	-0.304	-0.117	-0.012	-0.087
12	-0.083	-0.159	0.091	-0.198	-0.541
13	0.104	0.339	-0.001	-0.059	-0.099
14	0.365	0.026	0.018	0.015	0.003
15	0.075	-0.031	-0.110	-0.483	0.017
16	-0.090	0.031	-0.485	-0.040	0.130
17	-0.019	0.066	-0.360	-0.010	-0.031
18	0.230	-0.051	-0.076	0.543	-0.030
19	-0.006	-0.222	-0.361	-0.236	-0.041
20	0.004	0.201	-0.113	0.180	-0.279
21	0.112	-0.011	-0.314	0.066	0.026
22	-0.061	-0.004	0.043	0.014	-0.587
23	0.009	0.437	-0.059	-0.111	0.076
24	-0.077	0.123	-0.389	0.098	-0.047
25	0.146	-0.195	-0.203	-0.071	-0.131
26	0.406	0.075	0.178	0.077	-0.016
27	0.102	0.190	0.127	-0.449	0.022

cas de las mediciones de solo estos 5 pluviómetros en comparación con las de los datos completos. Los resultados expuestos en la Tabla 3 muestran como eliminando 80% de los pluviómetros se obtiene un promedio de error del 49%. En la tabla se muestran el valor promedio del día \bar{x} , la desviación estándar del día S , el coeficiente de variación CV que es la desviación estándar sobre el promedio. El error porcentual es calculado en base al valor del CV del nuevo conjunto de pluviómetros con respecto al valor del CV de todos los pluviómetros.

Tabla 3: Cálculo de estadísticas de precipitación para 6 días para el PCA completo

Número de Pluviómetros	Estadística	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Promedio
27	\bar{x}	49.50	68.14	31.48	28.34	27.90	11.99	-
	S	20.62	27.59	14.97	13.65	9.82	5.18	-
	CV	0.4165	0.4049	0.4757	0.4816	0.3521	0.4317	0.427
	Error(%)	0	0	0	0	0	0	0
5	\bar{x}	44.76	55.52	25.91	22.94	23.15	7.35	-
	S	9.68	32.61	17.92	13.72	9.67	6.93	-
	CV	0.2163	0.5872	0.6916	0.5982	0.4176	0.9431	0.576
	Error(%)	-48.07	45.02	45.38	24.20	18.61	118.44	49.95

3.3.2. Método de análisis con radios.

Se realizó el segundo método para los 36 días (85% del total) consiguiéndose las puntuaciones para los 27 pluviómetros. Se realizó primero el método con radios que iban de 1 a 27 *km* así consiguiéndose abarcar a todos los pluviómetros con cada centro analizado. Los resultados arrojaron que unos pocos pluviómetros tenían puntuaciones cercanas a 1 mientras que la mayoría quedaban con puntuaciones menores al 0,1. Al analizarse los resultados se descubrió que a partir de utilizar 10 *km* como límite de radio los mismos pluviómetros se repetían como pluviómetros dominantes haciendo que de los rangos de los demás pluviómetros sean cada vez menores haciendo que el ranking no tenga una escala practica para su análisis. Por esta razón se realizaron los PCA comenzando con radios de 1 *km* y aumentando de a 1 *km* hasta llegar a los 10 *km*.

En la Tabla 5 se observan las veces que cada pluviómetro fue analizado y las veces que fue seleccionado como pluviómetro dominante. Estos datos son importantes para poder calificar de manera más justa a pluviómetros que se encuentran en lugares de más densidad de la red y a pluviómetros que se encuentran en lugares más apartados. Con estas dos mediciones se calculó el rango de cada pluviómetro que definirá su importancia en explicar la variación del sistema. Estos resultados están expuestos en la Tabla 5 en donde se

muestran, para cada grupo de pluviómetros, el valor promedio del día \bar{x} , la desviación estándar del día S , el coeficiente de variación CV que es la desviación estándar sobre el promedio. El error porcentual es calculado en base al valor del CV del nuevo conjunto de pluviómetros con respecto al valor del CV de todos los pluviómetros.

El pluviómetro 20 es el de mayor puntaje con un puntaje de 1 lo que significa que fue seleccionado como pluviómetro dominante 100% de las veces que fue analizado con otros pluviómetros, mientras que el pluviómetro 7 tiene un puntaje de 0.

Tabla 4: Rangos de pluviómetros

Número de pluviómetro	Número de veces dentro del radio	Número de veces como pluviómetro principal	Rango
20	41	41	1.000
26	48	44	0.917
12	41	21	0.512
25	14	7	0.500
24	62	28	0.452
3	60	22	0.367
1	58	17	0.293
15	28	8	0.286
6	71	18	0.254
11	31	7	0.226
9	30	5	0.167
23	51	7	0.137
5	51	6	0.118
22	43	5	0.116
8	52	5	0.096
13	42	3	0.071
16	42	3	0.071
19	47	3	0.064
21	64	4	0.063
27	53	3	0.057
18	39	2	0.051
2	66	3	0.045
14	68	3	0.044
4	47	2	0.043
17	60	2	0.033
10	67	1	0.015
7	44	0	0.000

Tabla 5: Cálculo de estadísticas de precipitación para 6 días para el PCA con Radios

Rango mayor a:	Número de Pluviómetros	Estadística	Día 1	Día 2	Día 3	Día 4	Día 5	Día 6	Promedio
-	27	\bar{x}	49.50	68.14	31.48	28.34	27.90	11.99	-
		S	20.62	27.59	14.97	13.65	9.82	5.18	-
		CV	0.417	0.405	0.476	0.482	0.352	0.432	0.427
		Error(%)	0	0	0	0	0	0	0
0	26	\bar{x}	49.73	66.32	30.91	27.99	27.95	12.08	-
		S	20.99	26.43	14.97	13.79	10.02	5.25	-
		CV	0.422	0.399	0.484	0.493	0.358	0.435	0.432
		Error(%)	1.36	-1.59	1.81	2.32	1.80	0.73	1.60
0.05	21	\bar{x}	50.43	63.45	30.39	28.60	28.58	11.25	-
		S	23.39	27.00	16.32	15.08	10.86	5.39	-
		CV	0.464	0.426	0.537	0.527	0.380	0.479	0.469
		Error(%)	11.34	5.09	12.88	9.46	7.93	10.95	9.61
0.1	14	\bar{x}	54.76	64.00	28.27	29.85	29.45	12.43	-
		S	25.80	24.62	14.18	17.71	11.30	4.06	-
		CV	0.471	0.385	0.501	0.593	0.384	0.326	0.443
		Error(%)	13.10	-5.00	5.41	23.14	9.00	-24.42	13.34
0.2	10	\bar{x}	49.11	66.83	26.69	25.64	31.32	12.80	-
		S	4.08	26.45	13.06	12.15	12.17	2.33	-
		CV	0.083	0.396	0.489	0.474	0.389	0.182	0.336
		Error(%)	-80.04	-2.26	2.86	-1.58	10.39	-57.76	25.81
0.3	6	\bar{x}	51.46	69.02	26.13	25.37	33.15	12.64	-
		S	3.37	32.02	14.31	16.06	15.80	1.19	-
		CV	0.066	0.464	0.548	0.633	0.477	0.094	0.380
		Error(%)	-84.26	14.58	15.11	31.44	35.41	-78.23	43.17
0.4	5	\bar{x}	51.85	69.64	24.73	25.49	34.68	12.22	-
		S	3.62	35.76	15.53	17.95	17.16	0.67	-
		CV	0.070	0.514	0.628	0.704	0.495	0.055	0.411
		Error(%)	-83.23	26.81	32.02	46.26	40.54	-87.25	52.69
0.5	3	\bar{x}	52.07	56.39	31.80	22.48	27.21	12.06	-
		S	4.89	43.59	15.59	20.03	3.11	0.52	-
		CV	0.094	0.773	0.490	0.891	0.114	0.043	0.401
		Error(%)	-77.45	90.86	3.03	84.98	-67.57	-89.96	68.97
0.9	2	\bar{x}	54.75	68.31	38.70	19.21	25.89	12.23	-
		S	2.19	54.29	14.14	27.17	2.98	0.61	-
		CV	0.040	0.795	0.365	1.414	0.115	0.050	0.463
		Error(%)	-90.38	96.28	-23.23	193.63	-67.31	-88.46	93.22

3.4. Evaluación de métodos y comparación con selección aleatoria

Para la evaluación de los métodos se compararon los resultados utilizando los datos de las mediciones de lluvia excluidos del análisis de componentes principales. Se compararon los datos completos con los datos de los pluviómetros retenidos utilizando métodos de interpolación (IDW) y regresión (regresión lineal múltiple).

3.4.1. Mapas de calor

En el caso del método de PCA completo se utilizaron mapas de calor creados utilizando IDW para poder crear una matriz de datos de mediciones para toda Gran Asunción. Estos mapas de calor muestran los patrones de lluvia recolectados por los pluviómetros para un día. En la Fig. 16 se muestran estos patrones formados con la cantidad máxima de información disponible, la de los 27 pluviómetros originales.

Para el caso de retener los 5 pluviómetros existen dos maneras de utilizarlos para predecir los patrones de lluvia originales, incluso sin contar con la información de los 22 pluviómetros restantes. La primera, expuesta en la Fig. 17, se basa simplemente en utilizar solo 5 pluviómetros e ignorar por completo los otros puntos de medición. De esta forma se logra de cierta forma recrear estos patrones de forma exagerada ya que los picos de valores se encuentran en estos 5 puntos de medición. La segunda manera se basa en utilizar estos 5 pluviómetros e interpolar los valores de mediciones para los demás 22 pluviómetros en el mismo día. Se utilizan estos datos interpolados como mediciones reales y se utilizan estos 27 puntos para el mapa de calor de la Fig. 18. Este parece presentar de mejor manera los patrones de los datos originales, teniendo más picos que el primer método.

De manera a una mejor diferenciación de estas dos maneras de interpolar los resultados se calcularon las diferencias entre los datos originales y los dos métodos. Las diferencias o errores del primer y segundo fueron expresados utilizando su valor absoluto en dos mapas de calor. La manera de utilizar simplemente 5 pluviómetros (Fig. 19) presenta errores grandes en donde se ubican pluviómetros que registraron picos en ese día, sin embargo parece ser bueno al predecir los valores al tener valores bajos (menores a 10 *mm*) en un gran área de la ciudad. En el segundo método de interpolación (Fig. 20) se presentan

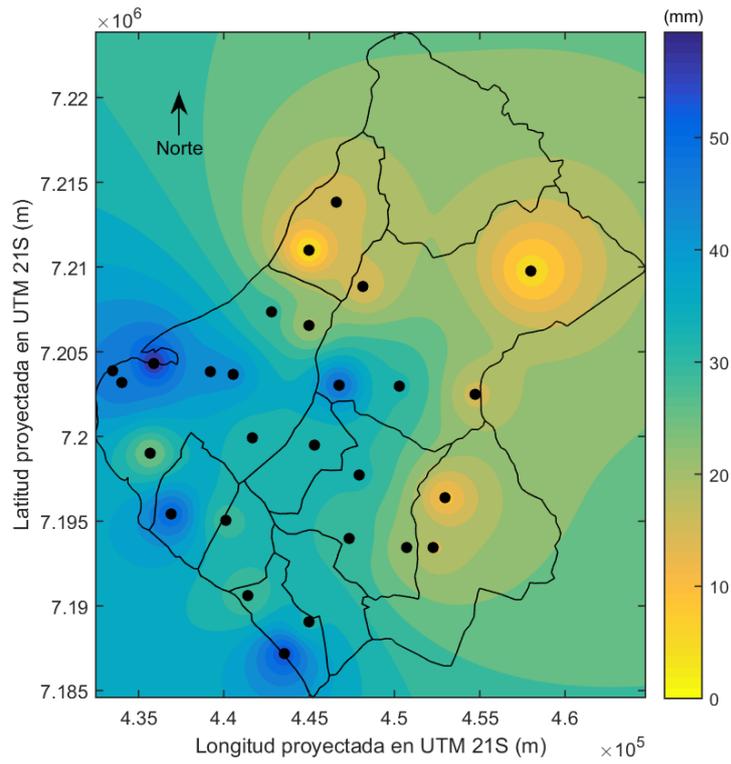


Figura 16: Mapa de calor creado utilizando las mediciones de 27 pluviómetros de un solo día excluido del análisis en PCA.

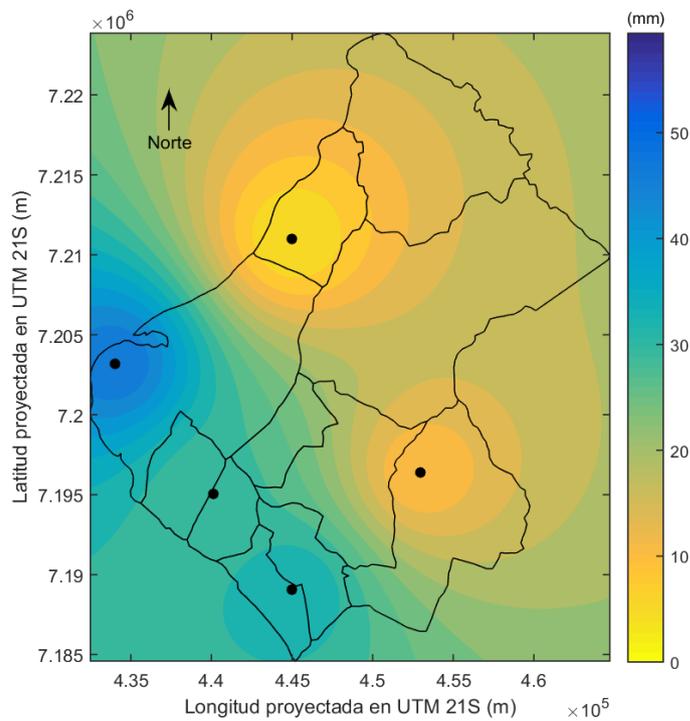


Figura 17: Mapa de calor creado utilizando las mediciones de 5 pluviómetros en un solo día excluido del análisis de PCA.

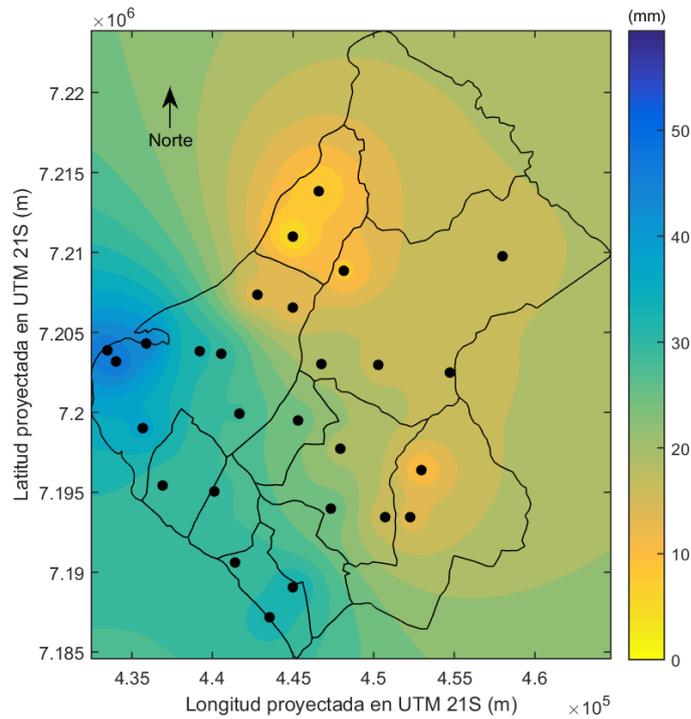


Figura 18: Mapa de calor creado utilizando las mediciones reales de 5 pluviómetros más las mediciones interpoladas en los otros 22 pluviómetros de un solo día excluido del análisis de PCA.

características similares aunque este parece por segunda vez presentar mejores resultados ya que las manchas de errores mayores a 10 mm se encuentran más contenidas hacia los pluviómetros en donde se registraron picos originalmente. Lo que apunta a que el segundo método genera menores errores a la hora de predecir valores en el mapa.

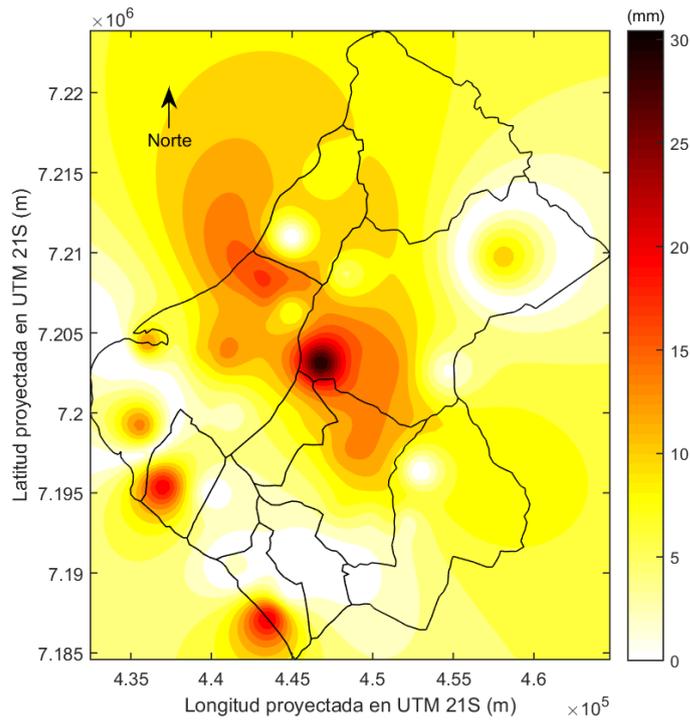


Figura 19: Mapa de calor creado utilizando las diferencias absolutas entre las interpolaciones con datos originales de la Fig. 16 y las interpolaciones de la Fig. 17.

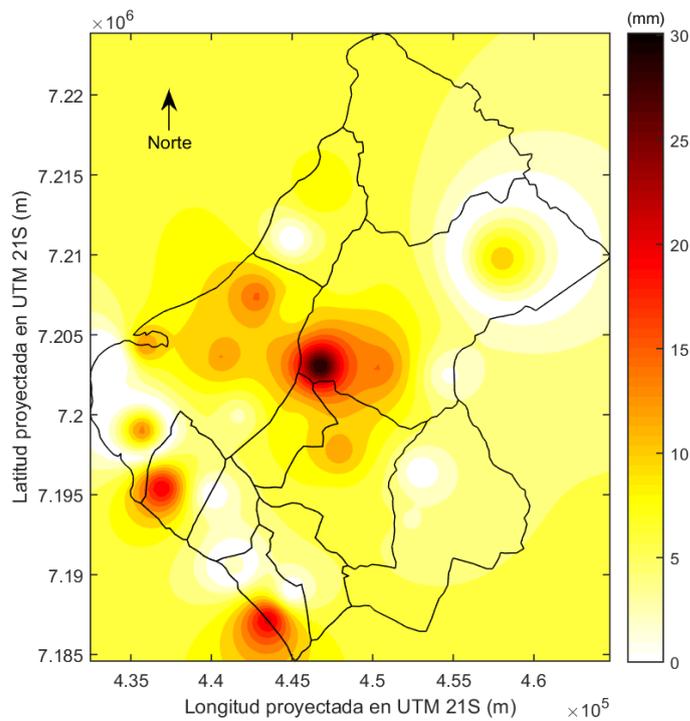


Figura 20: Mapa de calor creado utilizando las diferencias absolutas entre las interpolaciones con datos originales de la Fig. 16 y las interpolaciones de la Fig. 18.

3.4.2. Selección aleatoria de pluviómetros

Para la validación las selecciones de pluviómetros se realizaron comparaciones de estas con combinaciones hechas de manera aleatoria. La base de las comparaciones fue su potencial de predecir las mediciones en los demás pluviómetros no seleccionados. En la Fig. 21 se observa la distribución de promedios de errores absolutos para 1000 combinaciones aleatorias de pluviómetros. Aunque la selección dada por el PCA completo no es la que presenta menor error, esta posee un valor de 12 mm que esta solo a 2 mm del menor error promedio de las combinaciones. Sin embargo, el error de la selección de 5 pluviómetros del método de PCA con radios presenta un promedio de error mucho más alto. Esto refleja los datos estadísticos previos en donde este método presentaba mayor error al seleccionar la misma cantidad de pluviómetros.

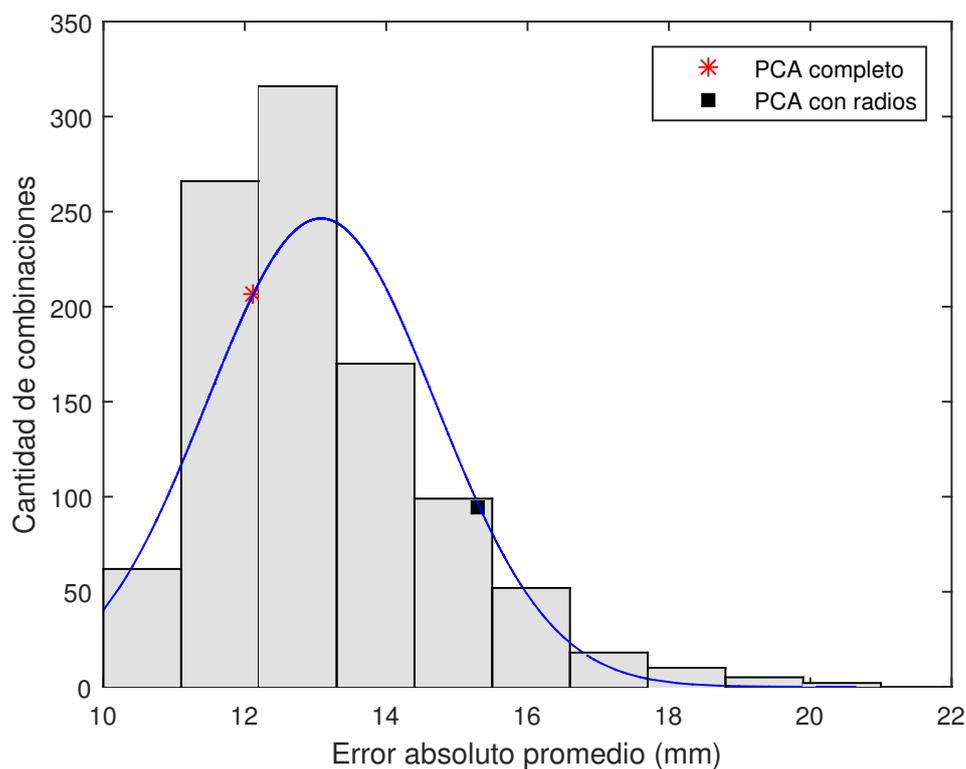


Figura 21: Distribución de medias de errores al regresar 1000 combinaciones aleatorias de 5 pluviómetros.

Se comparó también el caso del método de PCA con radios para el caso de retener más pluviómetros, en este caso 20. Se realizó la comparación utilizando 1000 combinaciones aleatorias y midiendo sus errores análogamente al método de PCA completo. Los resultados pueden verse en la Fig. 22. En el caso de retener más pluviómetros el método de PCA con radios parece comportarse de la misma manera que el método de PCA completo para

5 pluviómetros seleccionados.

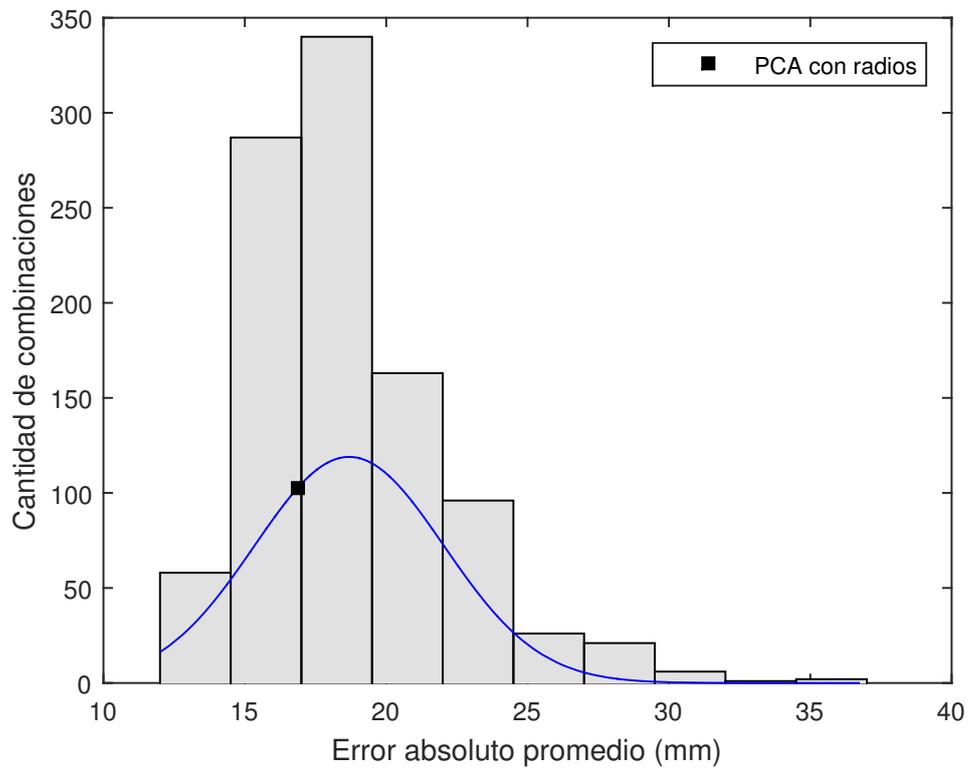


Figura 22: Distribución de medias de errores al regresar 1000 combinaciones aleatorias de 20 pluviómetros.

Capítulo 4

Conclusión

El presente estudio provee herramientas, basadas en pruebas analíticas, para la ayuda en la toma de decisiones y evaluación del mantenimiento de una red de pluviómetros que posea la cantidad de datos suficientes para su análisis. El conjunto de análisis realizado en este estudio fue llevado a cabo sobre información recolectada por una red de pluviómetros que fue creada basándose en factores que no necesariamente optimizan la eficiencia del mismo. El estudio buscó evidenciar la posibilidad de aumentar la eficiencia del mismo eliminando recursos, creando una red más simple sin perder mucha información en el proceso. Para esto se optó por la adaptación del PCA, utilizado en la compresión de sistemas de datos multivariantes.

En el procesamiento de los datos de mediciones fue posible solucionar la falta de datos con la implementación del método de interpolación IDW. Para esto se utiliza las mediciones tomadas en un mismo día y las coordenadas de los puntos de medición. Esto dio como resultado un campo de predicciones de precipitación para cada punto del área de Gran Asunción. Los resultados ayudaron a obtener una mejor interpretación de los patrones de la lluvia para cada día.

Basándose en los principios del PCA y al hecho de que mediciones cercanas de lluvia presentan una alta relación entre sí, se desarrollaron dos métodos de aplicación de PCA. El primer método utilizó la información de variación presente en todos los pluviómetros para poder reducir una red de 27 pluviómetros a una selección de solo 5 pluviómetros. El método logró comprimir la variabilidad del sistema en 5 componentes principales que en su conjunto explicaron 82 % de la varianza del sistema original. Estos 5 medidores seleccionados fueron los más influyentes dentro del espacio de estos 5 componentes principales.

En el segundo método utilizo PCA en muchos subconjuntos de pluviómetros delimitados por radios diferentes. Se dio prioridad solo al primer componente que normalmente lograba

explicar 50 a 90 % de la varianza del subconjunto analizado. De esta forma se logró crear un sistema de puntaje que se basa en la cantidad de veces que un pluviómetro es el más influyente dentro del primer componente y la cantidad de veces que es analizado. Esto permitió valorar a cada pluviómetro y obtener un método de compararlos unos con otros individualmente.

Para ambos métodos se calcularon los errores de aspecto estadístico causados por la eliminación de pluviómetros. Para el primer método se logró obtener un promedio de 49 % de error promedio en el coeficiente de variabilidad teniendo una reducción del 81 % del tamaño de la red. Aunque el segundo método presentó un error promedio mayor para la misma cantidad de pluviómetros (al rededor de 4 *mm* de promedio de error absoluto), este permitió medir su error para distintas cantidades de pluviómetros a retener.

Se evaluaron estos resultados utilizando variaciones del método de interpolación IDW. Se analizaron los patrones de lluvia en las interpolaciones generadas por los pluviómetros seleccionados y se compararon los resultados consiguiéndose resultados satisfactorios. Se compararon también las selecciones de los dos métodos con combinaciones aleatorias de pluviómetros en base a su potencial de predecir mediciones reales. Por el método de interpolación utilizado y la forma de medir el error se puede esperar que existan combinaciones aleatorias con menor error que la selección del PCA, esto no quiere decir que estas representen mejor la variabilidad de las lluvias en todos los periodos de tiempo. Estas combinaciones aleatorias de menor error pueden cambiar en distintos periodos de tiempo, mientras que se espera que la selección de PCA se posicione siempre en un rango de las combinaciones de menor error. Los dos métodos presentaron niveles de error que se encuentran en el rango bajo de errores posibles. Se llegó a la conclusión de que el primer método se desempeña mejor para su cantidad de pluviómetros seleccionados mientras que el segundo método se desempeña mejor cuando la cantidad de pluviómetros a retener es mayor.

Dada la baja cantidad de tiempo de recolección de datos en el cual se basó el estudio, se recomienda una aplicación de la presente metodología a periodos de tiempo más extensos antes de tomar alguna decisión que altere dramáticamente la red de pluviómetros. Sin embargo los resultados obtenidos apuntan a la viabilidad de aplicación de ambos métodos para la reducción de pluviómetros. Estas herramientas proveen información importante para la toma de decisiones flexible para diferentes objetivos de reducción de variables consiguiendo una reducción de recursos sin comprometer en gran medida la información del sistema.

Apéndices

A. Distancias entre pluviómetros

Tabla 6: Distancias en metros entre pluviómetros

P1s	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	0	5843	3772	3435	9461	11812	17018	12688	7812	12766	11258	4956	10508	8216	19833	17175	14130	5477	14873	6822	9075	6108	7323	12799	19026	16179	15430
2	5843	0	3129	8096	6086	6347	12920	9634	12426	7900	9895	9203	9320	3675	14394	11536	8291	10437	9791	9696	3846	8261	6855	7109	16326	11882	10610
3	3772	3129	0	5115	5774	9461	15679	12285	11420	10628	8326	6108	11220	6620	16144	13602	10919	9154	11124	9635	5453	5229	8265	9320	15703	14017	13722
4	3435	8096	5115	0	9536	14418	20149	16039	9563	15906	9867	1531	13943	11142	20772	18466	16017	7238	15998	9738	10889	3667	10758	14317	17841	19332	18405
5	9461	6086	5774	9536	0	9781	16821	15123	17192	11100	4487	9745	15356	9140	11443	9618	8673	14910	6234	15233	3551	7140	12928	6392	10262	16200	14471
6	11812	6347	9461	14418	9781	0	7042	6687	16735	1323	14264	15551	8985	3876	11839	8604	4375	15269	9236	13081	6237	14412	8619	5349	18508	6516	4700
7	17018	12620	15679	20149	16821	7042	0	5334	19486	5730	21301	21483	9143	9081	16437	13519	9952	18780	15486	15479	13273	20878	11039	11821	25202	878	2429
8	12688	9634	12285	16039	15123	6687	5334	0	14200	5974	19362	17488	3820	6047	18370	15160	10073	13647	15871	10197	11755	17480	9399	11987	24753	4484	5303
9	7812	12426	11420	9563	17192	16735	19486	14200	0	17213	18971	10754	10548	12894	20810	23872	20212	2378	22156	4007	16193	13178	8574	19455	26818	18013	18773
10	12766	7900	10628	15506	11100	1323	5730	5974	17213	0	15585	16685	8741	4611	12395	9190	5055	15890	10242	13434	7558	15651	8853	6393	19686	5234	3376
11	11258	9895	8326	9867	4487	14264	21301	19362	18971	15585	0	9367	19166	13319	13967	12966	12904	16595	9161	17847	8028	6361	16433	10573	7971	20729	18957
12	4956	9203	6108	1531	9745	15551	21483	17488	10754	16685	9367	0	15463	12429	21147	19015	16842	8494	15940	11207	11084	3008	12278	15011	17291	20678	19071
13	10508	9320	11220	13943	15356	8985	9143	3820	10548	8741	19166	15463	0	6556	20784	17551	13334	10280	17510	6591	12432	16069	3185	13779	25489	8301	8976
14	8216	3675	6620	11142	9140	3876	9081	6047	12894	4611	13319	12429	6556	0	14721	11552	7526	11393	11020	9348	5946	11851	5151	7428	19052	8304	7265
15	19833	14394	16144	20772	11443	11839	16437	18370	26810	12395	13967	21147	20784	14721	0	3235	7464	24821	5212	23787	10765	18576	19870	7408	12110	16452	14309
16	17775	11536	13602	18466	9618	8604	13519	15160	23872	9190	12966	19015	17551	11552	3235	0	4229	21963	3845	20720	8148	16677	16702	4428	13078	13456	11291
17	14130	8291	10919	10017	8673	4375	9952	10973	20212	5035	12904	16842	13334	7526	7464	4229	0	18465	5549	16841	5845	14982	12637	2333	15375	9728	7573
18	5477	10437	9154	7238	14910	15269	18780	13647	2378	15890	16595	8494	10280	11393	24821	21963	18465	0	20052	3894	14113	10824	7746	17534	24453	17901	17796
19	14873	9791	11124	15598	6234	9236	15486	15871	22156	10242	9161	15940	17510	11020	5212	3845	5549	20052	0	19453	5967	13372	15993	3889	9880	15222	13087
20	6822	9696	9635	9738	15233	13081	15479	10197	4007	13434	17847	11207	6591	9348	23787	20720	16841	3894	19453	0	13540	12906	4639	16379	25333	14607	14810
21	9075	3846	5453	10389	3551	6237	13273	11755	16193	7558	8028	11084	12432	5946	10765	8148	5845	14113	5967	13540	0	9137	10410	3933	13107	12709	10933
22	6108	8261	5239	3667	7140	14412	20878	17480	13178	15651	6361	3008	16069	11851	18576	16677	14982	10824	13372	12906	9137	0	12946	12952	14301	20132	18852
23	7323	6855	8265	10758	12928	8619	11039	5939	8574	8853	16433	12278	3185	5151	19870	16702	12637	7746	15993	4639	10410	12946	0	12542	23180	10160	10200
24	12799	7109	9320	14317	6392	5349	11821	11987	19455	6393	10573	15011	13779	7428	7408	4428	2333	17534	3889	16379	3933	12952	12542	0	13381	11484	9306
25	19026	16326	15703	17841	10262	18508	25202	24753	26818	19086	7974	17291	25489	19052	12110	13078	15375	24453	9880	25333	13107	14301	23180	13381	0	24854	22777
26	16179	11882	14917	19332	16260	6516	878	4484	18613	5234	20729	20678	8301	8304	16452	13456	9728	17901	15222	14607	12709	20132	10160	11484	24854	0	2171
27	15430	10610	13722	18405	14471	4700	2429	5303	18773	3376	18957	19071	8076	7265	14309	11291	7573	17796	13087	14810	10833	18852	10200	9396	22777	2171	0

Referencias bibliográficas

- ABC Color (2018). Cables sueltos y calles inundadas tras las lluvias. *ABC Color*, N.A. Disponible en <http://www.abc.com.py/nacionales/lluvia-colapsa-calles-en-asuncion-y-peligra-electrocucion-1737285.html>.
- Anderson, T. W. & Darling, D. A. (1952). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The annals of mathematical statistics*, 193–212.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245–276.
- Christensen, J. H. & Christensen, O. B. (2003). Climate modelling: severe summertime flooding in europe. *Nature*, 421(6925), 805.
- Dirks, K., Hay, J., Stow, C., & Harris, D. (1998). High-resolution studies of rainfall on norfolk island: Part ii: Interpolation of rainfall data. *Journal of Hydrology*, 208(3-4), 187–193.
- Gangopadhyay, S., Das Gupta, A., & Nachabe, M. (2001). Evaluation of ground water monitoring network by principal component analysis. *Groundwater*, 39(2), 181–191.
- Grubbs, F. E. et al. (1950). Sample criteria for testing outlying observations. *The Annals of Mathematical Statistics*, 21(1), 27–58.
- Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094–1096). Springer.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1), 141–151.
- Kablouti, M., Ouerdachi, L., & Boutaghane, H. (2012). Spatial interpolation of annual precipitation in annaba-algeria-comparison and evaluation of methods. *Energy Procedia*, 18, 468–475.
- Li, J. & Heap, A. D. (2014). Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software*, 53, 173–189.
- Manly, B. F. (1986). Randomization and regression methods for testing for associations with geographical, environmental and biological distances between populations. *Researches on Population Ecology*, 28(2), 201–218.

- Ouyang, Y. (2005). Evaluation of river water quality monitoring stations by principal component analysis. *Water research*, 39(12), 2621–2635.
- Rummel, R. J. (1988). *Applied factor analysis*. Northwestern University Press.
- Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, (pp. 517–524). ACM.
- Smith, L. I. (2002). A tutorial on principal components analysis. Technical report.
- Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, 47(1-2), 123–138.
- Zhang, Q., Sun, P., Singh, V. P., & Chen, X. (2012). Spatial-temporal precipitation changes (1956–2000) and their implications for agriculture in china. *Global and Planetary Change*, 82, 86–95.