

**UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA CIVIL**



TESIS

**REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA
PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ**

PARA OBTENER EL TÍTULO PROFESIONAL DE INGENIERO CIVIL

ELABORADO POR

DANTE RAMIRO LUCANO ALARCO

ID: 0000-0003-1171-1716

ASESOR

Ing. CARLOS MANUEL IPARRAGUIRRE ORTIZ

ID: 0000-0002-5343-7375

LIMA - PERÚ

2024

© 2024, Universidad Nacional de Ingeniería. Todos los derechos reservados.
“El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”

Lucano Alarco, Dante Ramiro

dlucanoa@uni.pe

956151109

DEDICATORIA

A mi familia, en especial a mi madre Diana, por siempre creer en lo que podía llegar a hacer y su gran apoyo en infinidad de ocasiones y formas.

AGRADECIMIENTOS

Agradezco a todas las personas que se involucraron de diferentes formas en el desarrollo de la presente investigación, desde amigos, conocidos y familiares; siempre tendré presente cada aporte que pude recibir de su parte.

Un gran agradecimiento a mi madre Diana quien siempre creyó en mí, me impulsó a tomar retos y constantemente me hace notar que siempre puedo dar un poco más de mí. Así mismo, a los integrantes de mi familia, por siempre apoyarme en todos los aspectos posibles para llevar de la mejor manera mi etapa universitaria y actualmente profesional, darme una visión más amplia de las cosas y transmitirme sus ánimos para culminar la tesis.

Toda mi gratitud para la consultora Anddes Asociados S.A.C. por confiar en mi potencial para liderar el desarrollo de la presente investigación, en particular a mi asesor Carlos Iparraguirre y los profesionales Miguel Huamán, Pablo Cárdenas y Sheila Aguilar por compartir conmigo sus conocimientos y guiarme en el desarrollo de la parte técnica de la presente tesis, asimismo, agradecer de igual manera a Selene Quispe por su orientación y tutela a lo largo del proceso de la investigación.

A mi alma máter, la Universidad Nacional de Ingeniería, por brindarme una educación de excelencia mediante sus profesores quienes me formaron como profesional y trasladaron hacia mí el deseo de aportar con mis conocimientos y potencial al desarrollo de mi país.

ÍNDICE

Resumen	4
Abstract.....	6
Prólogo.....	8
Lista de tablas.....	10
Lista de figuras	12
Lista de símbolos y siglas	17
Capítulo I: Introducción.....	19
1.1 Introducción	19
1.2 Descripción del problema de investigación	20
1.3 Objetivos del estudio.....	21
1.3.1 Objetivo General	21
1.3.2 Objetivos Específicos.....	21
1.4 Antecedentes	21
1.5 Hipótesis del estudio	24
1.5.1 Hipótesis General	24
1.5.2 Hipótesis Específicas.....	24
1.6 Metodología	24
Capítulo II: Marco teórico y conceptual	26
2.1 Precipitación máxima probable (PMP)	26
2.1.1 Duración de lluvia	26
2.1.2 Periodo de retorno	26
2.1.3 Lámina de lluvia	26
2.1.4 Lluvia de diseño	26
2.2 Método estadístico para el cálculo de la PMP	27
2.3 Metodología de Stowhas (inspirado en el método estadístico para el cálculo de la PMP)	30

2.4	Análisis de consistencia de una serie de datos de precipitación.....	37
2.4.1	Análisis visual	37
2.4.2	Análisis de <i>outliers</i>	38
2.4.3	Análisis de estacionariedad.....	39
2.4.4	Análisis de homogeneidad	41
2.4.5	Análisis de independencia serial	41
2.5	Clasificación regional basada en <i>clustering</i>	42
2.5.1	Índices estadísticos.....	43
2.6	Regresión de datos	58
2.6.1	Regresión con Máquina de Vectores de Soporte	58
Capítulo III: Análisis exploratorio y de consistencia de datos de precipitación.....		60
3.1	Datos de precipitación de estaciones terrenas	60
3.1.1	Datos de precipitación de SENAMHI.....	60
3.1.2	Datos de precipitación de la ANA.....	62
3.2	Datos de precipitación de estaciones terrenas utilizados para la formación del producto grillado PISCO (crudo / completado / <i>raster</i>) ...	64
3.3	Mapas y shapes de SENAMHI	65
3.3.1	Regiones Homogéneas de Precipitación Máxima	65
3.3.2	Mapa de Clasificación Climática del Perú	66
3.4	Formación de series de máximas anuales	66
3.4.1	Selección de series confiables de máximas anuales de 24; 48 y 72 horas.....	66
3.4.2	Base de datos de series máximas anuales	69
3.5	Análisis de consistencia de series de máximas anuales	70
3.5.1	Análisis visual	70
3.5.2	Detección de <i>outliers</i>	72
3.5.3	Análisis de estacionariedad o tendencia	84

3.5.4	Análisis de homogeneidad	84
3.5.5	Análisis de independencia serial	85
3.5.6	Resultados de los análisis de consistencia	85
Capítulo IV: Clasificación regional y determinación de envolventes.....		87
4.1	Clasificación regional de las estaciones pluviométricas por el método de agrupamiento (<i>clustering</i>)	88
4.2	Análisis exploratorios para la determinación de envolventes.....	92
4.3	Determinación de las envolventes regionales de K	101
4.4	Determinación de las envolventes de máximos y PMP	108
Capítulo V: Cálculo de la PMP		114
5.1	Cálculo de la PMP para cada estación.....	114
5.1.1	Verificación y corrección de valores de PMP.....	114
5.2	Regresión espacial y generación de imágenes rasterizadas de la PMP	116
5.3	Consideraciones en la elaboración de mapas debido a las zonas con información escasa	120
Capítulo VI: Análisis y discusión de resultados.....		125
6.1	<i>Clusters</i>	125
6.2	Envolventes	134
6.3	Valores puntuales de la PMP en comparación con precipitaciones máximas observadas en el mundo.....	137
6.4	PMP y P24h de 10 000 años.....	140
Conclusiones		143
Recomendaciones		146
Referencias bibliográficas		148
Anexos		157

Resumen

En la presente tesis, se ha realizado un estudio para la estimación de la Precipitación Máxima Probable (PMP) de 24; 48 y 72 horas en el Perú, se ha profundizado principalmente en la realización de análisis de consistencia con el fin de verificar la calidad de la información pluviométrica recolectada de las diversas fuentes oficiales y depurar información no apropiada para el análisis, la formación de *clusters* de estaciones definidos por criterios hidrológico-estadísticos y de posición geográfica, la elaboración de envolventes del factor K (factor que para el cálculo de la PMP) de las estaciones pluviométricas y finalmente en llevar los valores puntuales de PMP a la mayor parte del territorio nacional posible mediante una regresión espacial.

Se partió de la realización de un análisis exploratorio de datos de precipitación diaria de todas las estaciones pluviométricas instaladas en el país, cuyos datos se encuentran disponibles en repositorios de los sitios web del Senamhi y de la Autoridad Nacional del Agua (ANA), a los datos se les aplicó una serie de análisis de consistencia los cuales fueron el análisis visual, prueba de detección de datos atípicos (*outliers*), y los análisis de tendencias, homogeneidad e independencia serial.

Dada la amplia variedad climática y de mecanismos de producción de precipitaciones máximas, se buscó identificar grupos de estaciones con un comportamiento similar que nos lleve a definir de manera más específica las curvas del factor K que le corresponde a cada uno, ello se logró mediante el *clustering* (agrupamiento) de estaciones basado en estadísticas de las series de máximas anuales (media, máximo, coeficiente de variación) e información de la ubicación espacial de las estaciones (latitud, longitud y altitud), para determinar el número ideal de *clusters*, o regiones, y las estaciones pluviométricas que los conforman.

En busca de encontrar curvas envolventes del factor K que sean apropiadas para el Perú en 24; 48 y 72 horas, se siguió principalmente la metodología aplicada por Stowhas (1983) en su estudio de la PMP en Chile, la cual tiene una base estadística sólida y permite estimar el valor más extremo que podría llegar a tener el factor K mediante una doble maximización.

A continuación, se calculó la PMP mediante la fórmula establecida originalmente por Hershfield (1961) haciendo uso del factor K estimado a partir de las curvas envolventes definidas para el Perú, permitiendo ello tener una base de datos de valores puntuales en la mayor parte del país.

Buscando extender los valores de la PMP estimados a los lugares con mayor escasez de información, se realizó una regresión espacial al coeficiente de variación de los máximos anuales, los parámetros de ubicación (longitud, latitud y altitud), obteniendo así, *rasters* de dichas variables que sirvieron como insumo para realizar finalmente la regresión espacial de la PMP.

Asimismo, se busca que la información generada contribuya como una herramienta de apoyo y consulta referencial para el diseño de obras hidráulicas de protección y control; para conseguir diseños más seguros, y por tanto se minimizando la probabilidad de falla o excedencia de alguna de estas obras, a su vez, ello reducirá la cantidad de inundaciones, pérdida de numerosas vidas humanas, daño otras obras civiles y al medio ambiente.

Palabras clave: Precipitación máxima probable (PMP), factor K , regresión espacial, *support-vectors machine*

Abstract

In this thesis, a study has been carried out for the estimation of the Probable Maximum Precipitation (PMP) of 24, 48 and 72 hours in Peru, it has been mainly focused on the performance of consistency analysis in order to verify the quality of the rainfall information collected from various official sources and to refine information not appropriate for the analysis, the formation of station clusters defined by hydrological-statistical and geographic position criteria, the elaboration of envelopes of the K factor (factor used to calculate the PMP) of the rain gauge stations and finally in bringing the specific PMP values to as much of the national territory as possible through a spatial regression.

The starting point was to carry out an exploratory analysis of daily precipitation data from all the rain gauge stations installed in the country, whose data are available in repositories on the websites of Senamhi and the National Water Authority (ANA), a series of consistency analyzes were applied to the data, which were visual analysis, atypical data detection test (outliers), and trends, homogeneity and serial independence analyses.

Given the wide variety of climate and maximum rainfall production mechanisms, we sought to identify groups of stations with similar behavior that would lead us to define more specifically the K factor curves that correspond to each one. This was achieved by the clustering (grouping) of stations based on statistics of the series of annual maximums (mean, maximum, coefficient of variation) and information on the geographic position of the stations (latitude, longitude and altitude), to determine the ideal number of clusters, or regions, and the rain gauge stations that make them up.

In search of finding K factor envelope curves that are appropriate for Peru in 24; 48 and 72 hours, the methodology applied by Stowhas (1983) in his study of the PMP in Chile was mainly followed, which has a solid statistical basis and allows estimating the most extreme value that the K factor could have through a double maximization.

Next, the PMP was calculated using the formula originally established by Hershfield (1961) using the K factor estimated from the envelope curves defined for Peru, allowing us to have a database of specific values in most of the country.

Seeking to extend the estimated PMP values to the places with the greatest scarcity of information, a spatial regression was carried out on the coefficient of variation of the annual maximums, the location parameters (longitude, latitude and altitude), thus obtaining rasters of said variables that served as input to finally carry out the spatial regression of the PMP.

Likewise, it is sought that the information generated contributes as a support and reference consultation tool for the design of hydraulic protection and control works; to achieve safer designs, and therefore minimizing the probability of failure or exceedance of any of these works, in turn, this will reduce the amount of flooding, loss of numerous human lives, damage to other civil works and to the environment.

Keywords: Probable maximum precipitation (PMP), K factor, spatial regression, support-vectors machine

Prólogo

La PMP según su definición es “la mayor lámina de precipitación para una duración determinada de tormenta, que es físicamente posible en un área de tormenta de tamaño dado, en una ubicación geográfica particular y en una época determinada del año” (Hansen et al., 1982; WMO, 1986, 2009). Este concepto es el parámetro de diseño más empleado para el diseño de presas y estructuras hidráulicas, en las que su falla implica un gran riesgo en pérdidas de vidas, y en gran daño económico y ambiental. En guías de diseño de presas de relaves (CDA, 2019; ANCOLD, 2012) se prescribe la PMP para casos de riesgo extremo; pero en el estándar global GISTM (ICMM et al., 2020) se refiere al evento de 10 000 años de periodo de retorno.

Existen seis métodos reconocidos para evaluar la PMP, cinco de los cuales requieren el desarrollo de modelos meteorológicos, y un método estadístico. Este es el más accesible para el nivel de información con que se suele contar más a menudo: las series de precipitaciones máximas. El método requiere un análisis de envolventes para despejar la estructura de las relaciones entre estadísticas de precipitaciones máximas, que se sustenta en que esas relaciones están vinculadas a los mecanismos propios de producción de lluvias intensas.

El método estadístico se refiere en esencia al trabajo de Hershfield (1961) que consiste en despejar la PMP en un punto con una ecuación similar a la propuesta por Ven Te Chow (1951), donde se expresa como la suma de la media de la muestra más la desviación estándar de la muestra multiplicada por un factor de frecuencia máximo K_{max} determinado a partir de una envolvente con representatividad regional, cuyo trazado requiere buena cobertura espacial de pluviómetros con largas series de datos de precipitación.

Hershfield estableció primero el valor de K_{max} como un límite superior igual a 15 (Hershfield, 1961); luego propuso a K_{max} como variable entre 5 y 20 en función de la media de las precipitaciones máximas anuales (μ) y la duración (Hershfield, 1965). Interesantemente Hershfield señaló que $K_{max} = 15$ es muy grande para áreas muy lluviosas y muy bajo para áreas áridas, y que es muy alto para duraciones menores a 24 horas.

Muchos trabajos han sido desarrollados para evaluar el factor K_{max} en diversas regiones del planeta, entre ellos varios en Hispanoamérica (Bazzano, 2019; Casas

et al., 2008, 2010; Lagos Zuñiga, 2012; Stowhas, 1983), dirigidos generalmente a la duración de 24 horas. En la mayoría se trató de delinear envolventes particulares, lineales por tramos, o no lineales en el plano K_{max} versus μ . Especialmente intuitivo fue el análisis de Stowhas (1983) para Chile, donde indagó la influencia del coeficiente de variación (Cv) sobre el factor K , teniendo en cuenta que Cv se presenta más grande en zonas áridas que en zonas lluviosas, lo que aprovechó para realizar una doble maximización de K_{max} por μ y Cv , controlando el cálculo con la relación de los récords de precipitaciones máximas en Chile y en el mundo con respecto a μ .

Las evaluaciones de K_{max} para duraciones mayores a 24 horas han sido pocas para 48 y 72 horas (Koutsoyiannis & Xanthopoulos, 1999), aun cuando los criterios de seguridad para el diseño son cada vez más exigentes, pasando a indicar el uso de la duración crítica, que puede llegar a alcanzar 48, 72 horas, y hasta 7 días, dependiendo de la climatología y de la confianza en las acciones de respuesta.

En consecuencia, es muy importante la evaluación propia de la PMP para el Perú, pues el factor K_{max} debe ser determinado de forma exclusiva para cada zona de evaluación.

Se aprovechó la mayor información hidrológica ya presente en el país, y los métodos de inspección, conceptos y significados de las relaciones entre K_{max} , μ , Cv , μ_{max} y otros, proporcionados por las investigaciones previas, para el delineado de las envolventes del factor K_{max} . Como el Perú presenta gran diversidad climática se incluyeron técnicas de agrupamiento estadístico para determinar las diversas envolventes para diversas zonas del país. En coherencia con las necesidades para el diseño se evaluaron PMP para duraciones de 24; 48 y 72 horas, y se puso atención a las precipitaciones máximas de 10 000 años de periodo de retorno para la determinación o corroboración de las envolventes. Se recurrió a técnicas de interpolación espacial para elaborar mapas de PMP de utilidad práctica.

Ing. Carlos Manuel Iparraguirre Ortiz

Asesor

Lista de tablas

Tabla 1	<i>Valores I_{cr} para la prueba de Mann - Kendall</i>	41
Tabla 2	<i>Estaciones sospechosas de tener datos que fueron errores de registro o fuera de escala física creíble</i>	81
Tabla 3	<i>Cantidad de estaciones que cumplen con los análisis de consistencia</i>	85
Tabla 4	<i>Índices estadísticos empleados para estimar el mejor número de clusters</i>	88
Tabla 5	<i>Estaciones que conforman los Clusters 7 y 8</i>	90
Tabla 6	<i>Identificación del valor extremo de K para 24 horas</i>	99
Tabla 7	<i>Identificación del valor extremo de K para 48 horas</i>	100
Tabla 8	<i>Identificación del valor extremo de K para 72 horas</i>	100
Tabla 9	<i>$K_{maxprobable}$ determinado para 24; 48 y 72 horas</i>	100
Tabla 10	<i>Errores identificados en los valores de PMP</i>	115
Tabla 11	<i>Parámetros para la regresión del C_v</i>	117
Tabla 12	<i>Métricas del proceso de calibración del C_v</i>	118
Tabla 13	<i>Métricas del proceso de validación del C_v</i>	118
Tabla 14	<i>Parámetros para la regresión de μ</i>	118
Tabla 15	<i>Métricas del proceso de calibración de μ</i>	118
Tabla 16	<i>Métricas del proceso de validación de μ</i>	119
Tabla 17	<i>Parámetros para las regresiones de la PMP</i>	119
Tabla 18	<i>Métricas del proceso de calibración de la PMP</i>	119
Tabla 19	<i>Métricas del proceso de validación de la PMP</i>	120
Tabla 20	<i>Lluvias puntuales de mayor intensidad observadas en el mundo</i>	137
Tabla 21	<i>Valores máximos de PMP por cluster para el caso de 24 horas</i>	139

Tabla 22	<i>Valores máximos de PMP por cluster para el caso de 48 horas.....</i>	140
Tabla 23	<i>Valores máximos de PMP por cluster para el caso de 72 horas.....</i>	140
Tabla 24	<i>Estaciones con PMP de 24 horas más altas del Perú, comparación con precipitación de TR=10 000 años según SENAMHI</i>	141

Lista de figuras

Figura 1	<i>Metodología</i>	25
Figura 2	<i>Factor de ajuste de μ de la serie de máximas anuales</i>	28
Figura 3	<i>Factor de ajuste de σ de la serie de máximas anuales</i>	28
Figura 4	<i>Factores de ajuste de μ y σ de la serie de máximas anuales basado en la cantidad de años disponibles</i>	29
Figura 5	<i>Parámetro K en función del C_v de precipitación en Chile</i>	31
Figura 6	<i>C_v en función de μ de precipitación en Chile</i>	32
Figura 7	<i>P_m en función de μ de precipitación en Chile</i>	33
Figura 8	<i>Amplificación de la envolvente de la P_m mediante el C_v para aplicarla al parámetro K en Chile</i>	34
Figura 9	<i>Amplificación de la envolvente de P_m mediante C_v y K muestral de estaciones en Chile</i>	35
Figura 10	<i>Precipitaciones máximas registradas en Chile y a nivel internacional</i>	36
Figura 11	<i>Margen entre clases: el criterio que los SVM intentan optimizar</i>	59
Figura 12	<i>Estructura del archivo de la estación Caravelí</i>	61
Figura 13	<i>Estructura del data frame con los datos diarios de las estaciones del SENAMHI</i>	62
Figura 14	<i>Estructura del archivo de la estación Alao</i>	63
Figura 15	<i>Estructura del data frame con datos diarios de las estaciones de la ANA</i>	64
Figura 16	<i>Precipitación diaria de la estación Tinajones del Observatorio del Agua a Nivel Nacional de la ANA</i>	71
Figura 17	<i>Máximos anuales diarios de la estación Santa Rita de Castilla del Observatorio del Agua a Nivel Nacional de la ANA</i>	72
Figura 18	<i>C_v versus μ de máximos anuales para 24 horas, antes de análisis de outliers</i>	75

Figura 19	<i>C_v versus μ de máximos anuales para 24 horas, luego de análisis de outliers</i>	76
Figura 20	<i>y_{max cyn} versus μ de máximos anuales para 24 horas, antes de análisis de outliers</i>	76
Figura 21	<i>y_{max cyn} versus μ de máximos anuales para 24 horas, luego de análisis de outliers</i>	77
Figura 22	<i>y_{max cyn} en orden creciente de μ de máximos anuales para 24 horas, antes de análisis de outliers</i>	77
Figura 23	<i>y_{max cyn} en orden creciente de μ de máximos anuales para 24 horas, luego de análisis de outliers</i>	78
Figura 24	<i>Boxplots de y_{max cyn} en orden creciente de μ de máximos anuales para 24 horas, antes de análisis de outliers</i>	78
Figura 25	<i>Boxplots de y_{max cyn} en orden creciente de μ de máximos anuales para 24 horas, luego de análisis de outliers</i>	79
Figura 26	<i>Máximo versus μ de máximos anuales para 24 horas, antes de análisis de outliers</i>	79
Figura 27	<i>Máximo versus μ de máximos anuales para 24 horas, luego de análisis de outliers</i>	80
Figura 28	<i>K_{muestral} versus μ de máximos anuales para 24 horas, antes de análisis de outliers</i>	80
Figura 29	<i>K_{muestral} versus μ de máximos anuales para 24 horas, luego de análisis de outliers</i>	81
Figura 30	<i>(a) Precipitación media anual (1998–2012) estimada con TRMM P^a 2A25. (b) Estaciones meteorológicas y precipitación media anual. (c) Precipitación media anual por estación (1975–2005).</i>	83
Figura 31	<i>Número óptimo de clusters según el Método del Codo</i>	89
Figura 32	<i>Ubicaciones de las estaciones de los Clusters 7 y 8</i>	91
Figura 33	<i>Clusters para el caso de 24 horas</i>	92

Figura 34	<i>K</i> muestral versus μ de máximos anuales y envolventes de otros estudios para 24 horas.....	93
Figura 35	<i>K</i> muestral versus <i>Cv</i> de los clusters 4; 5; 7 y 8 con su envolvente, para 24 horas.....	95
Figura 36	<i>K</i> muestral versus <i>Cv</i> de los clusters 1 y 3 con su envolvente, para 24 horas.....	95
Figura 37	<i>K</i> muestral versus <i>Cv</i> del cluster 2 con su envolvente, para 24 horas.....	96
Figura 38	<i>K</i> muestral versus <i>Cv</i> del cluster 6 con su envolvente, para 24 horas.....	96
Figura 39	<i>Cv</i> versus μ de los clusters 4; 5; 7 y 8 con su envolvente, para 24 horas.....	97
Figura 40	<i>Cv</i> versus μ de los clusters 1 y 3 con su envolvente, para 24 horas.....	98
Figura 41	<i>Cv</i> versus μ del cluster 2 con su envolvente, para 24 horas.....	98
Figura 42	<i>Cv</i> versus μ del cluster 6 con su envolvente, para 24 horas.....	99
Figura 43	<i>Pm</i> versus μ de los clusters 4; 5; 7 y 8 con su envolvente, para 24 horas.....	101
Figura 44	<i>Pm</i> versus μ de los clusters 1 y 3 con su envolvente, para 24 horas.....	102
Figura 45	<i>Pm</i> versus μ del cluster 2 con su envolvente, para 24 horas.....	102
Figura 46	<i>Pm</i> versus μ del cluster 6 con su envolvente, para 24 horas.....	103
Figura 47	Amplificación de envolvente de los clusters 4; 5; 7 y 8, para 24 horas.....	104
Figura 48	Amplificación de envolvente de los clusters 1 y 3, para 24 horas.....	104
Figura 49	Amplificación de envolvente del cluster 2, para 24 horas.....	105
Figura 50	Amplificación de envolvente del cluster 6, para 24 horas.....	105
Figura 51	Envolvente de <i>K</i> versus μ de los clusters 4; 5; 7 y 8, para 24 horas.....	106

Figura 52	<i>Envolvente de K versus μ de los clusters 1 y 3, para 24 horas.....</i>	107
Figura 53	<i>Envolvente de K versus μ del cluster 2, para 24 horas</i>	107
Figura 54	<i>Envolvente de K versus μ del cluster 6, para 24 horas</i>	108
Figura 55	<i>Pmax versus μ y envolvente generada de los clusters 4; 5; 7 y 8, para 24 horas.....</i>	109
Figura 56	<i>Pmax versus μ y envolvente generada de los clusters 1 y 3, para 24 horas.....</i>	109
Figura 57	<i>Pmax versus μ y envolvente generada del cluster 6, para 24 horas.....</i>	110
Figura 58	<i>Pmax versus μ y envolvente generada del cluster 2, para 24 horas.....</i>	110
Figura 59	<i>PMP versus μ y envolventes generadas de los clusters 4; 5; 7 y 8, para 24 horas.....</i>	111
Figura 60	<i>PMP versus μ y envolventes generadas de los clusters 1 y 3, para 24 horas.....</i>	112
Figura 61	<i>PMP versus μ y envolventes generadas del cluster 6, para 24 horas.....</i>	112
Figura 62	<i>PMP versus μ y envolventes generadas del cluster 2, para 24 horas.....</i>	113
Figura 63	<i>Ubicación de todas las estaciones pluviométricas a nivel nacional</i>	121
Figura 64	<i>Ubicación de las estaciones pluviométricas restantes para el caso de 24 horas</i>	122
Figura 65	<i>Rampa de color Turbo</i>	123
Figura 66	<i>Censura adoptada en el mapa del Perú.....</i>	124
Figura 67	<i>Clusters de estaciones para 24 horas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú.....</i>	126
Figura 68	<i>Clusters de estaciones para 48 horas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú.....</i>	127

Figura 69	<i>Clusters de estaciones para 72 horas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú.....</i>	128
Figura 70	<i>Clusters de estaciones para 24 horas en el Mapa de Clasificación Climática del Perú (2020) componente de lluvias.....</i>	129
Figura 71	<i>Clusters de estaciones para 48 horas en el Mapa de Clasificación Climática del Perú (2020) componente de lluvias.....</i>	130
Figura 72	<i>Clusters de estaciones para 72 horas en el Mapa de Clasificación Climática del Perú (2020) componente de lluvias.....</i>	131
Figura 73	<i>Clusters de estaciones para 24 horas en las zonas pluviométricas del Perú.....</i>	132
Figura 74	<i>Clusters de estaciones para 48 horas en las zonas pluviométricas del Perú.....</i>	133
Figura 75	<i>Clusters de estaciones para 72 horas en las zonas pluviométricas del Perú.....</i>	134
Figura 76	<i>Curvas envolventes de los clusters para 24 horas</i>	135
Figura 77	<i>Curvas envolventes de los clusters para 48 horas</i>	136
Figura 78	<i>Curvas envolventes de los clusters para 72 horas</i>	136
Figura 79	<i>Comparación de episodios de precipitaciones puntuales notables en Australia con los más importantes observados en el mundo.....</i>	138

Lista de símbolos y siglas

Símbolos

μ_y	: Media de una serie de “y” datos.
μ_{n-1}	: Media de una serie de “n” datos excluyendo el dato mayor.
μ	: Media de la serie de máximos anuales.
σ_y	: Desviación estándar de una serie de “y” datos.
σ_{n-1}	: Desviación estándar de una serie de “n” datos excluyendo el dato mayor.
R^2	: Coeficiente de determinación.
P_m	: Precipitación máxima adimensional de una serie de datos.
P_{max}	: Precipitación máxima de una serie de datos.
Cv	: Coeficiente de variación de la serie de máximos anuales.
ϕ_y	: Factor de frecuencia de una serie de “y” datos.
K	: Factor K que se le asigna a una estación dependiendo del <i>cluster</i> o grupo de <i>clusters</i> al que pertenece, se calcula a partir de la curva envolvente del factor K que le corresponda a la estación.
$K_{muestral}$: Factor K calculado a partir de la serie de precipitaciones máximas de la estación a la que pertenece.
$K_{maxprobable}$: Valor máximo probable de K, valor estimado con una maximización simple del coeficiente de variación.
K_{max}	: Valor máximo de K, estimado mediante una doble maximización.
y_{maxcyn}	: Valor máximo centralizado y normalizado de una serie.
σ	: Desviación estándar.
σ	: El parámetro controla la anchura de la “campana”. Si se utiliza un valor grande las características conocidas se extenderán más. Por otro lado, si se utiliza un valor pequeño, los valores de los puntos de nuestras características conocidas tendrán un dominio de valores más reducido.

C : El parámetro de costo, también conocido como el costo de la clasificación errónea, determina cuánta flexibilidad se debe permitir con los datos.

Siglas

ANA : Autoridad Nacional del Agua

ANCOLD : Australian National Committee On Large Dams

CCC : Criterio de Agrupamiento Cúbico (*Cubic Clustering Criterion*)

CDA : Canadian Dam Association

GISTM : Global Industry Standard on Tailings Management

ICMM : International Council on Mining and Metals

MAE : Error Absoluto Medio (*Mean Absolute Error*)

NASA : National Aeronautics and Space Administration

NGA : National Geospatial-Intelligence Agency

PISCO : Peruvian Interpolated data of the SENAMHI's Climatological and hydrological Observations

PMP : Precipitación Máxima Probable

PRI : Principles for Responsible Investment

RMSE : Raíz del Error Cuadrático Medio (*Root Mean Square Error*)

SENAMHI : Servicio Nacional de Meteorología e Hidrología del Perú

SNIRH : Sistema Nacional de Información de Recursos Hídricos

SVM : Máquina de vectores de soporte (*Support Vectors Machine*)

TFPW : Pre-blanqueamiento Sin Tendencias (*Trend Free Pre-Whitening*)

UNEP : United Nations Environment Programme

WMO : World Meteorological Organization

Capítulo I: Introducción

1.1 Introducción

El término Precipitación Máxima Probable (en adelante PMP) se origina de lo que solía llamarse Máxima Precipitación Posible (MPP) la cual se definía como un máximo límite superior. Sin embargo, hubo reportes de que la MPP en algunos casos si se llegó a exceder (Benson, 1973) dado esto, se le cambió el nombre a “Precipitación Máxima Probable (PMP)”. Este concepto se desarrolló para el diseño de obras hidráulicas cuya falla tuviese consecuencias económicas y pérdidas de vidas muy elevadas, con el objetivo de minimizar la posibilidad de falla.

A causa de la falla de varias presas en los años 70s se hizo una revisión de las normas de diseño, en particular de la componente pluvial que se usaba en esa época, la cual estaba basada en lo estudiado por Hershfield (1961). Con base a la revisión, gran cantidad de organismos dedicados a seguridad de presas recomiendan explícitamente, el uso de la PMP como lámina de tormenta para diseño de grandes obras hidráulicas, cuya destrucción involucre riesgos importantes (Sugai & Fill, 1990).

En la actualidad países como Reino Unido, Estados Unidos, Australia y Sudáfrica usan la PMP como parámetro de diseño para sus presas y obras relacionadas, asimismo, resaltan la importancia de mantener actualizada la data con la que se calcula este parámetro (Johnson & Smithers, 2019). La PMP tiene periodos de retorno bastante altos, según estudios, de acuerdo con el National Research Council (1994) estos corresponden a valores entre 10^5 años y 10^9 de años y de acuerdo con Nathan & Weinmann (2019) estos corresponden a valores entre 10^4 años y 10^7 años.

Hay varios métodos para evaluar la PMP: maximización, transposición, combinación, inferencial, generalizado y estadístico, de los cuales los cinco primeros son de fundamento hidrometeorológico y el último de fundamento estadístico, más accesible a condiciones de escasez de datos. El método estadístico o de Hershfield (Campos, 1998; Hershfield, 1961, 1965; WMO, 1971, 1986, 2009) se basa en la estructura de la ecuación propuesta por Ven Te Chow (Chow, 1951) para evaluar precipitaciones máximas en función de su frecuencia, donde el valor de la PMP se obtiene de la suma de la media y la desviación

estándar multiplicada por un factor K , donde este no es un factor de frecuencia sino un límite máximo, el cual fue determinado hasta la duración crítica de 24 horas en función de la precipitación máxima de 24 horas promedio anual, con base de datos de EE. UU. Y otras partes del mundo.

1.2 Descripción del problema de investigación

Actualmente los especialistas dedicados a los campos de la hidráulica y la hidrología aplicada, así como especialistas de investigación, no han profundizado en el tema del fundamento del cálculo de la PMP para el Perú.

En nuestro país, el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) ha realizado investigaciones relacionadas a las precipitaciones extremas en todo nuestro territorio nacional. Como resultado han desarrollado una plataforma web titulada “Modulo para la estimación de curvas de Intensidad – Duración – Frecuencia (IDF)”, cuya interfaz proporciona intensidades de precipitación para periodos de retorno de 2 años a 1 000 años, y duraciones de 1 hora a 24–horas, de cualquier parte del país (SENAMHI, 2021).

En la plataforma web en mención no hay información de precipitaciones cuyo periodo de retorno sea más de 1 000 años o de precipitaciones de 48 y 72 horas de duración que podrían ser críticas.

Por otro lado, las otras fuentes de información de precipitaciones en su mayoría nos ofrecen datos crudos sin un control de calidad como puede ser el caso de los datos de precipitaciones diarias provenientes de estaciones meteorológicas convencionales y automáticas de la web del SENAMHI (s.f.) o los datos de precipitación cada 12 y 24 horas provenientes del Observatorio del Agua a Nivel Nacional del SNIRH de la ANA (s.f.), a excepción de los datos pertenecientes al producto PISCO (Peruvian Interpolated data of the SENAMHI’s Climatological and hydrological Observations), el cual nos ofrece datos de precipitación diarios y mensuales, tanto en su versión cruda (la usada para realizar la investigación) como en su versión procesada; además de datos grillados interpolados diarios y mensuales que fueron el resultado de la investigación realizada por Aybar et al. (2020).

La falta de datos de precipitación de altos periodos de retorno y de 48 y 72 horas de duración que hayan sido procesados y cuenten con un adecuado control de calidad es una limitación para que los especialistas puedan generar diseños más

seguros de obras civiles importantes como son las presas u otras grandes obras hidráulicas para protección y control contra avenidas.

Si se tiene en cuenta, además, que el estado del arte de la hidrología a nivel mundial exige utilizar la PMP, o precipitaciones máximas de muy altos periodos de retorno de 10 000 años, o más como principal parámetro de diseño para este tipo de obras, podemos decir que existe en el Perú una brecha de información y falta de investigación de la PMP en el país.

1.3 Objetivos del estudio

1.3.1 Objetivo General

Estimar la Precipitación Máxima Probable (PMP) de 24; 48 y 72 horas de duración mediante la regionalización del parámetro K de Hershfield en el Perú.

1.3.2 Objetivos Específicos

- Definir regiones hidrológicamente homogéneas en base a criterios hidrológico-estadísticos y de posición geográfica.
- Elaborar envolventes de K para 24; 48 y 72 horas para cada región homogénea, y luego aplicar la curva envolvente para estimar la PMP de 24; 48 y 72 horas de duración de cada estación pluviométrica de la base de datos.
- Evaluar el factor K de las estaciones pluviométricas seleccionadas para duraciones de 24; 48 y 72 horas.
- Evaluar la PMP de 24; 48 y 72 horas en ubicaciones sin datos en el Perú mediante una herramienta geoespacial.

1.4 Antecedentes

El manual de la World Meteorological Organization (WMO) para el cálculo de la PMP (WMO, 2009) expone aplicaciones para algunos países como EE. UU., Canadá, China, Australia e India, y para cuencas ubicadas en regiones orográficas y tropicales. Aparte, varios países desarrollaron sus propios manuales para estimar la PMP y la Inundación Máxima Probable (PMF, del inglés Probable Maximum Flood), aunque el manual de la WMO ha sido la guía principal (Salas et al., 2020).

Investigaciones en diversas partes del mundo han definido variantes a la propuesta original de Hershfield para adaptar la evaluación de la PMP a la realidad estadística y la particularidad de áreas homogéneas desde el punto de vista hidrológico. Stöwhas (1983), analizando 190 estaciones pluviométricas con un total de 6504 estaciones-años de registro, concluye que los coeficientes maximalistas K de Hershfield, que van de 5 a 20, tienden a sobreestimar las PMP en Chile, donde predominan lluvias de gran variabilidad, con altos coeficientes de variación, y sugiere utilizar un factor máximo constante $K = 11$.

Casas et al. (2008) en su estudio de la PMP para Catatonia, España, para la estimación de la PMP identificaron una curva del factor K basada en la data de precipitación de 145 estaciones de la región, luego aplicaron el método de interpolación de Cressman dentro de un modelo GIS (Geographic Information Systems) para obtener la PMP de 24 horas a una resolución espacial de 1 km², variando esta entre 200 y 500 milímetros. Se encontró que el 90 % de las PMP halladas les correspondían a periodos de retorno entre 10 000 y 100 000 000 de años y se pudo notar una gran diferencia respecto a la precipitación máxima anual que se había calculado previamente.

Casas et al. (2010) en el estudio de la PMP en Barcelona, España, expusieron la evaluación de la PMP en Barcelona para duraciones entre 5 minutos hasta 30 horas utilizando dos métodos, para luego compararlos; estos métodos fueron el método físico basado en la maximización de la precipitación actual y el método estadístico de Hershfield, donde, se pudo evidenciar coincidencia en los resultados de los métodos, y un comportamiento incremental apreciable en el rango de las duraciones de 2 a 9 horas y uno no muy notorio desde las 12 horas en adelante. Adicionalmente se pudo calcular que el periodo de retorno equivalente para la PMP de 24 horas era de 360 000 años.

Lagos (2012) en su tesis de maestría sobre el impacto del cambio climático en eventos extremos, analizó los efectos de los escenarios climáticos A2 y B1 (2045 – 2065) en la Precipitación Máxima Probable (PMP) determinada estadísticamente en las regiones chilenas de Coquimbo y Maule, e hidrometeorológicamente en la Cuenca Puclaro. Los resultados obtenidos mostraron que para las cuencas Puclaro y Colbúnse se proyectan los incrementos más severos de precipitación de 20% y 30% respectivamente. Se pudo notar que los aumentos mayores al nivel regional se dan en las zonas de mayor elevación. En el caso del embalse Puclaro

se proyectaron los incrementos más severos aún, de 85% y 81% en las crecidas de 1 000 y 10 000 años. Todos estos resultados, que además incluyen un análisis de incertidumbre estadística, proyectan que existe una vulnerabilidad en los embalses Puclaro y Colbún frente a escenarios de cambio climático, siendo esta tendencia generalizable a regiones de similares proyecciones climáticas y régimen hidrológico mixto en Chile.

Campos (2014) en su estudio de precipitaciones máximas en 24 horas utilizó el método estadístico de Hershfield para ciudades o provincias en el estado de Zacatecas, México y procesó los datos de 98 estaciones que tuvieran más de 25 años de información para evaluar la PMP puntual de 24 horas. Apoyándose con análisis regionales de frecuencias en dos zonas geográficas de Zacatecas, estimó la PMP en sitios o localidades de la misma región donde no había datos.

Bazzano (2019) en su tesis de doctorado sobre lluvias máximas para diseño hidrológico propuso una metodología para la determinación de la lluvia de diseño, tomando como sistema experimental a la provincia de Tucumán, Argentina. A fin de calcular la PMP propuso una metodología novedosa para estimarla estadísticamente, basada en series aleatorias sintéticas e independiente de los registros locales, incorporando un análisis de la incidencia del lapso de medición de la lluvia, producto del uso de pluviómetros de alta frecuencia. Una vez establecida la magnitud de la lluvia se definió su distribución temporal a través de un Hietograma Tipo. Luego, ajustó un algoritmo de atenuación, que convierte la lluvia puntual en una media areal. Por último, propuso la regionalización de los parámetros de los modelos de predicción empleando herramientas de la geoestadística.

Ball et al. (2019) en un manual de estimación de escorrentías en Australia expusieron el caso de la realidad de Australia. En el manual se indica que hay algunas situaciones de diseño en las que parece que la duración crítica de interés puede ser tan larga como 168 horas o 7 días, la que se relaciona con los límites meteorológicos asociados con eventos de precipitación únicos para Australia, por lo cual se hace la recomendación de hacer el cálculo de precipitaciones de estas duraciones para las obras que requieran máxima confiabilidad. Si bien es cierto, estas lluvias son sumamente excepcionales, pero en países como Australia sí se pueden llegar a dar en condiciones extremas, sin embargo, esto se puede tomar

como ejemplo para otros países al tener en cuenta el uso de precipitaciones de más de 24 horas de duración para los diseños.

1.5 Hipótesis del estudio

1.5.1 Hipótesis General

Se puede calcular la PMP de 24; 48 y 72 horas de duración para cada estación pluviométrica del Perú, a partir de los datos de precipitación pertenecientes al SENAMHI y la ANA, aplicando la fórmula propuesta por Hershfield (1961), pero utilizando un parámetro K obtenido a partir de curvas envolventes de los valores $K_{muestra}$, determinadas para cada región hidrológica identificada. El resultado se extendería a lugares sin presencia de estaciones por medio de una regresión espacial e hidrológica.

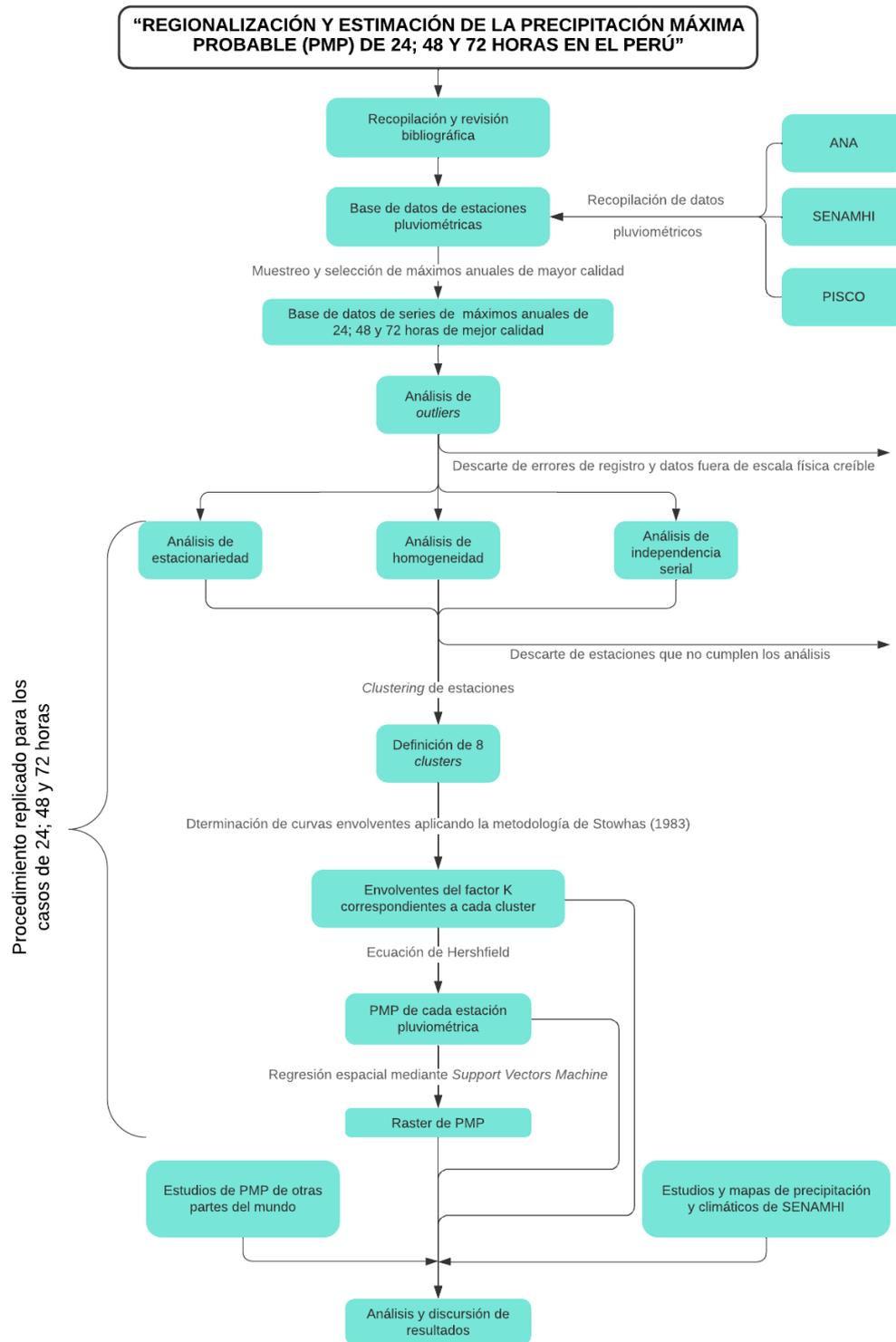
1.5.2 Hipótesis Específicas

- Se postula que los límites de las regiones hidrológicas para el factor K siguen las delimitaciones identificadas en el Mapa de Clasificación Climática Nacional (Senamhi, 2020^a) y el Mapa de la Regionalización de Precipitaciones Máximas en el Perú (SENAMHI, 2017).
- Existen 10 regiones climáticas basadas en las Precipitaciones Máximas Probables (PMP) tal como en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú (SENAMHI, 2017).
- Las precipitaciones máximas de 24 horas de 10 000 años de periodo de retorno que se pueden hallar en la web del SENAMHI titulada “Modulo para la estimación de curvas de Intensidad Duración Frecuencia (IDF)” (SENAMHI, 2021) serán inferiores a los valores de PMP que se puedan calcular para lugares similares.
- No habrá mucha diferencia de magnitud entre la PMP de 48 horas y la PMP de 72 horas en las distintas estaciones, pero la distribución espacial de los valores diferirá, de manera que el modo de representación espacial puede no ser similar para esos casos.

1.6 Metodología

En la Figura 1 se muestra la metodología seguida para el desarrollo de la presente tesis.

Figura 1 Metodología



Fuente: Elaboración propia.

Capítulo II: Marco teórico y conceptual

El fundamento teórico de la presente tesis abarca las teorías y conceptos necesarios para comprender el desarrollo de la presente tesis, abarcando el cálculo de la Precipitación Máxima Probable (PMP) mediante el Método estadístico o de Hershfield y una adaptación del mismo; la clasificación regional basada en clustering (agrupamiento) y una regresión espacial de datos.

2.1 Precipitación máxima probable (PMP)

Es “la mayor profundidad de precipitación para una duración determinada de tormenta, que es físicamente posible en un área de tormenta de tamaño dado, en una ubicación geográfica particular y en una época determinada del año” (Hansen et al., 1982; WMO, 1986, 2009).

2.1.1 Duración de lluvia

Es el lapso en el cual se desarrolla el evento climatológico de precipitación, puede ir desde unos minutos hasta varios días.

2.1.2 Periodo de retorno

Se define como el lapso o número de años que en promedio, se cree que será igualado o excedido, es decir, es la frecuencia con la que se presenta un evento (Méllice & Reason, 2007).

2.1.3 Lámina de lluvia

Se le conoce a la lámina de lluvia como el espesor de una capa de agua que se formaría al recolectar la lluvia en una superficie horizontal. La altura de 1 milímetro de lámina de lluvia equivale a 1 L de agua de lluvia por m².

2.1.4 Lluvia de diseño

Es el evento hipotético pluvial más intenso, estadísticamente previsible, para una duración dada y una recurrencia asignada que, al alimentar un modelo lluvia – escorrentía permite evaluar la Creciente de Proyecto, elemento con el que se diseñan tanto las medidas estructurales (obras) como las no estructurales (planificación territorial).

2.2 Método estadístico para el cálculo de la PMP

El manual de la WMO describe 6 métodos para estimar la PMP, estos son: (a) el método local (modelo de maximización de tormentas locales) (b) el método de transposición (modelo de transposición de tormentas) (c) el método combinado (maximización temporal y espacial de la tormenta) (d) el método inferencial (modelo teórico) (e) el método generalizado, y (f) el método estadístico (WMO, 2009).

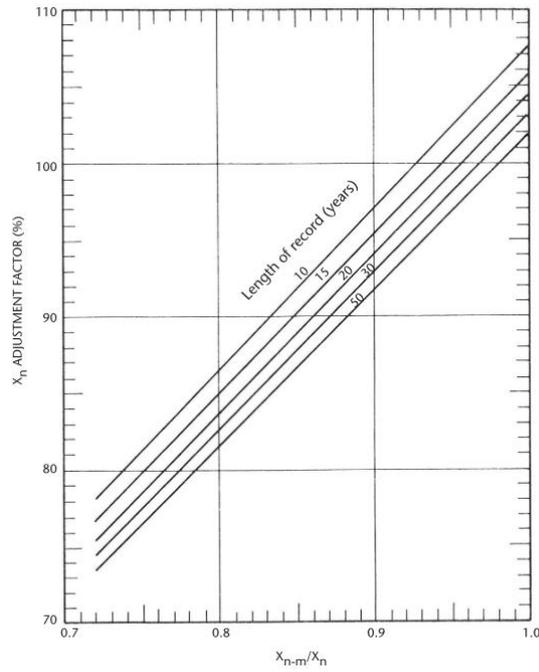
El método (f) corresponde al método propuesto por Hershfield (1961, 1965), el cual es de práctico empleo, puesto que la metodología es bastante sencilla y los datos e información que se requieren para su ejecución son menores en comparación a los demás, corresponden a datos que podemos obtener con relativa facilidad de los registros pluviométricos, además de ello, es uno de los más difundidos y explorados a lo largo de los años.

El Método de Hershfield se basa en minimizar la probabilidad de ocurrencia de la tormenta, maximizando el factor de frecuencia Φ_y en la ecuación general de Chow (1951).

$$y = \mu_y + \Phi_y * \sigma_y \quad (2.1)$$

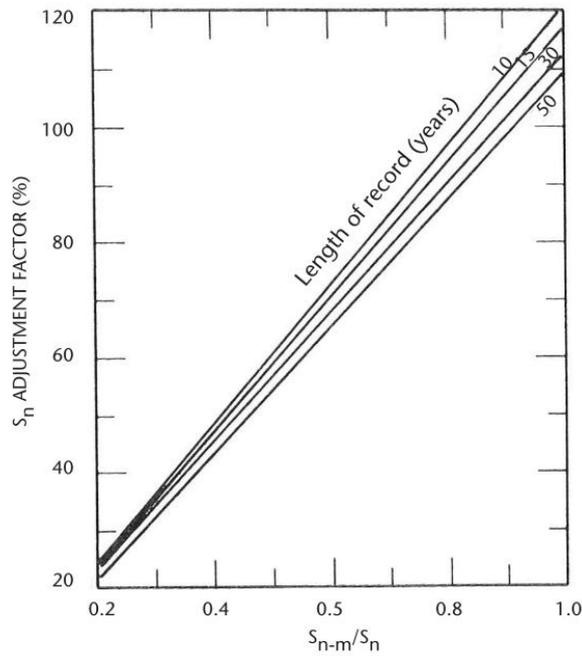
A los valores de la media (μ) y la desviación estándar (σ) se les debe aplicar factores de corrección, el primero es de acuerdo con los cocientes $\frac{\mu_{n-1}}{\mu}$ y $\frac{\sigma_{n-1}}{\sigma}$, dichos factores pueden ser estimados a partir de la Figura 2 y Figura 3 para cada variable respectivamente, adicionalmente se deben aplicar los factores de corrección obtenidos a partir de la Figura 4, basados en la cantidad de años de información disponible.

Figura 2 Factor de ajuste de μ de la serie de máximas anuales



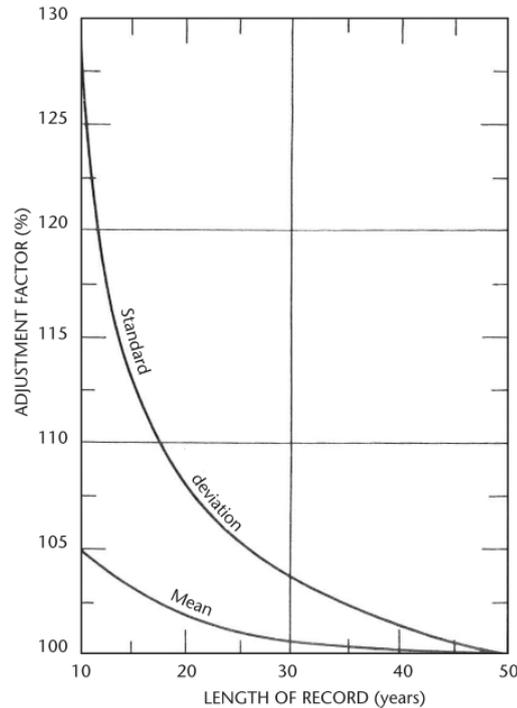
Fuente: Hershfield (1961)

Figura 3 Factor de ajuste de σ de la serie de máximas anuales



Fuente: Hershfield (1961)

Figura 4 Factores de ajuste de μ y σ de la serie de máximas anuales basado en la cantidad de años disponibles



Fuente: Hershfield (1961)

Asumido que la altura de la lámina de lluvia tiende a un límite superior (la PMP), con una probabilidad bajísima de excedencia, habrá un tope en la cantidad de desvíos estándar en que el dato supera a la media, es decir, un valor máximo mediante el cual se podrá calcular la PMP, a este factor de frecuencia se le llama K . Para el cálculo de la PMP se utilizará la siguiente ecuación:

$$PMP = \mu_y + K * \sigma_y \quad (2.2)$$

Donde K se estima a partir de Φ_{n-1} el cual está definido de la siguiente manera:

$$\Phi_{n-1} = K = \frac{y_m - \mu_{n-1}}{\sigma_{n-1}} \quad (2.3)$$

Donde el subíndice $n - 1$ en μ y σ simboliza que al calcular cada uno de ellos, se excluye del conjunto de n datos el de mayor valor, el cual es simbolizado por y_m . Buscando maximizar el valor de K en la ecuación (2.3).

A continuación, se determina la ecuación de una curva envolvente para los valores de K graficados en función de μ , la ecuación de la envolvente buscada tendrá como variable principal la media.

De acuerdo con la bibliografía, se sugiere una exponencial negativa (acotada) para expresar esta envolvente (Caamaño Nelli & García, 1999). La ecuación de esta curva tiene la siguiente forma:

$$K = K_0 * e^{-a*\mu} \quad (2.4)$$

Otra estructura para la envolvente es la siguiente:

$$K = -a * \ln(\mu) + b \quad (2.5)$$

La forma de la ecuación de la envolvente dependerá de la distribución de los puntos graficados correspondientes a cada agrupación de estaciones.

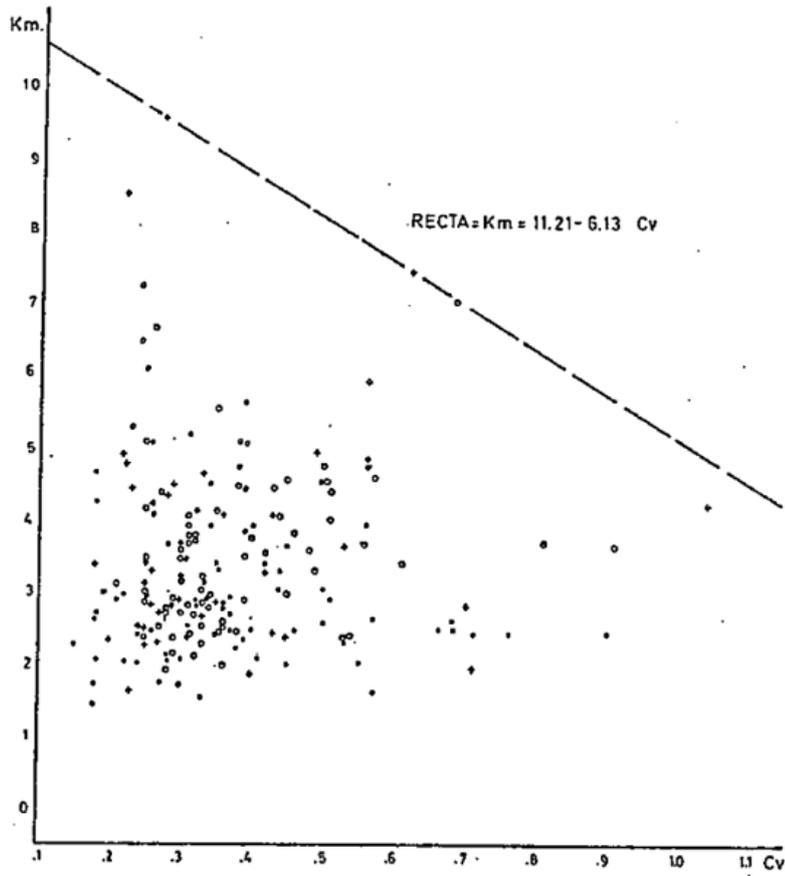
Habiendo sido determinada la ecuación de la envolvente adecuada para cada agrupación de estaciones, se calcula un nuevo valor de K reemplazando el valor de la media de cada estación en la ecuación de la envolvente que corresponda a la agrupación a la cual pertenece dicha estación, obteniendo así un nuevo y definitivo valor de K para cada estación pluviométrica.

Finalmente se procede a calcular el valor de la PMP correspondiente a cada estación pluviométrica de acuerdo con la ecuación (2.2) utilizando el nuevo valor de K .

2.3 Metodología de Stowhas (inspirado en el método estadístico para el cálculo de la PMP)

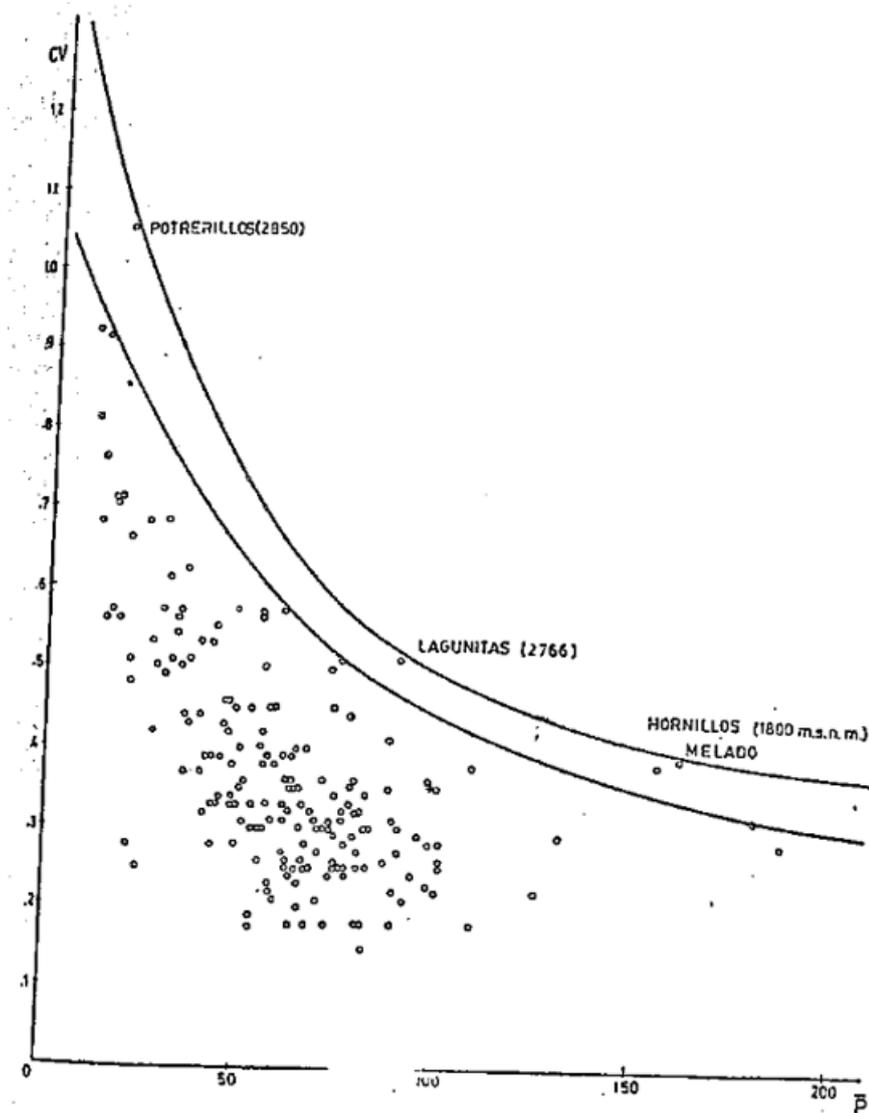
Stowhas (1983) luego de analizar 190 estaciones pluviométricas con un total de 6.504 años-estación de registro, llegó a graficar de manera independiente los valores del $K_{muestra}$ en función del coeficiente de variación (Cv), tal como se puede apreciar en la Figura 5 y los valores del coeficiente de variación (Cv) en función de la media de máximos anuales (μ_y) de cada estación, tal como se puede apreciar en la Figura 6 y de trazar una envolvente superior en cada gráfica, pudo establecer una relación lógica (la ecuación de la envolvente) entre el coeficiente de variación y el $K_{muestra}$ respectivamente con la media.

Figura 5 Parámetro K en función del C_v de precipitación en Chile



Fuente: Stowhas (1983).

Figura 6 *Cv en función de μ de precipitación en Chile*



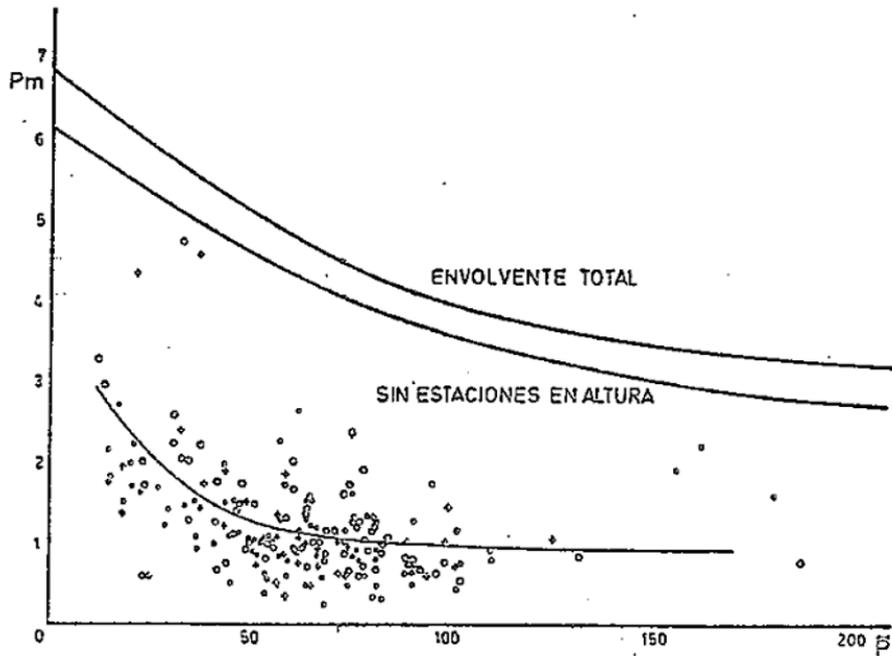
Fuente: Stowhas (1983).

Adicionalmente, dado que la precipitación máxima registrada es expresada de forma adimensional como:

$$P_m = \frac{P_{max} - \mu_y}{\mu_y} = \phi(\mu_y) \quad (2.6)$$

También graficó esta variable en función de la media, tal como se puede apreciar en la Figura 7 y trazó 2 envolventes superiores, una de ellas considerando estaciones en altura y otra sin considerar dichas estaciones.

Figura 7 P_m en función de μ de precipitación en Chile



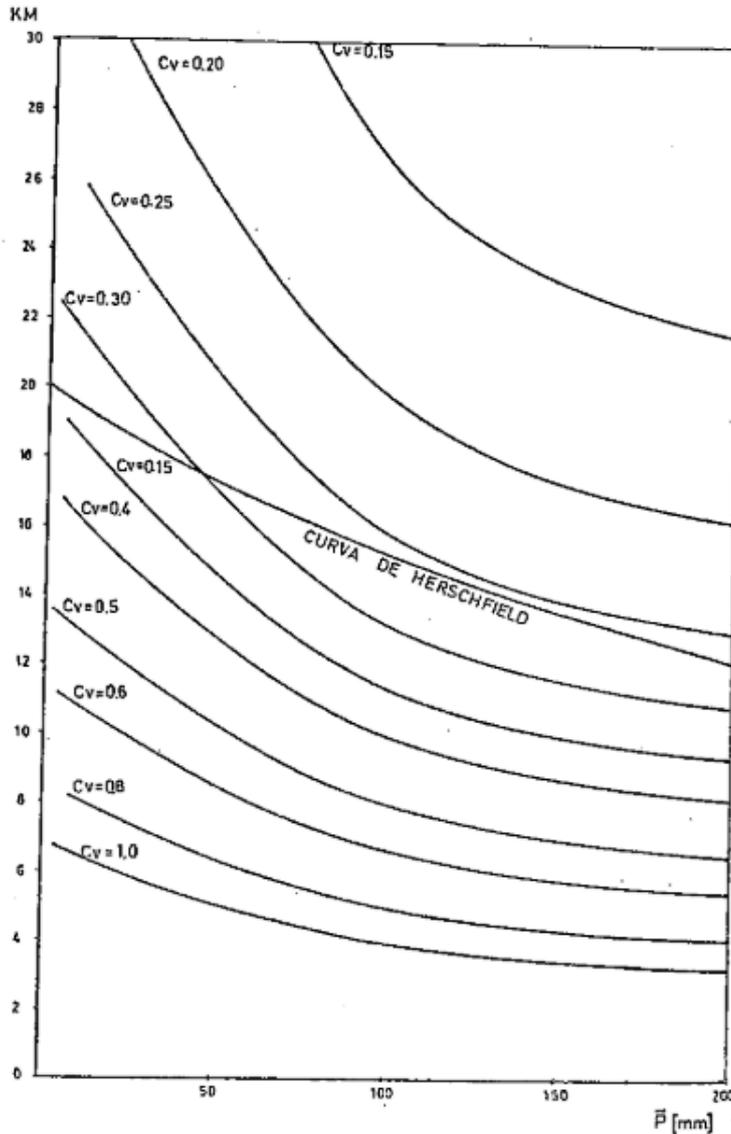
Fuente: Stowhas (1983).

Realizando una equivalencia algebraica, se puede llegar a la siguiente igualdad:

$$\frac{P_m}{Cv} = \frac{P_{max} - \mu_y}{\mu_y * Cv} = K \rightarrow \frac{P_m}{Cv} = K \rightarrow \frac{\phi(\mu_y)}{Cv} = K \quad (2.7)$$

De lo cual, se llega a formar una relación cruzada entre las 3 variables. Tal como se muestra en la Figura 8, definió un conjunto de envolventes para el parámetro K basadas en la amplificación de la envolvente de la precipitación máxima adimensional mediante la variación del coeficiente de variación, ello con el fin de encontrar una envolvente para el parámetro K . Dado ello, cuando $Cv = 1$ se da la igualdad entre el valor máximo adimensional (P_m) y el parámetro K .

Figura 8 Amplificación de la envolvente de la P_m mediante el Cv para aplicarla al parámetro K en Chile



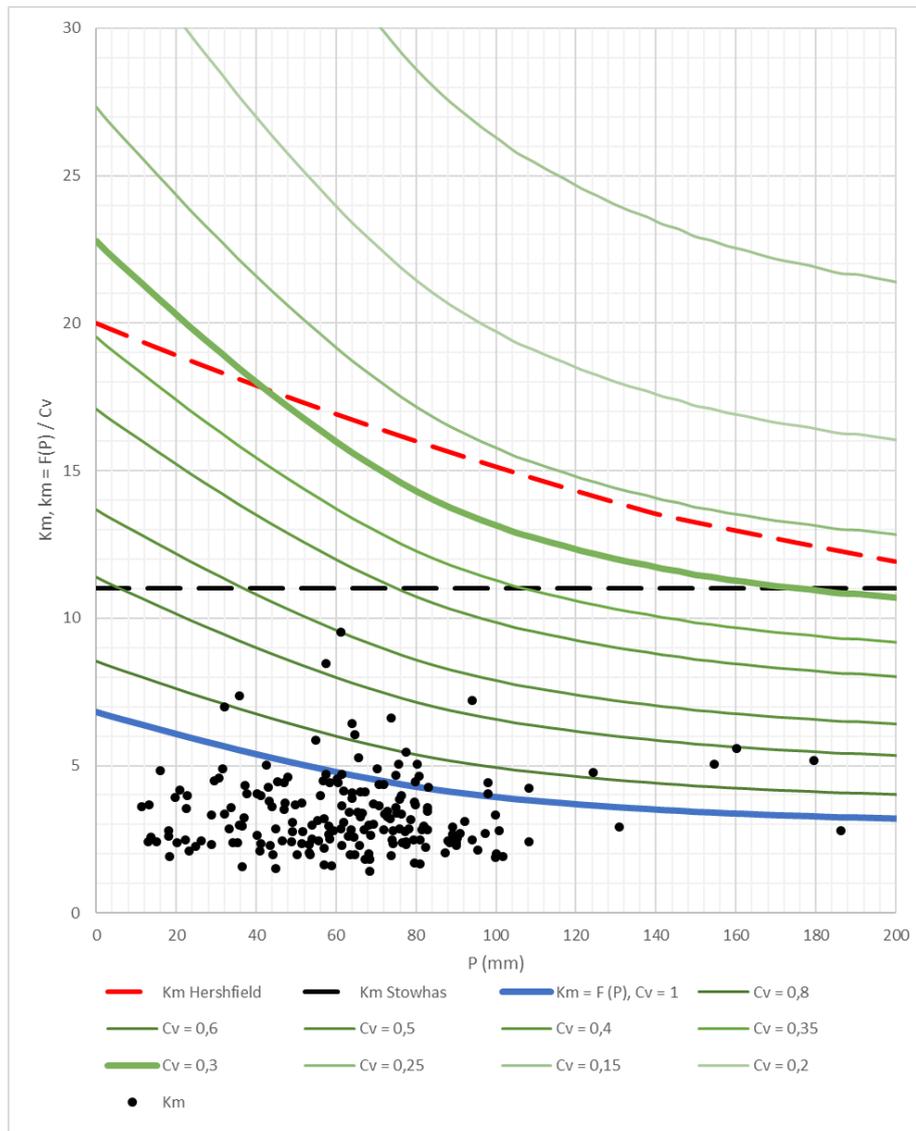
Fuente: Stowhas (1983).

Adicionalmente, si consideramos que la fórmula para calcular la PMP de Hershfield se puede reescribir de la siguiente manera:

$$PMP = \mu_y + K \sigma_y = \mu_y * (1 + K Cv) \quad (2.8)$$

En la Figura 9 se muestran graficados los valores de $K_{muestra}$ en función de la media en conjunto con las envolventes de la Figura 8, de lo cual se puede apreciar que pudo resultar suficiente para envolver a los datos, la curva que resulta al amplificar la envolvente original considerando un $Cv = 0.5$, a dicho valor de Cv le corresponde un valor de $K = 8,15$ de acuerdo con la envolvente de la Figura 5.

Figura 9 Amplificación de la envolvente de P_m mediante Cv y K muestral de estaciones en Chile



Fuente: Elaboración propia a partir de información de Stowhas (1983).

Sin embargo, se prefirió establecer que la envolvente adecuada sería la curva que corresponde a la que resulta de amplificar la envolvente original considerando un $Cv = 0.3$ (cuando Cv es menor K se maximiza, tal como se pudo observar en el comportamiento de los datos en la Figura 5 y Figura 6), a dicho valor de Cv le corresponde un valor de $K = 9,4$ de acuerdo con la envolvente de la Figura 5.

Para ser lo más conservador posible, decidió considerar los valores más bajos de Cv posibles siendo un aproximado de $Cv = 0.15$, al cual le corresponde un valor

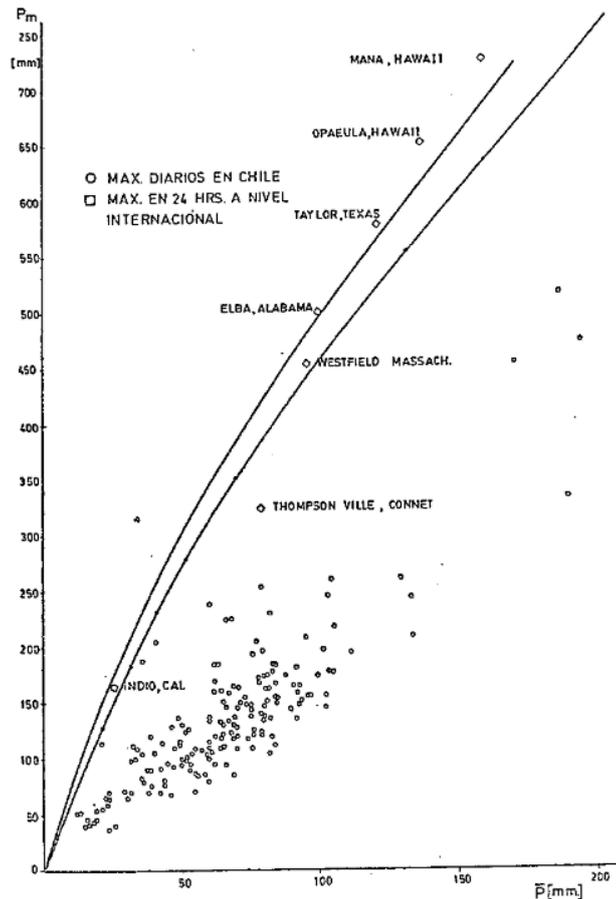
de $K = 10,3$ de acuerdo con la envolvente de la Figura 5, finalmente esto le llevaría a establecer un valor aún más conservador de $K = 11$.

Tras realizar el análisis mencionado anteriormente, se propuso que para estimar la PMP en Chile se use la ecuación:

$$PMP = \mu_y + 11 * \sigma_y \quad (2.9)$$

En la Figura 10, se graficó las precipitaciones máximas registradas a nivel de Chile y también aquellas registradas a nivel internacional a la fecha (Hershfield, 1962), para luego establecer 2 envolventes superiores para los valores de precipitación máxima, una de ellas considerando estaciones en altura y otra sin considerar dichas estaciones.

Figura 10 Precipitaciones máximas registradas en Chile y a nivel internacional



Fuente: Stowhas (1983).

Para escribir una ecuación que estime la PMP en Chile, tomó de referencia la gráfica mencionada, por sus envolventes, dado que de esa manera el resultado usando la expresión de Chile se controla con otra curva envolvente de P_{max} vs P

apoyado con los datos máximos registrados en Chile y en el mundo, prefiriendo el mayor resultado. Así, Stowhas (2017) propuso estimar la PMP con las expresiones:

$$PMP = \mu_y * (4 + 3.8 * e^{-0.0069 * \mu_y^{1.141}}) \quad (2.10)$$

$$PMP = \mu_y * (3.5 + 3.65 * e^{-0.0076 * \mu_y^{1.102}}) \quad (2.11)$$

Donde la ecuación (2.10) sería aplicable a estaciones pluviométricas cordilleranas y la ecuación (2.11) sería aplicable a estaciones pluviométricas no cordilleranas. Stowhas recomendó estimar la PMP con el método estadístico, utilizando el valor más conservador que resulte entre las ecuaciones (2.9) y la que corresponda entre (2.10) o (2.11).

Por lo tanto, se deduce que este método, que se basa en datos chilenos y no busca sustituir al método hidrometeorológico debido a la arbitrariedad inherente a cualquier método basado en envolventes, ofrece en la mayoría de los casos estimaciones de la Precipitación Máxima Probable (PMP) menos conservadoras que el método de Hershfield. Además, estas estimaciones están más en línea con las pocas estimaciones hidrometeorológicas de la PMP disponibles o con el uso de un valor probabilístico resultante para el K máximo probable igual a 11.

2.4 Análisis de consistencia de una serie de datos de precipitación

Este tipo de análisis se realizan en series de datos completas, a partir de ellos podemos identificar si existen inconsistencias causadas por la ubicación de la estación de medición, cambio de los métodos o condiciones de recolección de información, intervención de embalses o cambios en el uso del agua en la cuenca donde se ubica una o más estaciones de medición. Para detectar inconsistencias en la información se debe considerar los análisis visuales, análisis de doble masa y análisis estadísticos.

2.4.1 Análisis visual

El análisis visual es un medio que permite explorar y comprender los datos de manera rápida y eficaz. Permite respaldar y acelerar el proceso de análisis. Mediante este análisis podemos formular una pregunta, obtener la respuesta y hacernos preguntas de seguimiento. A medida que se realiza, se desarrolla una historia de un resumen visual a otro, más adelante, se puede volver sobre la historia para reflexionar sobre esta, explorar más y compartirla. El análisis visual

permite avanzar en cualquier dirección y seguir sus pensamientos mientras ve los datos e interactúa con ellos (Tableau, s.f.).

2.4.2 Análisis de outliers

Un análisis de *outliers* se realiza al evaluar una hipótesis de normalidad cuando las observaciones que se sospechan son *outliers* (datos atípicos, ya sean mucho más altos o mucho más bajos que el resto de los datos de un conjunto) se remueven del grupo original de observaciones. Para identificar datos atípicos en series de datos hidrológicos, diversos autores han empleado la prueba de Grubbs (1950), dicha prueba se puede realizar para un solo valor atípico, dos valores atípicos en una cola o dos valores atípicos en colas opuestas en una muestra pequeña de datos.

Para realizar la prueba se deben poner en orden los n datos de la serie y_i de menor a mayor, calcular la media (μ) y la desviación estándar (σ) del conjunto de datos, a continuación, se debe calcular el estadístico de prueba G usando una de las siguientes ecuaciones:

- Para una prueba de dos colas:

$$G = \frac{\max_{1, \dots, n} |y_i - \mu|}{\sigma}$$

- Para una prueba de cola izquierda:

$$G = \frac{\mu - y_{min}}{\sigma}$$

- Para una prueba de cola derecha:

$$G = \frac{y_{max} - \mu}{\sigma}$$

A continuación, se debe comparar el valor calculado de G con el valor crítico de G , el cual se puede extraer de diversas tablas de acuerdo con el valor de significancia que le está dando a la prueba y el tipo de prueba que se desee ejecutar (es decir, una prueba de una o dos colas). El valor crítico de G se puede calcular manualmente mediante la siguiente ecuación:

$$G_{crítico} = \frac{N - 1}{N} \sqrt{\frac{t_{\alpha/(2N), N-2}^2}{N - 2 + t_{\alpha/(2N), N-2}^2}}$$

Donde: $t_{\alpha/(2N), N-2}$ es el valor crítico superior de una distribución t con $N - 2$ grados de libertad.

Nota: Para la prueba de una cola, reemplace $\alpha/(2N)$ con α/N .

Finalmente, al hacer la comparación de los valores de G se puede determinar si se acepta o se rechaza el valor atípico, ello de acuerdo con la siguiente lógica:

- $G < G_{crítico}$: Mantenga el punto en el conjunto de datos; no es un caso atípico.
- $G > G_{crítico}$: Elimine el punto del conjunto de datos, es un caso atípico.

2.4.3 Análisis de estacionariedad

Las pruebas de estacionariedad son aplicadas a series de datos hidrológicos para determinar si dichas series presentan o no tendencias, dado que un requisito para analizar estadísticamente las series de datos es que los mismos no presenten tendencias. Para identificar aquellas tendencias que generan no estacionariedad en series de datos hidrológicos, numerosos autores han orientado sus estudios a la identificación de estas tendencias en series, la prueba más utilizada actualmente en la identificación de tendencias en series de variables hidrometeorológicas en general es la de Mann-Kendall, el inconveniente que esta prueba presenta es que la autocorrelación existente en algunas de estas series afecta los resultados de la prueba, causando problemas en la detección y evaluación del nivel de significación (Baraquet, 2018).

“La presencia de autocorrelación positiva tiende a aumentar la probabilidad de detección de tendencia cuando, en efecto, ésta no existe, mientras que la autocorrelación negativa tiende a disminuirla (López et al., 2007).”

Para corregir el efecto de la autocorrelación serial en la aplicación de la prueba Mann-Kendall, Yue, et al. (2002) presenta una técnica de pre-blanqueo libre de tendencia (Trend Free Pre-Whitening –TFPW- por sus siglas en inglés) que remueve la correlación serial sin alterar la tendencia presente en la serie. Consiste en remover de la serie estandarizada la tendencia, estimada por algún método robusto (Sen, 1968), antes de calcular el coeficiente de autocorrelación de retardo r_1 . De esta forma se resuelve el problema de la influencia de la tendencia en la estimación del coeficiente autocorrelación, mientras que la remoción de la tendencia no afecta al proceso autorregresivo subyacente.

La tendencia se asume lineal, esta se elimina de la siguiente manera:

$$Y_t = X_t - T_t = X_t - b_t$$

Se estima el coeficiente r_1 de la serie Y_t y si éste no es significativamente distinto de cero, la serie se considera independiente y la prueba Mann-Kendall se aplica a la serie original. Si r_1 es significativamente distinto de cero, la serie Y_t se blanquea:

$$Y'_t = Y_t - r_1 Y_{t-1}$$

La tendencia estimada y la serie Y'_t , compuesta por residuos que deben ser independientes, se combinan como como:

$$Y''_t = Y'_t + T_t$$

La prueba Mann-Kendall se finalmente se aplica a la serie Y''_t , evitando de esta forma el efecto de la correlación serial.

En la prueba de Mann-Kendall, para cada dato de la serie ordenada cronológicamente, se calcula cuantos datos posteriores son superiores a cada uno de los valores mediante lo siguiente:

$$S = \sum_{i=1}^{n-1} s(i)$$

En donde $s(i)$ corresponde al número de los valores de $X_j > X_i$, siendo $i < j \leq n$.

Por otro lado, se determinan cuantos valores posteriores son menores a cada uno de los valores mediante lo siguiente:

$$T = \sum_{i=1}^{n-1} t(i)$$

En donde $t(i)$ es el número de valores $X_j < X_i$, siendo $i < j \leq n$.

Con los valores de S y T calculados, se define el índice $I = S - T$, que debe ser próximo a cero si se verifica la estacionariedad. Se tienen las siguientes condiciones a verificar:

- $I < I_{\alpha}$, esto se refiere a que el I calculado sea menor que un valor de I crítico tabulado en función del tamaño de la muestra y según diferentes niveles de significación α .
- Para tamaños de muestra (n) mayores a 10, el valor del estadístico se aproxima a una distribución normal. Por lo que se debe utilizar una forma

estandarizada del estadístico en donde se transforma el índice I en el índice K, definido como:

$$K = \frac{I_{cr}-1}{\left(\frac{n \cdot (n-1) \cdot (2n+5)}{18}\right)^{0.5}} \quad \text{si } I > 0$$

$$K = 0 \quad \text{si } I = 0$$

$$K = \frac{I_{cr}+1}{\left(\frac{n \cdot (n-1) \cdot (2n+5)}{18}\right)^{0.5}} \quad \text{si } I < 0$$

En la Tabla 1 son presentados los valores de I_{cr} son obtenidos de una distribución Normal.

Tabla 1 Valores I_{cr} para la prueba de Mann - Kendall

I _{cr}	α
2.33	0.01
1.64	0.05
1.28	0.1

Fuente: Baraquet (2018)

2.4.4 Análisis de homogeneidad

Se trata de una prueba no paramétrica propuesta por Pettitt (1979), que se basa principalmente en rangos y hace caso omiso de la condición de normalidad de la serie que se está evaluando. Está basada en el orden de rangos de los valores y_i. El estadístico que se emplea para esta prueba se define como:

$$X_d = 2 \sum_{i=1}^d r_i - d(n + 1) \text{ para } d = 1, 2, \dots, n$$

Donde la variación fuerte en la serie analizada es detectada cuando el valor del año m cumple:

$$X_m = \max_{1 \leq d \leq n} |X_d|$$

2.4.5 Análisis de independencia serial

La prueba presentada por Ljung y Box (1978) también es conocida como la prueba Q de Ljung-Box, y está estrechamente relacionada con la prueba de Box-Pierce. La prueba estadística de Box-Pierce es una versión simplificada de la estadística de Ljung-Box para los cuales los estudios de simulación posteriores han demostrado un rendimiento deficiente.

La prueba de Ljung-Box es aplicable en la econometría y otras aplicaciones de análisis de series temporales, como puede ser el caso del análisis de series de datos de precipitación. La prueba parte de 2 hipótesis:

- H_0 : Los residuos se distribuyen de forma independiente; no exhiben correlación serial (hipótesis nula);
- H_A : Los residuos no se distribuyen de forma independiente; exhiben correlación serial.

Idealmente, nos gustaría no rechazar la hipótesis nula. Es decir, nos gustaría ver que el valor p de la prueba sea mayor que 0.05 porque esto significa que los residuos de nuestro modelo de series de tiempo son independientes, lo que a menudo es una suposición que hacemos al crear un modelo.

La medida estadística Q para la prueba de Ljung-Box es la siguiente:

$$Q = n(n + 2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n - k}$$

Dónde:

- n = tamaño de la muestra;
- $\hat{\rho}_k$ = autocorrelación de la muestra en el retraso k ;
- h = es el número de retardos que se están probando.

El estadístico de prueba Q sigue una distribución de chi-cuadrado con h grados de libertad. Dado un nivel de significación α , la región crítica para el rechazo de la hipótesis de aleatoriedad y decir que los residuos del modelo no se distribuyen independientemente si es:

$$Q > X_{1-\alpha, k}^2$$

Donde $X_{1-\alpha, k}^2$ es la α -cuantil de la distribución chi-cuadrado con h grados de libertad.

2.5 Clasificación regional basada en *clustering*

El *clustering* (agrupamiento) o análisis *cluster* es el método estadístico multivariable estándar utilizado para dividir un grupo de datos en diferentes grupos con características similares o que guarden una relación. Por lo general se utiliza este método para formar regiones en análisis de frecuencia. Un vector de datos característico es asociado a cada estación de medición de datos

hidrometeorológicos, y las estaciones son particionadas en diferentes grupos, dependiendo de la similitud entre sus vectores. Este vector de datos de cada estación puede incluir las estadísticas, características o combinaciones entre ambas. Se considera al análisis *cluster* basado en características de las estaciones como el método más práctico para formar regiones con elementos de características similares a partir de un gran número de datos o estaciones (Romesburg, 2004).

Existen múltiples *index* (índices) que se pueden tomar en cuenta para realizar las agrupaciones, por ejemplo, las propiedades geométricas, estadísticas de las series de datos, el número de objetos de datos y las mediciones de disimilitud o similitud (Charrad et al., 2014).

2.5.1 Índices estadísticos

A continuación, se describirán los índices considerados:

2.5.1.1 Índice CH

El índice de Calinski y Harabasz (1974) también conocido como índice CH se define por:

$$CH(q) = \frac{\text{trace}(Bq)/(q-1)}{\text{trace}(Wq)/(n-q)}$$

El valor de q , que maximiza $CH(q)$, se considera que especifica el número de *clusters*.

2.5.1.2 Índice Duda

Duda y Hart (1973) propusieron un criterio de relación $Je(2)/Je(1)$, donde $Je(2)$ es la suma de errores al cuadrado dentro de los *clusters* cuando los datos se dividen en dos grupos, y $Je(1)$ da los errores al cuadrado cuando solo hay un *cluster* presente. Se define de la siguiente manera:

$$Duda = \frac{Je(2)}{Je(1)} = \frac{W_k + W_l}{W_m}$$

Se supone que los cúmulos C_k y C_l se fusionan para formar C_m .

Gordon (1999) indicó que el número óptimo de *clusters* es el valor más pequeño q tal que:

$$Duda \geq 1 - \frac{2}{\pi p} - z \sqrt{\frac{2 \left(1 - \frac{8}{\pi^2 p}\right)}{n_m p}} = \text{critValue_Duda}$$

Donde z es una variable estandarizada. Se probaron varios valores para z y se obtuvieron los mejores resultados cuando el valor se estableció en 3,20 (Milligan & Cooper, 1985).

2.5.1.3 Índice Pseudot2

Duda y Hart (1973) propusieron también otro índice, Pseudo t_2 , que sólo puede aplicarse a métodos jerárquicos. Se calcula utilizando la siguiente ecuación:

$$Pseudot2 = \frac{V_{kl}}{\frac{W_k + W_l}{n_k + n_l - 2}}$$

Donde $V_{kl} = W_m - W_k - W_l$, si $C_m = C_k \cup C_l$.

Gordon (1999) indicó que el número óptimo de *clusters* es el valor más pequeño q tal que:

$$Pseudot2 \leq \left(\frac{1 - \text{critValue_Duda}}{\text{critValue_Duda}} \right) \times (n_k + n_l - 2)$$

2.5.1.4 Índice C

El índice C fue revisado en Hubert y Levin (1976). Se calcula utilizando la siguiente ecuación:

$$Cindex = \frac{S_w - S_{min}}{S_{max} - S_{min}}, S_{min} \neq S_{max}, Cindex \in (0, 1)$$

Donde:

- S_{min} = es la suma de las N_w distancias más pequeñas entre todos los pares de puntos de todo el conjunto de datos;
- S_{max} = es la suma de las N_w distancias más grandes entre todos los pares de puntos de todo el conjunto de datos.

El valor mínimo del índice se utiliza para indicar el número óptimo de *clusters* (Gordon, 1999; Milligan & Cooper, 1985).

2.5.1.5 Índice Gamma

El índice Gamma representa una adaptación del estadístico Gamma de Goodman y Kriskal para su uso en situaciones de agrupamiento (Baker & Hubert, 1975).

Se hacen comparaciones entre todas las disimilitudes dentro del grupo y todas las disimilitudes entre el grupo. Se considera que una comparación es concordante si una disimilitud dentro del grupo es estrictamente menor que una disimilitud entre grupos; Las igualdades entre los miembros de dos conjuntos de diferencias no se tienen en cuenta en la definición del índice (Gordon, 1999). Este índice se calcula utilizando la siguiente ecuación:

$$Gamma = \frac{s(+)-s(-)}{s(+)+s(-)}$$

Donde:

- $s(+)$ = representa el número de veces en que dos puntos no agrupados tenían una distancia mayor que dos puntos que estaban en el mismo grupo, es el número de comparaciones concordantes;
- $s(-)$ = representa lo opuesto, es el número de comparaciones discordantes.

El valor máximo del índice se utiliza para indicar el número óptimo de *clusters* (Milligan & Cooper, 1985).

2.5.1.6 Índice Beale

Baele (1969) propuso el uso de una relación F para probar la hipótesis de la existencia de *clusters* q_1 versus q_2 en los datos ($q_2 > q_1$). Este índice se calcula utilizando la siguiente ecuación:

$$Baele = F \equiv \frac{\left(\frac{V_{kl}}{W_k + W_l}\right)}{\left(\left(\frac{n_m - 1}{n_m - 2}\right) 2^{\frac{2}{p}} - 1\right)}$$

Donde $V_{kl} = W_m - W_k - W_l$. Se asume que los *clusters* C_k y C_l se fusionan para formar C_m .

El número óptimo de *clusters* se obtiene comparando F con una distribución $F_{p,(n_m-2)p}$. La hipótesis nula de un solo *cluster* se rechaza para valores significativamente grandes de F (Gordon, 1999).

2.5.1.7 Índice CCC

El Criterio de Agrupamiento Cúbico (CCC por sus siglas en inglés) es la prueba estadística proporcionada por el paquete de software SAS (Sarle, 1983). Este índice se calcula utilizando la siguiente ecuación:

$$CCC = \ln \left[\frac{1 - E(R^2)}{1 - R^2} \right] \frac{\sqrt{\frac{np^*}{2}}}{(0.001 + E(R^2))^{1.2}}$$

Donde:

- $R^2 = 1 - \frac{\text{traza}(X^T X - \bar{X}^T Z^T Z \bar{X})}{\text{traza}(X^T X)}$;
- $X^T X$ = Matriz de suma total de cuadrados y productos cruzados de muestra total de la matriz (p x p);
- $\bar{X} = (Z^T Z)^{-1} Z^T X$;
- Z es una matriz indicadora de *cluster* (nxq) con el elemento $z_{ik} = 1$ si la observación i pertenece al *cluster* k y $z_{ik} = 0$ en otro caso;
- $E(R^2) = 1 - \left[\frac{\sum_{j=1}^{p^*} \frac{1}{n+u_j} + \sum_{j=p^*+1}^p \frac{u_j^2}{n+u_j}}{\sum_{j=1}^p u_j^2} \right] \left[\frac{(n-q)^2}{n} \right] \left[1 + \frac{4}{n} \right]$;
- $u_j = \frac{s_j}{c}$;
- s_j = raíz cuadrada del valor propio j de $X^T X / (n - 1)$;
- $c = \left(\frac{v^*}{q} \right)^{\frac{1}{p^*}}$;
- $v^* = \prod_{j=1}^{p^*} s_j$;
- p^* se elige para ser el entero más grande que q tal que tal que u_{p^*} no es menor que uno.

El valor máximo del índice se utiliza para indicar el número óptimo de *clusters* en el conjunto de datos (Milligan & Cooper, 1985).

2.5.1.8 Índice Ptbiserial

Este índice, examinado por Milligan (1980, 1981) y Kraemer (1982), es una correlación punto-biserial entre la matriz de disimilitud de entrada bruta y una matriz correspondiente que consta solamente de 0 o 1. Se le asigna un valor de 0 si los dos puntos correspondientes están agrupados por el algoritmo, de lo contrario, se le asigna un valor de uno (Milligan, 1980). Dado que los valores

positivos más grandes reflejan un mejor ajuste entre los datos y la partición obtenida, el valor máximo del índice se utiliza para seleccionar el número óptimo de grupos en el conjunto de datos (Milligan & Cooper, 1985). El coeficiente de correlación biserial puntual se calcula utilizando la ecuación:

$$P_{tbiserial} = \frac{[\bar{S}_b - \bar{S}_w][N_w N_b / N_t^2]^{1/2}}{S_d}$$

Donde:

- $\bar{S}_w = S_w / N_w$;
- $\bar{S}_b = S_b / N_b$;
- S_d = desviación estándar de todas las distancias.

2.5.1.9 Índice Gplus

Este índice fue revisado por Rohlf (1974) y examinado por Milligan (1981). Se calcula utilizando la siguiente ecuación:

$$G_{plus} = \frac{2s(-)}{N_t(N_t - 1)}$$

Donde $s(-)$ es el número de comparaciones discordantes, en otras palabras, corresponde al número de veces en que dos puntos que estaban en el mismo grupo tenían una distancia mayor que dos puntos no agrupados (Milligan, 1981). Los valores mínimos del índice se utilizan para determinar el número óptimo de *clusters* en los datos (Milligan & Cooper, 1985).

2.5.1.10 Índice DB

El índice de Davies y Bouldin (1979) es una función de la relación de suma de la dispersión dentro del *cluster* a la separación entre grupos. Se calcula utilizando la siguiente ecuación:

$$BD(q) = \frac{1}{q} \sum_{k=1}^q \max_{k \neq l} \frac{\delta_k + \delta_l}{d_{kl}}$$

Donde:

- $k, l = 1, \dots, q$ = número de *clusters*;
- $d_{kl} = \sqrt[\nu]{\sum_{j=1}^{\nu} |c_{kj} - c_{lj}|^{\nu}}$ = distancia entre los centroides de los *clusters* C_k y C_l (para $\nu = 2$, d_{kl} es la distancia Euclídeana);

- $\delta_k = \sqrt[u]{\frac{1}{n_k} \sum_{i \in C_k} \sum_{j=1}^p |x_{ij} - c_{kj}|^u}$ = medida de dispersión de un *cluster* C_k (para $u = 2$, δ_k es la desviación estándar de la distancia de los objetos en el *cluster* C_k al centroide de este *cluster*).

Se considera que el valor de q minimiza $DB(q)$ especifica el número de *clusters* (Davies & Bouldin, 1979; Milligan & Cooper, 1985).

2.5.1.11 Índice Frey

El índice propuesto por Frey y Van Groenewoud (1972), sólo puede aplicarse a métodos jerárquicos. Como se muestra en la ecuación, es la relación de puntuaciones de diferencia de dos niveles sucesivos en la jerarquía. El numerador es la diferencia entre la media entre distancias de *cluster*, db , de cada uno de los dos niveles jerárquicos (nivel j y nivel $j + 1$) y el denominador es la diferencia entre la media dentro de las distancias del *cluster*, dw , de los dos niveles (nivel j y nivel $j + 1$).

Los autores propusieron, utilizando una puntuación de proporción de 1,00, identificar el nivel de *cluster* correcto. Las proporciones a menudo variaban por encima y por debajo de 1.00. Los mejores resultados se produjeron cuando se continuó con la agrupación hasta que la última relación cayó por debajo de 1,00. En este punto, el nivel de *cluster* anterior a este se tomó como partición óptima. Si la relación nunca cayó por debajo de 1.00, se asumió una solución de un *cluster* (Milligan & Cooper, 1985). Este índice se calcula utilizando la siguiente ecuación:

$$Frey = \frac{\bar{S}_{b_{j+1}} - \bar{S}_{b_j}}{\bar{S}_{w_{j+1}} - \bar{S}_{w_j}}$$

Donde:

- $\bar{S}_w = \frac{S_w}{N_w}$ = distancia media entre *clusters*;
- $\bar{S}_b = \frac{S_b}{N_b}$ = media de distancia dentro del *cluster*.

2.5.1.12 Índice Hartigan

El índice de Hartigan (Hartigan, 1975) se calcula utilizando la siguiente ecuación:

$$Hartigan = \left(\frac{\text{trace}(W_q)}{\text{trace}(W_{q+1})} - 1 \right) (n - q - 1)$$

Donde $q \in \{1, \dots, n - 2\}$. La diferencia máxima entre los niveles de jerarquía se toma como el número correcto de *clusters* en los datos (Milligan & Cooper, 1985).

2.5.1.13 Índice Tau

El índice Tau, fue revisado por Rohlf (1974) y probado por Milligan (1981), se calcula entre los elementos de dos matrices. La primera contiene las distancias entre los elementos y la segunda matriz de 0 o 1 indica, si cada par de puntos están o no dentro del mismo *cluster*. El índice Tau se calcula utilizando la siguiente ecuación:

$$Tau = \frac{s(+)-s(-)}{[(N_t(N_t-1)/2-t)(N_t(N_t-1)/2)]^{1/2}}$$

Donde:

- $s(+)$ = el número de veces en que dos puntos no agrupados tenían una distancia mayor que dos puntos que estaban en el mismo grupo, es decir, $s(+)$ es el número de comparaciones concordantes;
- $s(-)$ = el resultado inverso a $s(+)$ (Milligan, 1981), es el número de comparaciones discordantes;
- N_t = el número total de distancias y t es el número de comparaciones de dos pares de puntos donde ambos pares representan dentro de las comparaciones de conglomerados o ambos pares son entre comparaciones de grupos.

El valor máximo del índice se toma como indicando el número correcto de *clusters*.

2.5.1.14 Índice Ratkowsky

Ratkowsky y Lance (1978) propusieron un criterio para determinar el número óptimo de grupos basado en $\frac{\bar{S}}{q^{1/2}}$. El valor de \bar{S} es el promedio de las relaciones de $(BGSS_j/TSS_j)$ donde $BGSS$ significa la suma de cuadrados entre los *clusters* para cada variable y TSS para la suma total de cuadrados para cada variable (Hill, 1980).

El número óptimo de *clusters* es aquel valor de q para el que $\frac{\bar{S}}{q^{1/2}}$ tiene su valor máximo (Milligan & Cooper, 1985). Si el valor de q se hace constante, el criterio de Ratkowsky y Lance se puede reducir de $\frac{\bar{S}}{q^{1/2}}$ a \bar{S} (Hill, 1980).

Este índice se calcula utilizando la siguiente ecuación:

$$\text{Ratkowsky} = \frac{\bar{S}}{q^{1/2}}$$

Donde:

- $\bar{S} = \frac{1}{q} \sum_{j=1}^p \frac{BGSS_j}{TSS_j}$;
- $BGSS_j = \sum_{k=1}^q n_k (c_{kj} - \bar{x}_j)^2$;
- $TSS_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$.

2.5.1.15 Índice Scott

Scott y Symons (1971) introdujeron un índice, donde n es el número de elementos en el conjunto de datos, T es la suma total de cuadrados y W_q es la suma de cuadrados dentro de los q *clusters*. El índice Scott se calcula utilizando la siguiente ecuación:

$$\text{Scott} = n \log \frac{\det(T)}{\det(W_q)}$$

La diferencia máxima entre los niveles de jerarquía se utiliza para escoger el número correcto de *clusters* (Milligan & Cooper, 1985).

2.5.1.16 Índice Marriot

Marriot (1971) propuso el siguiente índice calculado utilizando la siguiente ecuación:

$$\text{Marriot} = q^2 \det(W_q)$$

La diferencia máxima entre niveles sucesivos se utiliza para determinar el número correcto de *clusters* (Milligan & Cooper, 1985).

2.5.1.17 Índice Ball

Ball y Hall (1965) propusieron un índice basado en la distancia media de los ítems a sus respectivos centroides de *cluster*. Se calcula utilizando la siguiente ecuación:

$$\text{Ball} = \frac{W_q}{q}$$

La mayor diferencia entre los niveles se utiliza para indicar el número óptimo de *clusters* (Milligan & Cooper, 1985).

2.5.1.18 Índice $Trcovw$

Este índice, examinado por Milligan y Cooper (1985), representa la traza de la matriz de covarianza agrupada dentro de los *clusters*. Se calcula utilizando la siguiente ecuación:

$$Trcovw = traza(COV(W_q))$$

Las puntuaciones de diferencia máxima entre niveles se utilizan para indicar el número óptimo de *clusters* (Milligan & Cooper, 1985).

2.5.1.19 Índice $Tracew$

Este índice ha sido uno de los índices más populares sugeridos para su uso en *clustering* (Edwards & Cavalli-Sforza, 1965; Friedman & Rubin, 1967; Fukunaga & Koontz, 1970; Milligan & Cooper, 1985; Orloci, 1967). Se calcula utilizando la siguiente ecuación:

$$Tracew = traza(W_q)$$

Dado que el criterio aumenta monótonamente con soluciones que contienen menos *clusters*, el máximo de las puntuaciones de las segundas diferencias se utiliza para determinar el número de *clusters* en los datos (Milligan & Cooper, 1985).

2.5.1.20 Índice Friedman

Este índice fue propuesto por Friedman y Rubin (1967), como base para un método de agrupamiento no jerárquico. Este índice se calcula utilizando la siguiente ecuación:

$$Friedman = traza(W_q^{-1}B_q)$$

La diferencia máxima de valores de este criterio se utiliza para indicar el número óptimo de *clusters* (Milligan & Cooper, 1985).

2.5.1.21 Índice McClain

El índice de McClain y Rao (McClain & Rao, 1975) consiste en la relación de dos términos, el primer término es el promedio dentro de la distancia del *cluster*, dividido por el número de distancias dentro del grupo, el segundo valor corresponde al del denominador y es el promedio entre la distancia del *cluster*

dividido por el número de distancias del *cluster*. Este índice se calcula utilizando la siguiente ecuación:

$$McClain = \frac{\bar{S}_w}{\bar{S}_b} = \frac{S_w/N_w}{S_b/N_b}$$

El valor mínimo del índice se utiliza para indicar el número óptimo de *clusters*.

2.5.1.22 Índice Rubin

Friedman y Rubin (1967) propusieron otro índice basado en la relación del determinante de la suma total de cuadrados y la matriz de productos cruzados al determinante de la matriz agrupada dentro del cúmulo. Este índice se calcula utilizando la siguiente ecuación:

$$Rubin = \frac{\det(T)}{\det(W_q)}$$

El valor mínimo de las segundas diferencias entre niveles se utiliza para seleccionar el número óptimo de *clusters* (Milligan & Cooper, 1985).

2.5.1.23 Índice KL

El índice KL fue propuesto por Krzanowski y Lai (1988) y está definido por la siguiente ecuación:

$$KL(q) = \frac{DIFF_q}{DIFF_{q+1}}$$

Donde:

$$DIFF_q = (q - 1)^{2/p} \text{traza}(W_{q-1}) - q^{2/p} \text{traza}(W_q)$$

El valor de q , que maximiza $KL(q)$, se considera que especifica el número óptimo de *clusters*.

2.5.1.24 Índice Silhouette

Rousseeuw (1987) introduced the silhouette index computed using Equation:

$$Silhouette = \frac{\sum_{i=1}^n S(i)}{n}, Silhouette \in [-1,1]$$

Donde:

- $S(i) = \frac{b(i)-a(i)}{\max\{a(i);b(i)\}}$

- $a(i) = \frac{\sum_{j \in \{C_s \setminus i\}} d_{ij}}{n_r - 1}$ es la disimilitud media del objeto i con respecto a todos los demás objetos del *cluster* C_r ;
- $b(i) = \min_{s \neq r} \{d_i C_s\}$;
- $d_i C_s = \frac{\sum_{j \in C_s} d_{ij}}{n_s}$ es la disimilitud media del objeto i con respecto a todos los demás objetos del *cluster* C_s .

El valor máximo del índice se utiliza para determinar el número óptimo de *clusters* en los datos (Kaufman & Rousseeuw, 1990).

2.5.1.25 Índice Gap

La estadística de brecha estimada propuesta por Tibshirani et al. (2001) y se calcula utilizando la siguiente ecuación:

$$Gap(q) = \frac{1}{B} \sum_{b=1}^B \log W_{qb} - \log W_q$$

Donde B es el número de conjuntos de datos de referencia generados mediante prescripción uniforme (Tibshirani et al., 2001) y W_{qb} es la matriz de dispersión interna definida como en el índice de Hartigan. El número óptimo de *clusters* se elige encontrando el q más pequeño tal que se cumpla lo siguiente:

$$Gap(q) \geq Gap(q + 1) - s_{q+1}, (q = 1, \dots, n - 2)$$

Donde:

- $s_q = sd_q \sqrt{1 + 1/B}$;
- sd_q es la desviación estándar de $\{\log W_{qb}\}, b = 1, \dots, B$: $sd_q = \sqrt{\frac{1}{B} \sum_{b=1}^B (\log W_{qb} - \bar{l})^2}$;
- $\bar{l} = \frac{1}{B} \sum_{b=1}^B \log W_{qb}$.

2.5.1.26 Índice Dindex

El índice Dindex (Lebart et al., 1995) se basa en la ganancia de agrupamiento en inercia intra-*cluster*. La inercia intra-*cluster* mide el grado de homogeneidad entre los datos asociados a un *cluster*. Calcula sus distancias en comparación con el punto de referencia que representa el perfil del *cluster*, es decir, el centroide del *cluster* en general. Está definido por la siguiente ecuación:

$$w(P^q) = \frac{1}{q} \sum_{k=1}^q \frac{1}{n_k} \sum_{x_i \in C_k} d(x_i, c_k)$$

Dadas dos particiones, P^{k-1} compuesta por $q-1$ clusters y P^k compuesta por k clusters, la ganancia de clustering en inercia intra-cluster se define como se muestra en la siguiente ecuación:

$$Gain = \omega(P^{q-1}) - \omega(P^q)$$

Esta ganancia del clustering debe minimizarse. La configuración óptima del cluster se puede identificar por una punta afilada que corresponde a una disminución significativa de las primeras diferencias de ganancia de agrupamiento frente al número de clusters. Esta punta afilada o gran salto de valores de ganancia se puede identificar por un pico significativo en las segundas diferencias de ganancia del clustering.

2.5.1.27 Índice Dunn

El índice de Dunn (Dunn†, 1974) define la relación entre la distancia mínima entre el inter-cluster y la distancia máxima entre el intra-cluster. Este índice viene dado por la siguiente ecuación:

$$Dunn = \frac{\min_{1 \leq i < j \leq q} d(C_i, C_j)}{\max_{1 \leq k \leq q} diam(C_k)}$$

Donde $d(C_i, C_j)$ es la función de disimilitud entre dos clusters C_i y C_j definido como $d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y)$ y $diam(C)$ es el diámetro de un cluster, que puede considerarse como una medida de dispersión del cluster. El diámetro de un cluster C se puede definir usando la ecuación:

$$diam(C) = \max_{x, y \in C} d(x, y)$$

Si el conjunto de datos contiene clusters compactos y bien separados, se espera que el diámetro de los clusters sea pequeño y que la distancia entre los clusters sea grande. Por lo tanto, el índice Dunn debe de maximizarse.

2.5.1.28 Índice Hubert

El estadístico Γ de Hubert (Hubert & Arabie, 1985) es el coeficiente de correlación serie puntual entre dos matrices cualesquiera. Cuando las dos matrices son

simétricas, Γ se puede escribir en su forma cruda como se muestra en la siguiente ecuación:

$$\Gamma(P, Q) = \frac{1}{N_t} \sum_{\substack{i=1 \\ i < j}}^{n-1} P_{ij} Q_{ij}$$

Donde:

- P es la matriz de proximidad del conjunto de datos;
- Q es una matriz $n \times n$ cuyo elemento (i, j) es igual a la distancia entre los puntos representativos (v_{ci}, v_{cj}) de los *clusters* donde pertenecen los objetos x_i y x_j .

Observamos que para $q = 1$ o $q = n$, el índice no está definido. La definición del estadístico normalizado $\bar{\Gamma}$ de Hubert viene dada por la siguiente ecuación:

$$\bar{\Gamma} = \frac{\sum_{\substack{i=1 \\ i < j}}^{n-1} (P_{ij} - \mu_P)(Q_{ij} - \mu_Q)}{\sigma_P \sigma_Q}$$

Donde $\mu_P, \mu_Q, \sigma_P, \sigma_Q$ son las respectivas medias y varianzas de las matrices P y Q .

Este índice toma valores entre -1 y 1 . Si P y Q no son simétricos, entonces todas las sumas se extienden sobre todas las entradas n^2 y $N_t = n^2$ (Bezdek & Pal, 1998).

Los valores altos de la estadística normalizada de $\bar{\Gamma}$ indica la existencia de grupos compactos. Así, en la gráfica de $\bar{\Gamma}$ normalizado versus q (q es el número de *clusters*), buscamos una punta significativa que corresponda a un aumento significativo de $\bar{\Gamma}$ normalizado ya que q varía de 2 a q_{max} , donde q_{max} es el número máximo posible de *clusters*. El número de *clusters* en los que se produce la punta es una indicación del número de *clusters* que subyacen a los datos (Halkidi et al., 2001).

2.5.1.29 Índice SD

La definición del índice de validez SD se basa en los conceptos de dispersión media para *clusters* y separación total entre *clusters*. Se calcula utilizando la siguiente ecuación:

$$SDindex(q) = \alpha Scat(q) + Dis(q)$$

El primer término, $Scat(q)$, indica la compacidad media de los grupos (es decir, la distancia intra-*cluster*). Un pequeño valor para este término indica *clusters* compactos. Este índice puede ser calculado utilizando la siguiente ecuación:

$$Scat(q) = \frac{\frac{1}{q} \sum_{k=1}^q \|\sigma^{(k)}\|}{\|\sigma\|}$$

Donde:

- $\sigma = (VAR(V_1), VAR(V_2), \dots, VAR(V_p))$ es el vector de varianzas para cada variable del conjunto de datos;
- $\sigma^{(k)} = (VAR(V_1^{(k)}), VAR(V_2^{(k)}), \dots, VAR(V_p^{(k)}))$ es el vector de varianzas para cada cúmulo C_k .

El segundo término $Dis(q)$, indica la separación total entre los q *clusters* (es decir, una indicación de la distancia entre grupos). Para ser calculado se utiliza la siguiente ecuación:

$$Dis(q) = \frac{D_{max}}{D_{min}} \sum_{k=1}^q \left(\sum_{z=1}^q \|c_k - c_z\| \right)^{-1}$$

Donde:

- $D_{max} = \max(\|c_k - c_z\|) \forall k, z \in \{1, 2, 3, \dots, q\}$ es la distancia máxima entre los centros de *cluster*;
- $D_{min} = \min(\|c_k - c_z\|) \forall k, z \in \{1, 2, 3, \dots, q\}$ es la distancia mínima entre los centros de *cluster*.

α es un factor de ponderación igual a $Dis(q_{max})$ donde q_{max} es el número máximo de *clusters* de entrada. El número de *clusters*, q , que minimiza el índice anterior, puede considerarse como un valor óptimo para el número de *clusters* presentes en el conjunto de datos.

2.5.1.30 Índice SDbw

La definición del índice de validez SDbw se basa en los criterios de compacidad y separación entre *clusters*. Se calcula utilizando la siguiente ecuación:

$$SDbw(q) = Scat(q) + Density.bw(q)$$

El primer término, $Scat(q)$, es el mismo calculado en el índice SD. El segundo término, $Density.bw(q)$, es la densidad entre grupos. Evalúa la densidad media en la región entre *clusters* en relación con la densidad de los *clusters* y se calcula utilizando la siguiente ecuación:

$$Density.bw(q) = \frac{1}{q(q-1)} \sum_{l=1}^{n_{ij}} \left(\sum_{j=1, i \neq j}^q \frac{density(u_{ij})}{\max(density(c_i), density(c_j))} \right)$$

Donde:

- u_{ij} es el punto medio del segmento de línea definido por los *clusters* centroides c_i y c_j ;
- $density(u_{ij})$ es calculado con la siguiente ecuación:

$$density(u_{ij}) = \sum_{l=1}^{n_{ij}} f(x_l, u_{ij})$$

Donde:

- n_{ij} es el número de tuplas que pertenecen a los cúmulos C_i y C_j ,
- $f(x_l, u_{ij})$ es igual a 0 si $d(x, u_{ij}) > Stdev$ y es igual a 1 en otro caso,
- $Stdev$ es la desviación estándar media de los *clusters* y se define por la siguiente ecuación:

$$Stdev = \frac{1}{q} \sqrt{\sum_{k=1}^q \|\sigma^{(k)}\|}$$

El número de *clusters* q que minimiza $SDbw$ se considera como el valor óptimo para el número de *clusters* en el conjunto de datos (Halkidi & Vazirgiannis, 2001).

Los resultados del análisis *cluster* (las regiones obtenidas) no se consideran como finales en la mayoría de los casos. Usualmente se realizan modificaciones y refinamientos de las regiones obtenidas con el análisis sobre la base de ajustes subjetivos para mejorar la coherencia física de las regiones y reducir la heterogeneidad en la medida de lo posible. Esto se hará basado en el criterio de la persona que está alisando los datos, tomando en cuenta sus conocimientos de los parámetros evaluados y experiencia previa en casos similares. Se recomiendan diferentes tipos de ajustes en las regiones (Hosking & Wallis, 1997):

- Trasladar una estación o un grupo de estaciones a otra región.
- Borrar una estación o algunas estaciones de la base de datos utilizada.
- Subdividir regiones.
- Eliminar una región, asignando las estaciones a diferentes regiones.
- Fusionar la región con otra región.
- Fusionar dos o más regiones.
- Utilizar más datos y redefinir regiones.

2.6 Regresión de datos

Para realizar una regresión se debe construir un modelo de regresión, el cual comprende un proceso iterativo que implica buscar modelar una variable dependiente a partir de un conjunto de variables explicativas. Un modelo de regresión se puede utilizar para predecir, explicar o controlar un fenómeno u evento observable, del cual se tenga información registrada.

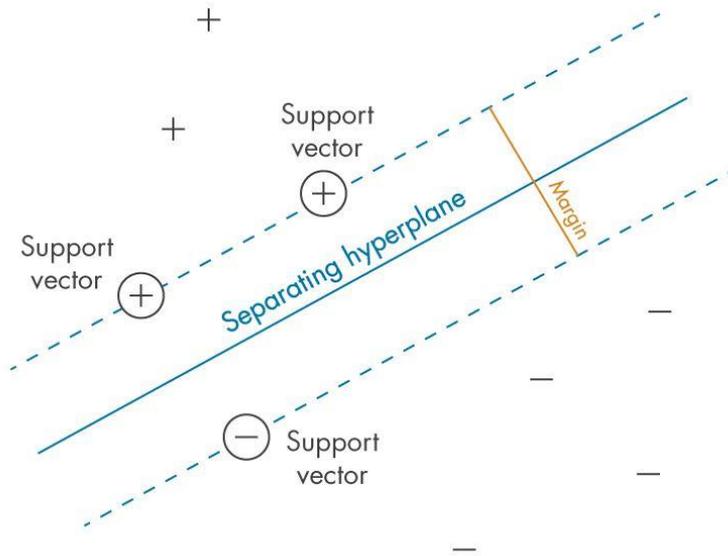
Hay diferentes tipos de modelos de regresión, tal como el lineal, el no lineal, el múltiple, el logístico, etc. Cada uno con sus propias consideraciones, parámetros de ajuste, características y supuestos. Es importante establecer y justificar la relación esperada entre cada variable explicativa y la variable dependiente antes del análisis. Dado ello, los modelos donde las relaciones establecidas no coinciden deben ser cuestionados.

2.6.1 Regresión con Máquina de Vectores de Soporte

Como se mencionó previamente, una regresión busca modelar una variable, para ello, se debe partir de datos concretos y aprender cual es la relación entre las variables explicativas y la variable dependiente, para lograr ello se pueden utilizar diferentes tipos de algoritmos, las máquinas de vectores de soporte (del inglés support-vector machines, SVM) son un conjunto de algoritmos asociados de aprendizaje supervisado que analizan datos para realizar tarea de clasificación y análisis de regresión (Boser et al., 1992; Cortes & Vapnik, 1995).

El objetivo del algoritmo SVM se basa en encontrar un hiperplano que separe de la mejor forma posible los datos puntuales en dos grupos a la vez, ello implica mantener el margen más amplio entre las dos clases, en la Figura 11 ello es representado por los signos más y menos.

Figura 11 Margen entre clases: el criterio que los SVM intentan optimizar



Fuente: MathWorks.

El margen es definido como la máxima división de la región paralela al hiperplano que no tiene datos puntuales interiores. El algoritmo SVM solo puede encontrar este hiperplano en problemas que permiten separación lineal; para otros casos, en la mayoría de los casos aplicables, el algoritmo llega a maximizar el margen flexible, debido a ello, permite un pequeño número de clasificaciones erróneas.

Capítulo III: Análisis exploratorio y de consistencia de datos de precipitación

En el presente capítulo se describe la formación de la base de datos unificada a partir de la información de precipitación de todas las estaciones que se encuentran dentro del territorio peruano provenientes de distintas fuentes; posterior a ello, se describe el procesamiento realizado para obtener las precipitaciones de 2 y 3 días a partir de las precipitaciones diarias; finalmente se expone los diferentes análisis de consistencia realizados, así como sus consideraciones y resultados.

3.1 Datos de precipitación de estaciones terrenas

3.1.1 Datos de precipitación de SENAMHI

Se descargó los datos de precipitación diaria del SENAMHI desde su web oficial en el apartado Datos, sección de Descarga de datos Meteorológicos a nivel nacional (Senamhi, s.f.), de la cual se pudieron descargar los datos de 292 estaciones pluviométricas de todas partes del Perú en archivos de texto individuales por cada estación.

En los archivos se pueden diferenciar 6 columnas, en la Figura 12 se puede apreciar la forma en la que están organizadas las columnas, las cuales corresponden al año, mes, día, valor de precipitación en milímetros, temperatura mínima en grados Celsius y temperatura máxima mínima en grados Celsius; los 2 últimos parámetros se descartaron ya que no formaban parte del análisis.

Figura 12 Estructura del archivo de la estación Caravelí

```
1963 10 1 -99.9 29 8.8
1963 10 2 -99.9 28 11.2
1963 10 3 -99.9 28.6 10.4
1963 10 4 -99.9 29 11.6
1963 10 5 -99.9 30 11.4
1963 10 6 -99.9 29 11.8
1963 10 7 -99.9 30 11.4
1963 10 8 -99.9 28.4 10.6
1963 10 9 -99.9 30 11.2
1963 10 10 -99.9 29.4 13
1963 10 11 -99.9 32.6 13
1963 10 12 -99.9 26.8 13
1963 10 13 -99.9 27.4 10.8
1963 10 14 -99.9 29 8.6
1963 10 15 -99.9 29 12.4
1963 10 16 -99.9 30 11.8
1963 10 17 -99.9 30.2 12.8
1963 10 18 -99.9 31.8 12
1963 10 19 -99.9 29.8 11
1963 10 20 -99.9 30 10.2
1963 10 21 -99.9 30 10
1963 10 22 -99.9 31.4 11
1963 10 23 -99.9 29.2 11.2
1963 10 24 -99.9 30 9.8
1963 10 25 -99.9 29.8 10.8
1963 10 26 -99.9 30.6 10
1963 10 27 -99.9 27.6 8.6
1963 10 28 -99.9 29 10
1963 10 29 -99.9 29 10.2
1963 10 30 -99.9 29 11.4
1963 10 31 -99.9 30 10
1963 11 1 -99.9 30.2 9.6
1963 11 2 -99.9 29.4 8.8
```

Fuente: Elaboración propia.

Se importaron los datos de precipitación diaria de las estaciones pluviométricas del SENAMHI al entorno de RStudio, para ello se leyeron los nombres y rutas de los archivos que contienen los datos, luego de eso se creó una función llamada `my_read_txt`, la cual es una variación de la función nativa de lenguaje R para leer archivos con formato de texto.

Utilizando la función `my_read_txt`, se importaron los datos de los archivos a partir de las rutas y luego se colocó la información en un *data frame* anidado (estructuras de datos de dos dimensiones que pueden contener datos de diferentes tipos y a su vez puede contener otros *data frames* dentro de sus propias celdas), el cual contía en su primera columna la ruta de ubicación del archivo, en su segunda columna el número del departamento seguido del nombre de la estación y en su tercera columna los *data frames* anidados de los datos de precipitación diaria.

Después de ello, se cargó la información de un archivo que contiene las características de las estaciones del SENAMHI y se acoplaron estas características al *data frame* con los datos de precipitación diaria, en la Figura 13 se muestra la estructura del *data frame* generado.

Figura 13 Estructura del data frame con los datos diarios de las estaciones del SENAMHI

depnombre	file	data	nombre	codigo	longitud	latitud	elevacion	departamento
1 01EL_PINTOR	input/01_SENAMHI/01EL_PINTOR.txt	1 variable	EL_PINTOR	152204	-78.516667	-5.750000	545.0	1
2 01MAGUNCHAL	input/01_SENAMHI/01MAGUNCHAL.txt	1 variable	MAGUNCHAL	152210	-78.183333	-5.883333	632.0	1
3 02BUENA_VISTA	input/01_SENAMHI/02BUENA_VISTA.txt	1 variable	BUENA_VISTA	435	-78.200000	-9.433333	419.0	2
4 02CHACCHAN	input/01_SENAMHI/02CHACCHAN.txt	1 variable	CHACCHAN	154107	-77.766667	-9.533333	2285.0	2
5 02MAYORARCA	input/01_SENAMHI/02MAYORARCA.txt	1 variable	MAYORARCA	150903	-77.416667	-10.166667	3200.0	2
6 02MILPO	input/01_SENAMHI/02MILPO.txt	1 variable	MILPO	150902	-77.233333	-9.883333	4400.0	2
7 02OCROS	input/01_SENAMHI/02OCROS.txt	1 variable	OCROS	155105	-77.400000	-10.400000	3230.0	2
8 02PIRA	input/01_SENAMHI/02PIRA.txt	1 variable	PIRA	154110	-77.700000	-9.583333	3570.0	2
9 02RECUAY	input/01_SENAMHI/02RECUAY.txt	1 variable	RECUAY	441	-77.450000	-9.716667	3462.0	2
10 02SIHUAS	input/01_SENAMHI/02SIHUAS.txt	1 variable	SIHUAS	154111	-77.650000	-8.566667	2716.0	2
11 03CURAHUASI	input/01_SENAMHI/03CURAHUASI.txt	1 variable	CURAHUASI	677	-72.733333	-13.550000	2763.0	3
12 04ANDAHUA	input/01_SENAMHI/04ANDAHUA.txt	1 variable	ANDAHUA	157310	-72.350000	-15.483333	3587.0	4

Fuente: Elaboración propia.

3.1.2 Datos de precipitación de la ANA

Se descargaron los datos de precipitación de 12 y 24 horas de la ANA, desde el Observatorio del Agua a Nivel Nacional, en el apartado de Oferta, sección de Pluviometría (Autoridad Nacional del Agua, s.f.), de la cual se pudieron descargar los datos de 962 estaciones pluviométricas de todas partes del Perú en archivos individuales por cada estación.

Los archivos, tal como se muestra en la Figura 14, contienen en sus primeras filas un resumen de la información de la estación pluviométrica a la que corresponden y en las demás filas se tienen 3 columnas, las cuales corresponden a la fecha, hora y valor de precipitación en milímetros.

Figura 14 Estructura del archivo de la estación Alao

	A	B	C
1			
2			
3			
4			
5			
6	Estación Alao (Codigo: 003308)		
7	Precipitación Acumulada En 12 Horas		
8			
9	Operador:	Servicio Nacional Meteorología E Hidrología	
10	WGS 84 Geográficas	Latitud: -6.518889 / Longitud: -76.73 / Altitud(msnm): 420	
11	Tipo:	Convencional / Climática	
12	Ambito Político	Dpto: San Martin / Prov: El Dorado / Dist.: San Martin	
13	Ambito Administrativo	AAA: Huallaga / ALA: Huallaga Central	
14	Unidad Hidrográfica:	Intercuenca Medio Alto Huallaga	
15	Nombre de la Fuente:	:	
16	Los datos descargados corresponden a información primaria sin control de calidad, la misma que está		
17			
18		FECHA	HORA
19		01/09/1972	19:00
20		02/09/1972	07:00
21		02/09/1972	19:00
22		03/09/1972	07:00
23		03/09/1972	19:00
24		04/09/1972	07:00
25		04/09/1972	19:00
26		05/09/1972	07:00
27		05/09/1972	19:00
28		06/09/1972	07:00
29		06/09/1972	19:00
30		07/09/1972	07:00
31		07/09/1972	19:00
32		08/09/1972	07:00
33		08/09/1972	19:00
34		09/09/1972	07:00
35		09/09/1972	19:00
36		10/09/1972	07:00

Fuente: Autoridad Nacional del Agua (s.f.).

Se importaron los datos de precipitación diaria de las estaciones pluviométricas al entorno de RStudio, para ello se leyeron los nombres y rutas de los archivos que contienen los datos, posterior a ello, se creó una función llamada `my_read_xlsx`, que es una variación de la función original utilizada para leer estos archivos.

Utilizando la función `my_read_xlsx` se importaron los datos y las características de las estaciones a partir de las rutas de los archivos, luego se agruparon los datos de 12 horas para obtener los datos diarios y se colocó la información en un *data frame*, en la Figura 15 se puede apreciar una parte de la tabla mencionada.

Figura 15 Estructura del data frame con datos diarios de las estaciones de la ANA

	estaciones	codigo	latitud	longitud	altitud	data
1	ABANCAY12	605	-13.608056	-72.87028	2750	1 variable
2	ACARI12	733	-15.400000	-74.61667	200	1 variable
3	ACJANAO12	5571	-13.150233	-71.68028	3487	1 variable
4	ACNOCOCHA12	646	-13.216667	-75.08333	4520	1 variable
5	ACO12	155106	-10.500000	-77.23333	3000	1 variable
6	ACOBAMBILLA12	156118	-12.666667	-75.31667	3795	1 variable
7	ACOMAYO12	687	-13.916900	-71.68360	3160	1 variable
8	ACOMBAMBA12	659	-12.837500	-74.56778	3236	1 variable
9	ACORA12	644	-13.783333	-75.36667	1890	1 variable
10	ACOSTAMBO12	625	-12.342333	-75.03562	3650	1 variable
11	AGUADA_BLANCA12	797	-16.233333	-71.35000	3855	1 variable
12	AGUAYTIA12	462	-9.033333	-75.50000	338	1 variable

Fuente: Elaboración propia.

3.2 Datos de precipitación de estaciones terrenas utilizados para la formación del producto grillado PISCO (crudo / completado / raster)

Se descargó datos de precipitación diaria desde la biblioteca de datos SENAMHI HSR PISCO del International Research Institute for Climate and Society: IRI (2019) en su versión estable más reciente, titulada “SENAMHI HSR PISCO Prec v2p1 stable daily Prec”, que contiene un *raster brick* (apilamiento de *rasters* con una resolución y extensión similar, que representan diferentes escenarios de una variable en la mayoría de casos) que abarca todo el Perú, donde están alojados los datos.

A partir de los *raster* de precipitación diaria del producto PISCO se generaron *rasters* de los máximos anuales de las precipitaciones diarias para cada año hidrológico disponible, los cuales se promediaron para generar un *raster* del valor medio de los máximos anuales. Este *raster* se elaboró para que sirva como insumo en un paso posterior para la interpolación espacial de los datos ya que nos proporciona información sobre cómo están distribuidos los valores de los máximos anuales en el territorio peruano.

Adicional a ello, se descargó la data de precipitación diaria cruda y procesada utilizada para la generación del producto grillado PISCOp (Aybar et al., 2020), estos datos se encontraron en una web propia del producto PISCOp V2.1.

Dado que se descargó la data cruda y procesada en archivos sin formato, se realizó una conversión mediante la página web FILExt (Simondi, 2000) para

acceder a los datos, tras esta conversión se obtuvieron archivos que podían ser leídos en RStudio.

Teniendo los archivos preparados, se pudieron importar los datos crudos y procesados de PISCO al entorno de RStudio, donde se le asignó una fecha a cada dato, puesto que estos se presentaban de forma secuencial ordenados cronológicamente, pero no se presentaba la fecha específica de cada dato, tomando en cuenta que los datos utilizados para PISCO parten desde el 01/01/1980 y observando la cantidad de datos, se les asignó la fecha correspondiente, las cuales llegaron hasta el 31/07/2018.

Dado que las series de datos de precipitación crudos y completados del producto PISCO no indicaban a que estación pluviométrica pertenecían específicamente, sino que solo se indican las coordenadas de la estación a la que pertenece cada serie histórica de datos, se hizo una asociación de series de datos.

La asociación consistió en identificar la estación pluviométrica a la que pertenecía cada serie de datos de la lista, inicialmente se unificó en una lista las estaciones de SENAMHI y la ANA asociándolas por un código único que la identifica, tomando como referencia los códigos establecidos por la ANA, posterior a ello, se definieron radios de búsqueda de 0.5° alrededor de las coordenadas de las series de datos de PISCO y se identificaron las 10 estaciones más cercanas de la lista unificada de SENAMHI y ANA que estuvieran dentro de dicho radio.

Posterior a ello, se compararon las medias de las series de datos diarios de la estación no identificada de PISCO con las cercanas estaciones identificadas, de esa manera se pudo asociar una a una aquellas estaciones de PISCO que tenían menor diferencia de medias en sus series con las estaciones de la lista unificada de SENAMHI y ANA, llegando a asociar exitosamente 440 de las 441 estaciones de PISCO. Aquella estación que no se pudo asociar fue añadida como una estación nueva a la lista unificada y se le asignó un código nuevo.

3.3 Mapas y shapes de SENAMHI

3.3.1 Regiones Homogéneas de Precipitación Máxima

Las Regiones Homogéneas de Precipitación Máxima (SENAMHI, 2020b) se descargaron del repositorio del Catálogo Nacional de Metadatos del Perú, ya que

se utilizará más adelante en capítulos posteriores para compararlas con los *clusters* de las estaciones formados en la presente tesis.

3.3.2 Mapa de Clasificación Climática del Perú

El Mapa de Clasificación Climática del Perú (Senamhi, 2020a) se descargó del Catálogo de Metadatos Cartográficos, ya que se utilizará más adelante en capítulos posteriores para comparar las regiones basadas en la componente de lluvias con los *clusters* de las estaciones formados en la presente tesis.

3.4 Formación de series de máximas anuales

3.4.1 Selección de series confiables de máximas anuales de 24; 48 y 72 horas

Luego de realizar la evaluación de las precipitaciones de 1, 2 y 3 días a partir de las precipitaciones diarias con los datos del SENAMHI, de la ANA y de PISCO tanto con su versión cruda como completada, se procedió a extraer los máximos mensuales y anuales de la base de datos más confiable, con el empleo de funciones que califican la completitud del registro en los meses de la temporada de lluvias:

- `year_hidrologico`: Esta función identifica a qué año hidrológico pertenecen las fechas asociadas a los datos diarios;
- `by_month`: Esta función contiene a la función `year_hidrologico` y agrupa los datos diarios por meses, identificando si cada uno de los meses de los años cuenta con información completa, es decir, si no tiene más de 6 días con datos faltantes, marca con un `TRUE` o `FALSE` si los años en cuestión cuentan con los meses críticos (Se consideran meses críticos a los meses de cada año hidrológico en los que se presenta mayor cantidad de precipitaciones, estos son diciembre, enero, febrero, marzo y abril (Endara, 2016)) completos o si cuenta con los meses completos, finalmente asigna un `score`, será 0 si el año hidrológico no tiene los meses críticos completos, será 1 si el año hidrológico tienen al menos los meses críticos completos y será 2 si el año hidrológico tiene sus 12 meses completos;
- `by_year`: Esta función contiene a la función `year_hidrologico` y agrupa los datos diarios por años, extrayendo los datos máximos anuales y las

fechas correspondientes a los máximos anuales, además calcula la cantidad de días con datos que tiene el año y su suma;

- *by_year1*: Esta función recolecta la información generada por las funciones *by_month* y *by_year* en un *data frame* anidado, dejando en este *data frame* solamente las columnas de año, precipitación máxima anual, fecha correspondiente al máximo y el score asignado, además, asigna un punto más de score al dato si es que el máximo del año hidrológico se encuentra dentro de los meses críticos.

Primero, se les aplicó a cada conjunto de datos de estaciones del SENAMHI, ANA y PISCO de forma independiente la función *by_month*, luego la función *by_year* y finalmente la función *by_year1*. Esto con el fin de identificar los valores máximos anuales de los años hidrológicos con información disponible, así como identificar sus fechas exactas de ocurrencia y de esa forma, asignarles un score para ayudar a discernir el dato de mayor confiabilidad, medida en que la base de datos más completa con mayor seguridad habrá registrado el máximo anual que prevalecería sobre los datos de las demás fuentes a fin de tener una base de datos unificada con los mejores datos posibles.

A continuación, se creó la función *results*, la cual funciona de la siguiente manera:

- *results*: Esta función selecciona los datos de acuerdo con su score, si el score del dato es 0, elimina el dato directamente, en otro caso escoge de entre las distintas fuentes, el dato con el mejor score posible o en caso el score sea igual para un año hidrológico escoge el dato mayor.

Para seleccionar los datos más confiables por cada año y estación, a los scores de los datos provenientes del SENAMHI, ANA y PISCO crudo se les multiplicó por 10, porque se tratan de datos sin procesar, de manera que los posibles scores de estas fuentes sean 0, 10, 20 o 30 para que al aplicar la función *results* a los conjuntos de datos, ella seleccione preferentemente a los datos de SENAMHI, ANA y PISCO crudo y que solo en caso que no se tenga un dato de ninguna de estas fuentes, tome el dato de PISCO completado, se toma esta fuente como última opción ya que sus datos han sido procesados, es decir, cuentan con un control de calidad y ello es el motivo de que en determinados casos no logran reflejar los eventos máximos de precipitación.

Previo a aplicar dicha función, se juntó la información de los *data frames* de las diferentes fuentes en un solo *data frame*, mediante la identificación de las

estaciones con códigos coincidentes, resultando en un conjunto de 989 estaciones y se procedió a aplicar la función *results* para obtener una base de datos unificada de los datos de las estaciones de las diferentes fuentes.

A continuación, se aplicó la función *results* y se empleó un pre-filtro de 30 años mínimos de información por estación para evitar que estaciones con escasa información pasen por el siguiente paso que correspondió al análisis de *outliers* de manera innecesaria.

En estudios previos de la PMP realizados en otros países (por tanto, bajo condiciones diferentes) se han considerado distintas cantidades de años (15, 25, 50) mínimos de información por estación (Campos Aranda, 2014; Casas et al., 2008; Rezacova et al., 2005), sin embargo, la WMO establece que se requiere un mínimo de 50 años de información para estimar la PMP de manera adecuada (WMO, 2009).

En el caso del Perú, si se consideraran 50 años como longitud mínima de información, se descartaría demasiadas estaciones, lo cual no es deseable para el análisis, por ello, tras hacer distintas pruebas con la cantidad de estaciones restantes y su distribución se decidió considerar 30 años como longitud mínima de información por estación para tener series de datos lo más amplias posible que puedan definir bien las medias y desviación estándar, buscando que los sucesos máximos atípicos no sean distorsionantes de la muestra.

Tras aplicar el pre-filtro de 30 años mínimos de información, restaron 636 estaciones.

Para la formación de las series de máximas anuales se tomaron las siguientes consideraciones:

- Que cuando se hace referencia a un año, se está tomando en cuenta el concepto de un año hidrológico, es decir, el año hidrológico en el Perú es un periodo de 12 meses, que abarca desde el 01 de septiembre del año anterior hasta el 31 de agosto del presente año (SENAMHI, 2020c);
- Que, en el caso de tener datos de 12 horas tomados a las 7 am y 7 pm de cada día, estos se agruparán de manera que el dato de las 7 am del día presente sumando con el dato de las 7 pm del día anterior corresponde la precipitación diaria del día presente, conforme a la práctica de la agencia meteorológica nacional, el SENAMHI;

- Que un mes con información completa es aquel mes al que no le falte más de 6 días de información, dado que por no ser objetivo de estudio las precipitaciones mensuales, simplemente se considera que la información perdida no contuvo los máximos diarios anuales;
- Se descartaron los datos de cada estación que pertenezcan a un año hidrológico del cual no se tengan los meses críticos con información completa y por tanto tampoco los 12 meses completos;
- A la estación Yuracmayo, proveniente del repositorio de la ANA, se le asignó el código 000000 ya que no le fue identificado un código asignado.
- Le fue colocado un 1 al final del código de la estación San Lorenzo que se ubica en el departamento de Loreto, y un 2 al final del código de la estación San Lorenzo que se ubica en el departamento de Tumbes, ya que el código anotado de estas estaciones diferentes es repetido, 152102;
- A la estación Cuñumbuque del repositorio de la ANA se le asignó la altitud de 450 msnm ya que no tenía altitud anotada en su ficha de registro, la cual se obtuvo con las coordenadas de la ubicación de la estación en Google Earth;

3.4.2 Base de datos de series máximas anuales

Se formó una base de datos de precipitaciones diarias a modo de tabla en RStudio, porque así en esta plataforma se hace más sencilla la manipulación de los datos de manera masiva, en ordenamiento, segmentación, filtrado, operaciones, generación de análisis estadísticos y gráficas.

A partir de la información diaria se obtuvieron las series de precipitaciones máximas de 1, 2 y 3 días de duración; y a partir de ellas, las series de 24, 48 y 72 horas de duración.

En primera instancia se descargaron los datos de las 3 fuentes, estas son el SENAMHI, la ANA y el producto PISCO.

Previo a la importación de datos al entorno de Rstudio, se creó la función `convertir_varios_dias` para agrupar los datos de precipitación diaria y estimar las precipitaciones de 2 y 3 días, ello para posteriormente obtener las precipitaciones de 24; 48 y 72 horas a partir de cada una de las antes mencionadas respectivamente.

3.5 Análisis de consistencia de series de máximas anuales

En cuanto al análisis de consistencia de las series de máximas anuales, se realizaron distintos tipos de evaluaciones, los cuales fueron el análisis visual, prueba de detección de datos atípicos (*outliers*), y los análisis de tendencias, homogeneidad e independencia serial, con el fin de verificar la calidad de la información recolectada de las diversas fuentes, retirar datos gruesamente atípicos por sospecha de mal registro, pero con la salvedad de conservar los registros estadísticamente inusuales, y depurar estaciones con series de datos no congruentes.

Previo a realizar los análisis estadísticos las series de máximos anuales fueron multiplicadas por un factor de transformación de acuerdo a la duración de exposición ante el evento de precipitación, para transformar a equivalentes de haberse dispuesto de datos con duraciones horarias o menores; es decir, para transformar de datos obtenidos con ventanas de tiempo con duración fija a datos obtenidos con ventanas de tiempo con duración móvil (WMO, 1994). Se utilizaron los siguientes factores de transformación para cada caso:

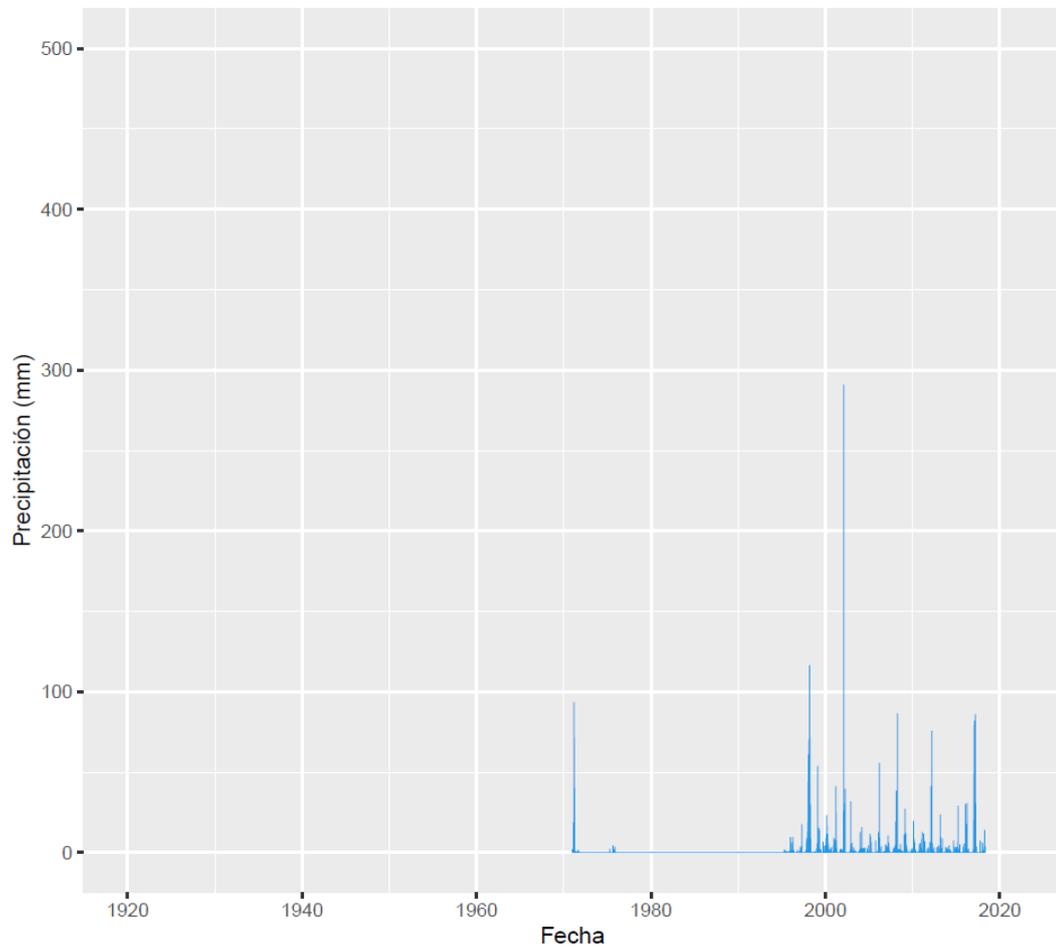
- El factor 1.13 para transformar precipitaciones diarias en precipitaciones de 24 horas;
- El factor 1.04 para transformar precipitaciones de 2 días en precipitaciones de 48 horas;
- El factor 1.03 para transformar precipitaciones de 3 días en precipitaciones de 72 horas.

3.5.1 Análisis visual

El análisis visual, consistió en una inspección con el principal propósito de detectar *outliers* muy exagerados, puesto que, al revisar las gráficas de los máximos anuales, se pueden identificar fácilmente.

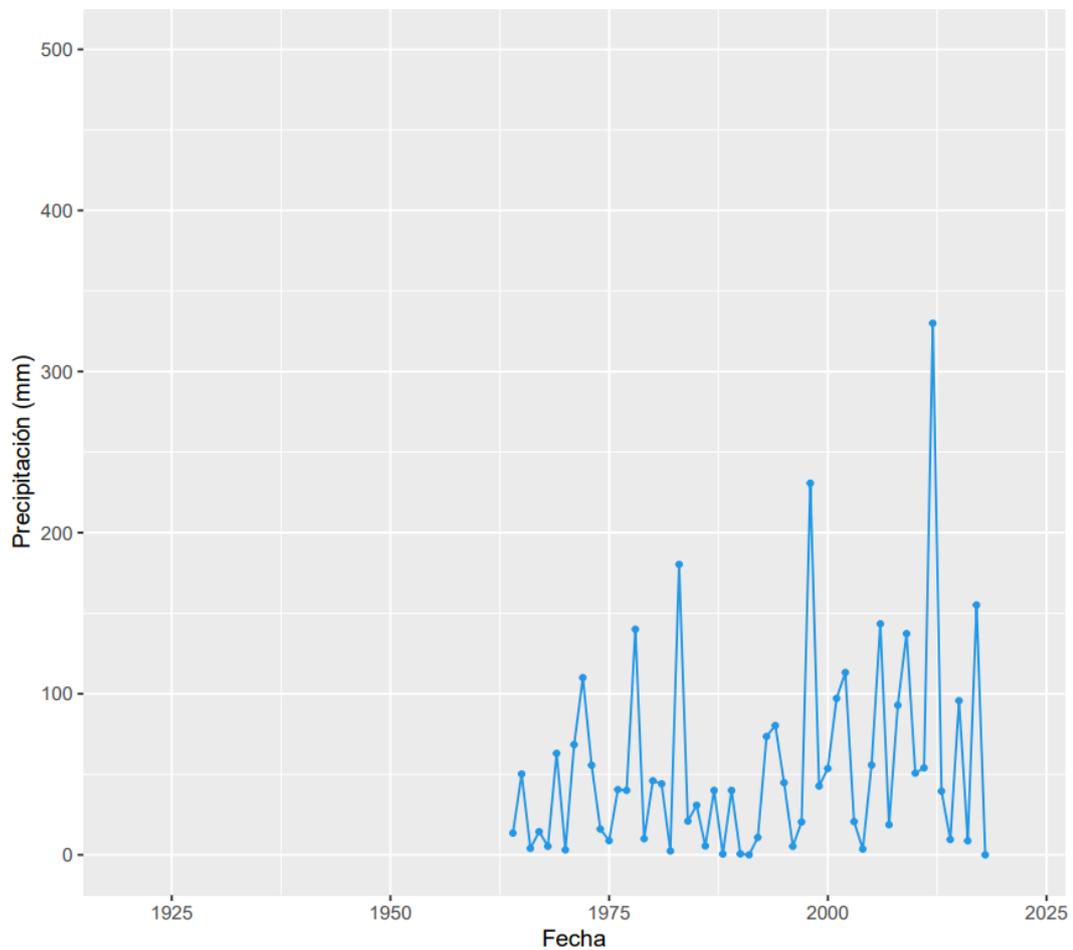
Se graficaron los datos de las distintas duraciones de todas las estaciones pluviométricas de las diferentes fuentes de datos por separado, como se observa en la Figura 16 y la Figura 17, para observar el comportamiento de los registros de lluvia e identificar estaciones que potencialmente tendrían información incongruente o atípica, dada la grandísima cantidad de estaciones por revisar, solo se observó aquellas estaciones que tuvieran gráficas muy irregulares y con saltos muy notorios.

Figura 16 Precipitación diaria de la estación Tinajones del Observatorio del Agua a Nivel Nacional de la ANA



Fuente: Elaboración propia.

Figura 17 Máximos anuales diarios de la estación Santa Rita de Castilla del Observatorio del Agua a Nivel Nacional de la ANA



Fuente: Elaboración propia.

Una vez obtenidas las gráficas antes mencionadas, se procedió a analizar de manera visual cada una de ellas y se pudo encontrar que en ciertas estaciones se presentan valores picos inexplicables, estos gráficos también fueron revisados en pasos posteriores para la detección de outliers con el fin de depurar datos fuera de la escala física creíble que puedan estar alterando drásticamente el análisis.

3.5.2 Detección de outliers

El análisis de atípicos (*outliers*) tuvo como objetivo principal detectar los valores fuera de escala física creíble en las series de máximos anuales, teniendo en cuenta como máximos posibles los máximos anuales de las estaciones Quincemil (código ANA: 6401), San Gabán (código ANA: 693) y Aguaytía (código ANA: 462); las cuales son estaciones que se tiene conocimiento que cuentan con las

precipitaciones más altas del país (Chavez & Takahashi, 2017; Espinoza et al., 2015).

El procedimiento de detección de *outliers* por lo general se realiza de manera visual o mediante un algoritmo previamente planteado que se ajuste mejor a las características de las series de datos con las que se va a trabajar, en el caso presente, ya que se trata de un análisis de precipitaciones máximas, no se pueden eliminar datos máximos de la manera en cómo se haría con una serie de otro tipo, ya que se busca que en la serie final se puedan tener estos datos superiores a lo esperado ya que dichos datos son los que finalmente revelarán la PMP.

Por lo antes expuesto, se hizo un análisis visual de diagramas boxplot de cada estación para descartar datos que salgan de toda lógica y puedan identificarse al ver el diagrama, sin embargo, en el caso de este análisis, se tuvo que haber revisado visualmente 636 diagramas bloxplot, lo cual no es práctico ni eficiente, por tal motivo se optó por elaborar un procedimiento que pueda descartar datos atípicos de cada serie.

Para realizar este análisis se elaboraron múltiples gráficas basadas en la media (μ), el máximo, el máximo centralizado y normalizado ($y_{max\ cyn} = \frac{y_{max} - \bar{y}}{\bar{y}}$), el coeficiente de variación (Cv) y el K muestral de las series de máximos anuales, con el fin de identificar incongruencias en el comportamiento de las gráficas, tomando como comportamiento esperado (delimitado por una envolvente o con una tendencia marcada), mostrado en las gráficas de investigaciones previas de la PMP (Campos Aranda, 2014; Casas et al., 2008, 2010; Hershfield, 1965; Stowhas, 1983), los gráficos que se elaboraron para este procedimiento fueron:

- Gráfico de puntos del coeficiente de variación en función de la media de los máximos anuales de las estaciones pluviométricas, acompañado de la envolvente de Stowhas para 24 horas. Esta gráfica permite visualizar el comportamiento de la nube de puntos, resaltando la forma de la frontera superior, la cual se llega a asemejar a la envolvente mencionada. Se muestra en la Figura 18 el caso previo al análisis de atípicos (*outliers*) y en la

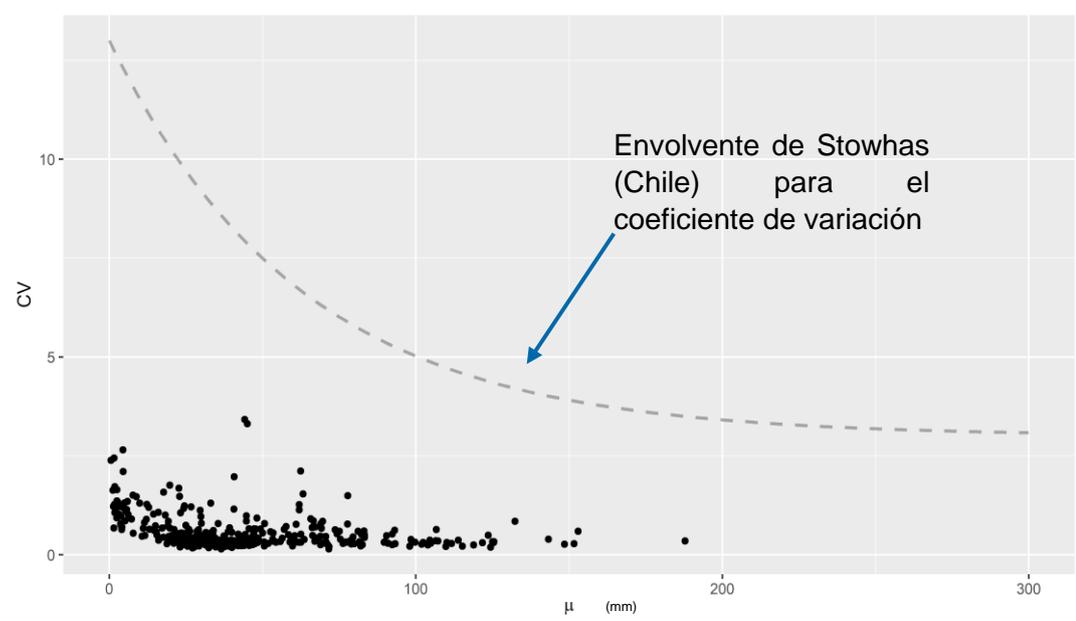
- Figura 19 el caso posterior al análisis de atípicos (*outliers*) para la duración de 24 horas;
- Gráfico de puntos del valor máximo centralizado y normalizado en función de la media de los máximos anuales de las estaciones pluviométricas, acompañado de la envolvente de Stowhas para 24 horas. Esta gráfica permite visualizar un comportamiento asintótico de la nube de puntos tanto al eje X como al Y, siendo la forma de la frontera superior no del todo similar a la envolvente mencionada. Se muestra en la Figura 20 el caso previo al análisis de atípicos (*outliers*) y en la Figura 21 el caso posterior al análisis de atípicos (*outliers*) para la duración de 24 horas;
- Gráfico de línea del valor máximo centralizado y normalizado de las estaciones pluviométricas en orden creciente de su media de máximos anuales. Esta gráfica permite visualizar el comportamiento esperado de los máximos centralizados y normalizados, el cual corresponde a mayores valores cuando la media es baja y valores cada vez más bajos conforme esta aumenta. Se muestra en la Figura 22 el caso previo al análisis de atípicos (*outliers*) y en la Figura 23 el caso posterior al análisis de atípicos (*outliers*) para la duración de 24 horas;
- Gráfico de boxplots de las series de máximos anuales de las estaciones en orden creciente de su media de máximos anuales. Esta gráfica permite visualizar el comportamiento esperado de las series de datos centralizados y normalizados, el cual corresponde a mayores valores cuando la media es baja y valores cada vez más bajos conforme esta aumenta, en contraste con el tipo de gráfica anterior, esta permite visualizar con mayor claridad la cantidad de datos atípicos que puedan existir. Se muestra en la Figura 24 el caso previo al análisis de atípicos (*outliers*) y en la Figura 25 el caso posterior al análisis de atípicos (*outliers*) para la duración de 24 horas;
- Gráfico de puntos del valor máximo en función de la media de los máximos anuales de las estaciones pluviométricas. Esta gráfica permite visualizar si la magnitud de los máximos guarda una coherencia dentro del rango físico creíble respecto a la media de su serie de máximos anuales, se puede apreciar una tendencia pseudo lineal de la mayoría de los puntos. Se muestra en la Figura 26 el caso previo al análisis de

atípicos (*outliers*) y en la Figura 27 el caso posterior al análisis de atípicos (*outliers*) para la duración de 24 horas;

- Gráficos de puntos del valor de $K_{muestral}$ de las estaciones en función de la media de máximos anuales de las estaciones pluviométricas, acompañado de la curva envolvente de Hershfield de 24 horas y otras envolventes generadas a partir de la ecuación de Koutsoyiannis & Xanthopoulos (1999) para 48 y 72 horas. Esta gráfica permite visualizar que el comportamiento de la mayoría de los puntos es similar a la envolvente de Hershfield pero en una magnitud más baja, dado que las curvas mostradas fueron propuestas para otra ubicación. Se muestra en la Figura 28 el caso previo al análisis de atípicos (*outliers*) y en la Figura 29 el caso posterior al análisis de atípicos (*outliers*) para la duración de 24 horas.

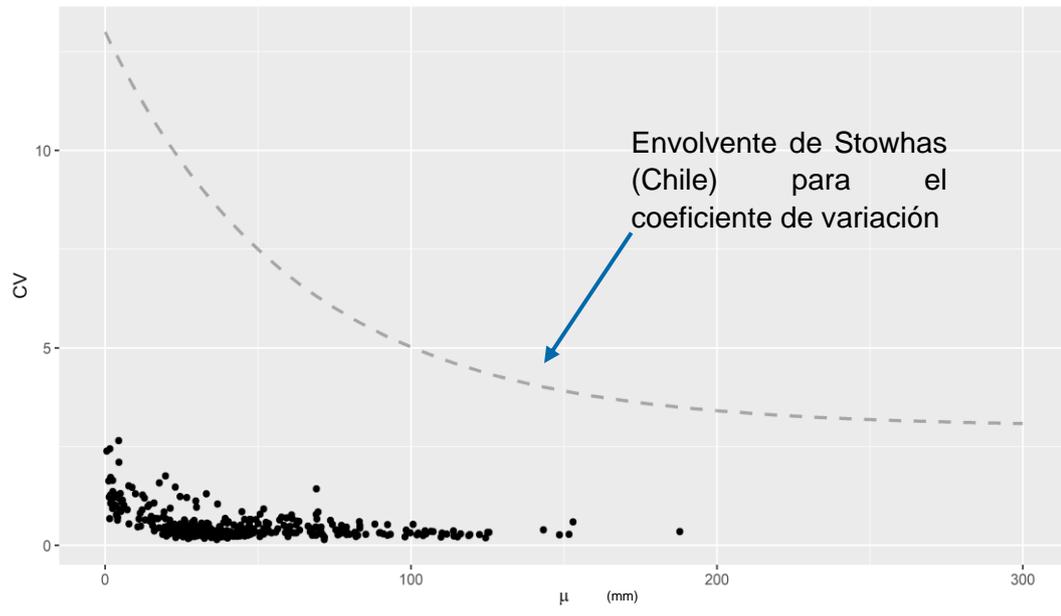
A aquellas estaciones que se identificaron como sospechosas, mostradas en la Tabla 2, se les aplicó la prueba de Grubbs con significancia de 0.5% repetidas veces para eliminar todos los datos fuera de escala física creíble y se volvieron a elaborar las mismas gráficas mencionadas para verificar que se hallan eliminado las inconsistencias.

Figura 18 Cv versus μ de máximos anuales para 24 horas, antes de análisis de outliers



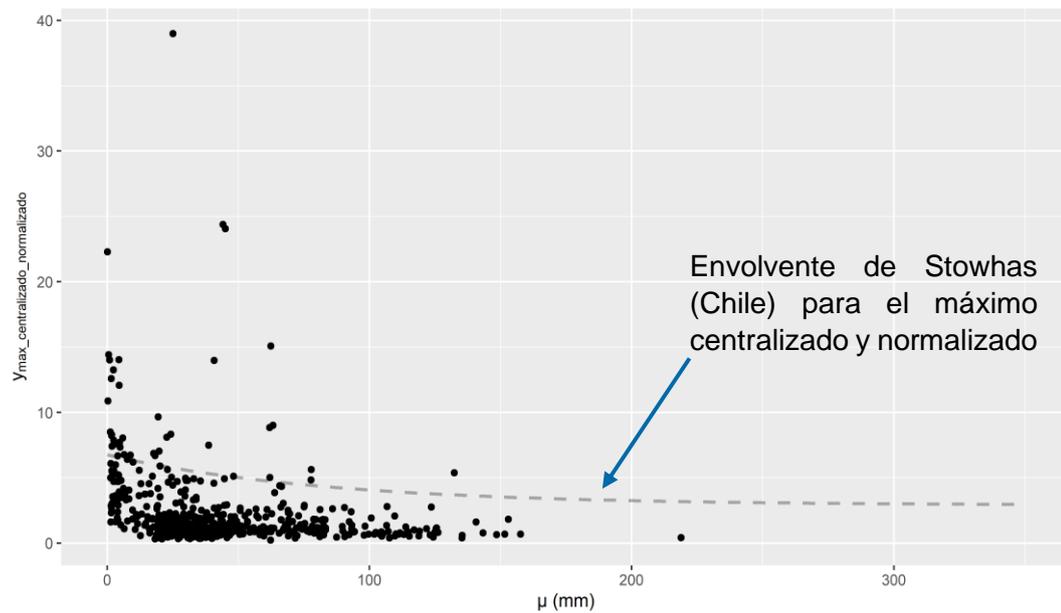
Fuente: Elaboración propia.

Figura 19 Cv versus μ de máximos anuales para 24 horas, luego de análisis de outliers



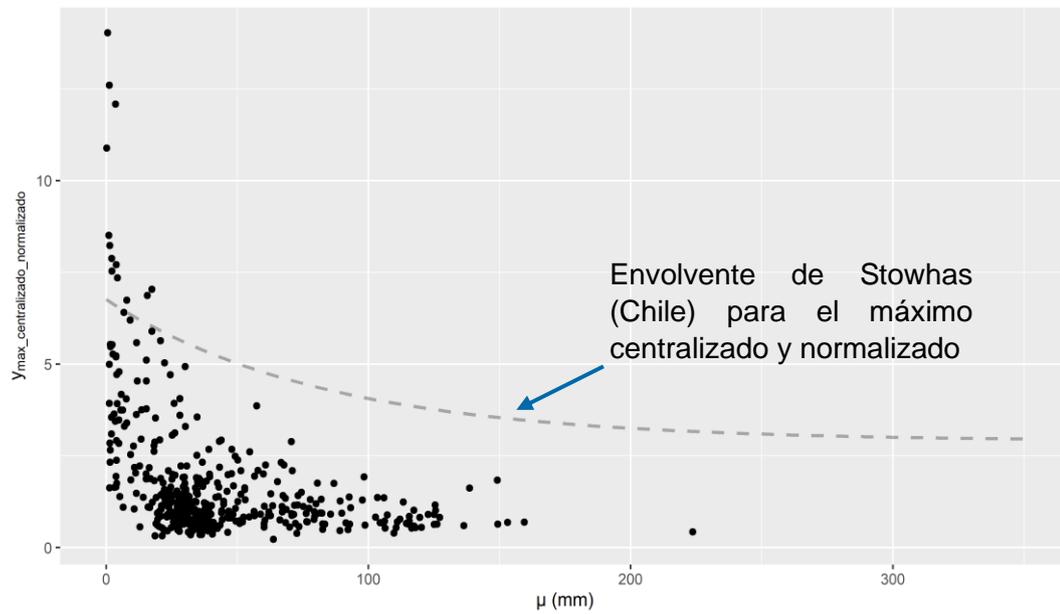
Fuente: Elaboración propia.

Figura 20 $y_{max\ cyn}$ versus μ de máximos anuales para 24 horas, antes de análisis de outliers



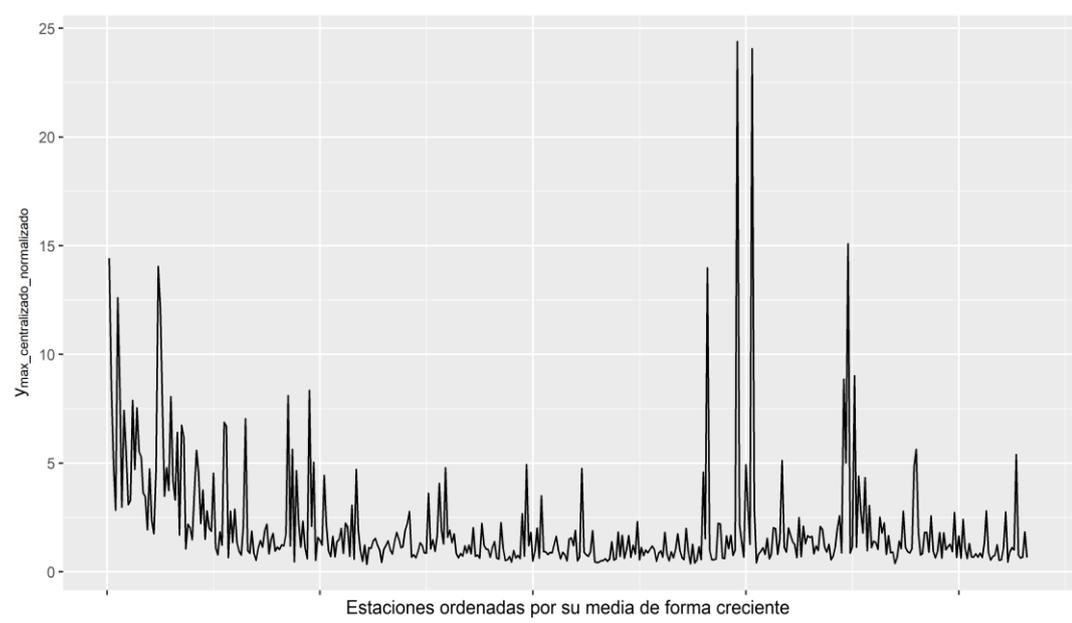
Fuente: Elaboración propia.

Figura 21 $y_{max\ cyn}$ versus μ de máximos anuales para 24 horas, luego de análisis de outliers



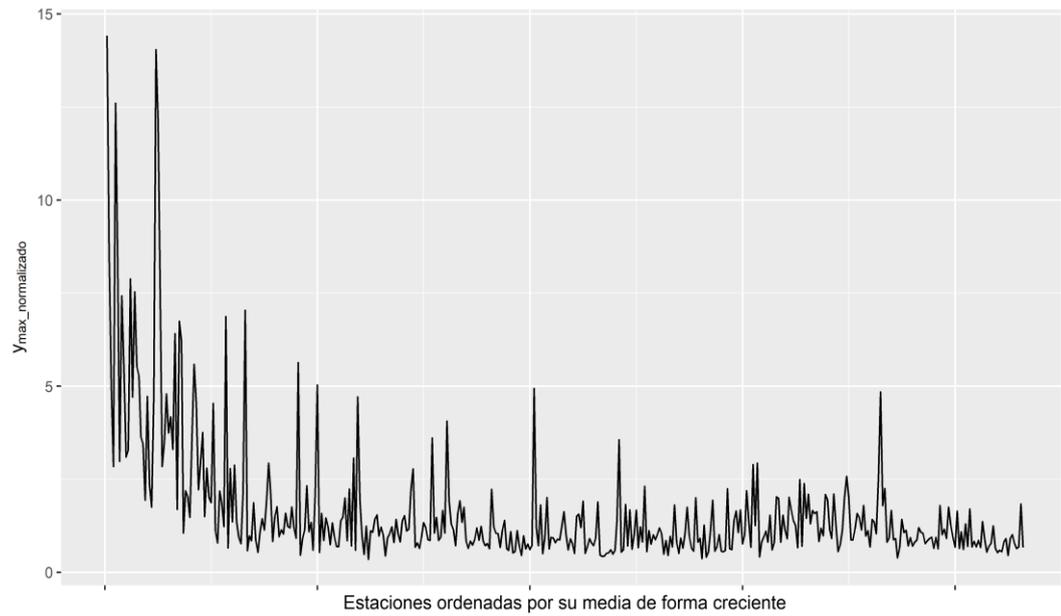
Fuente: Elaboración propia.

Figura 22 $y_{max\ cyn}$ en orden creciente de μ de máximos anuales para 24 horas, antes de análisis de outliers



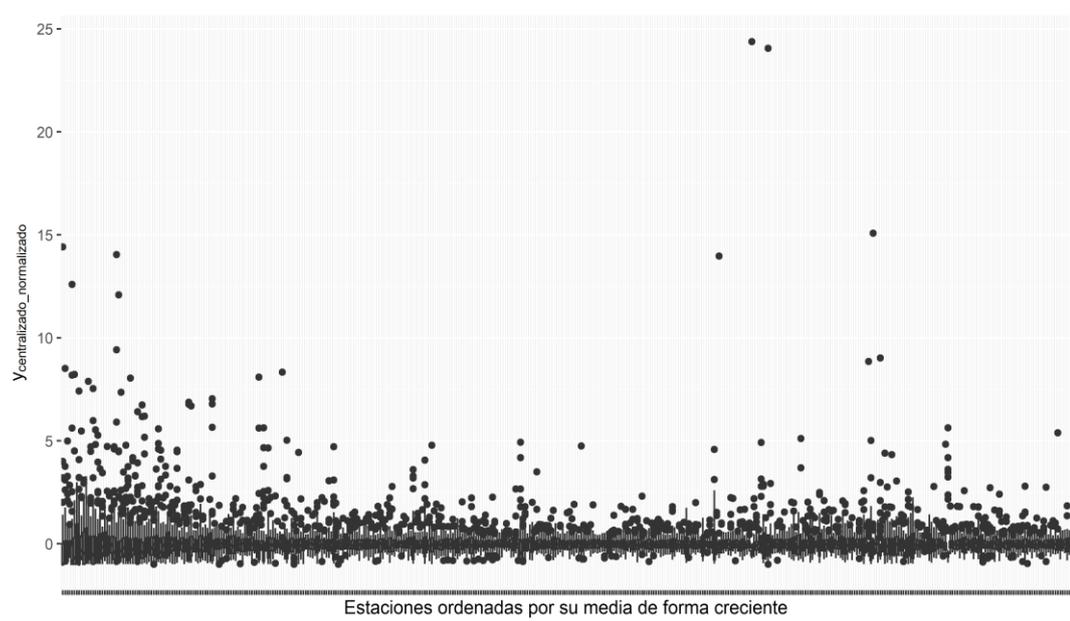
Fuente: Elaboración propia.

Figura 23 $y_{max\ cyn}$ en orden creciente de μ de máximos anuales para 24 horas, luego de análisis de outliers



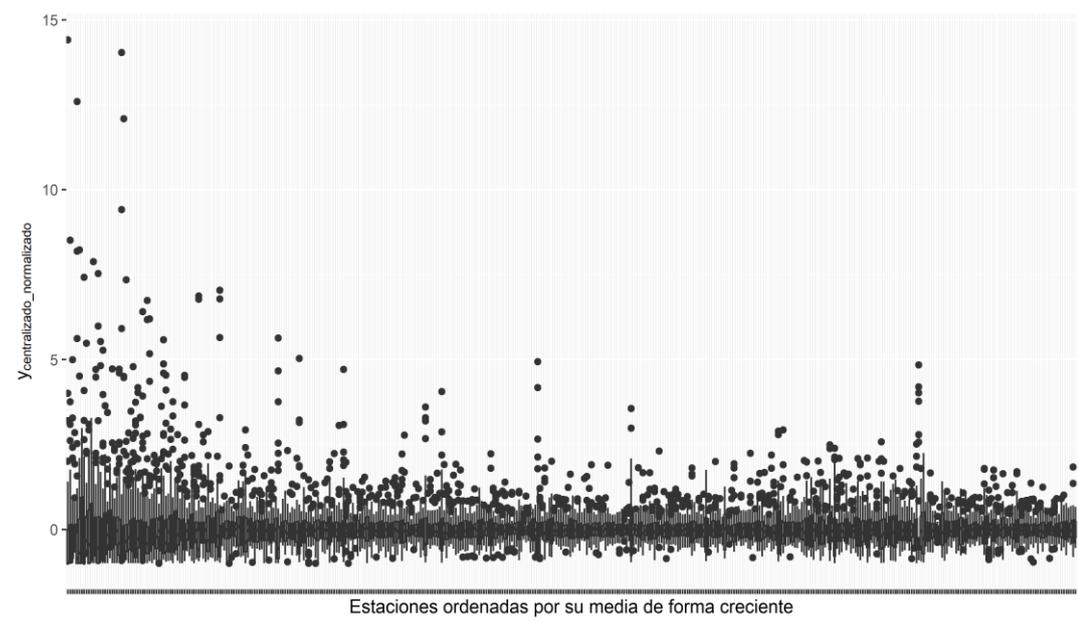
Fuente: Elaboración propia.

Figura 24 Boxplots de $y_{max\ cyn}$ en orden creciente de μ de máximos anuales para 24 horas, antes de análisis de outliers



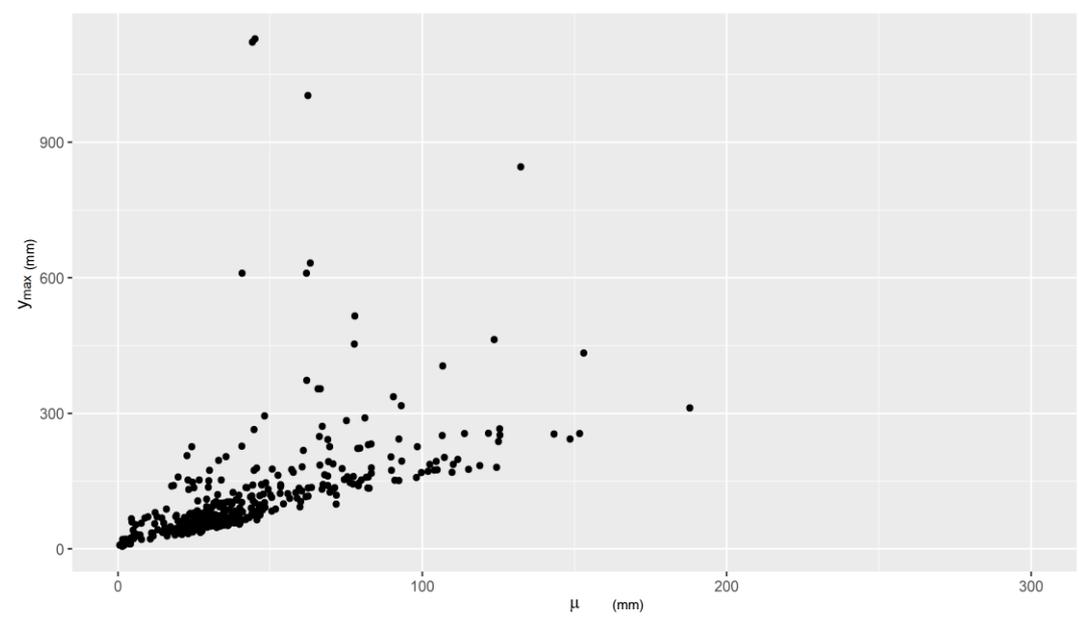
Fuente: Elaboración propia.

Figura 25 *Boxplots de $y_{max\ cyn}$ en orden creciente de μ de máximos anuales para 24 horas, luego de análisis de outliers*



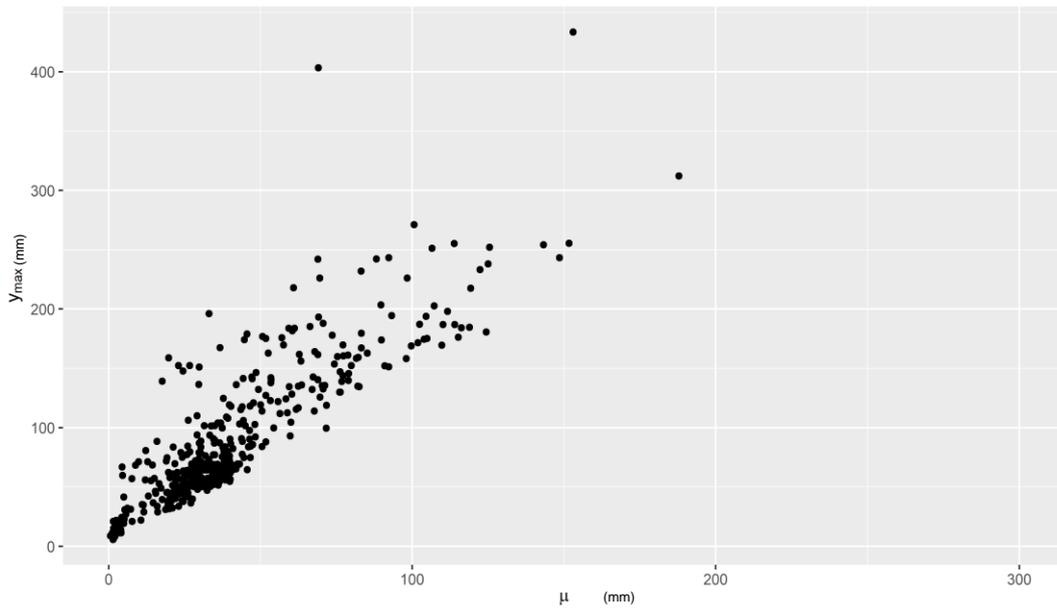
Fuente: Elaboración propia.

Figura 26 *Máximo versus μ de máximos anuales para 24 horas, antes de análisis de outliers*



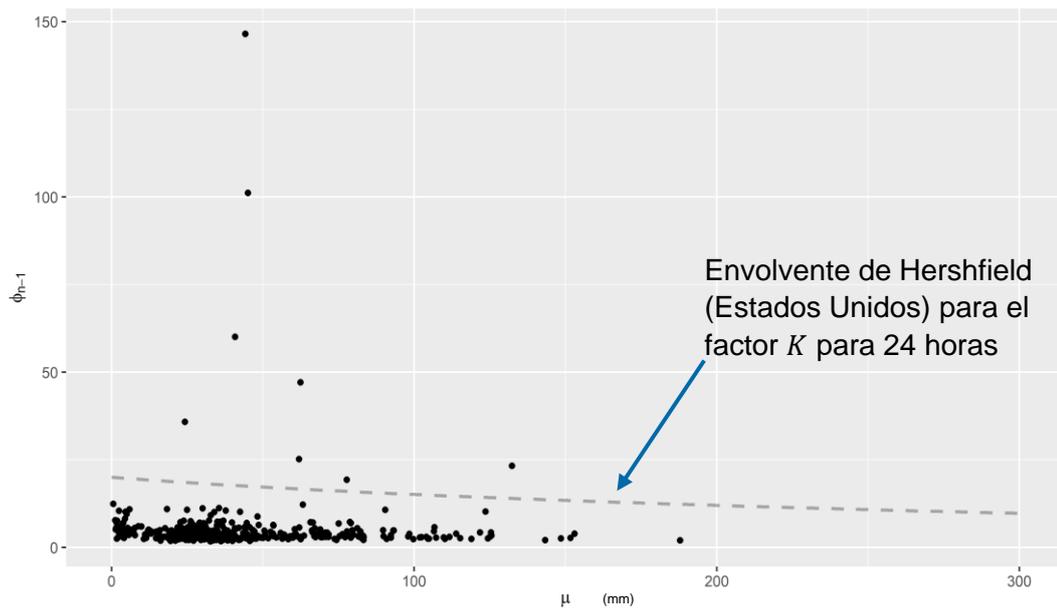
Fuente: Elaboración propia.

Figura 27 *Máximo versus μ de máximos anuales para 24 horas, luego de análisis de outliers*



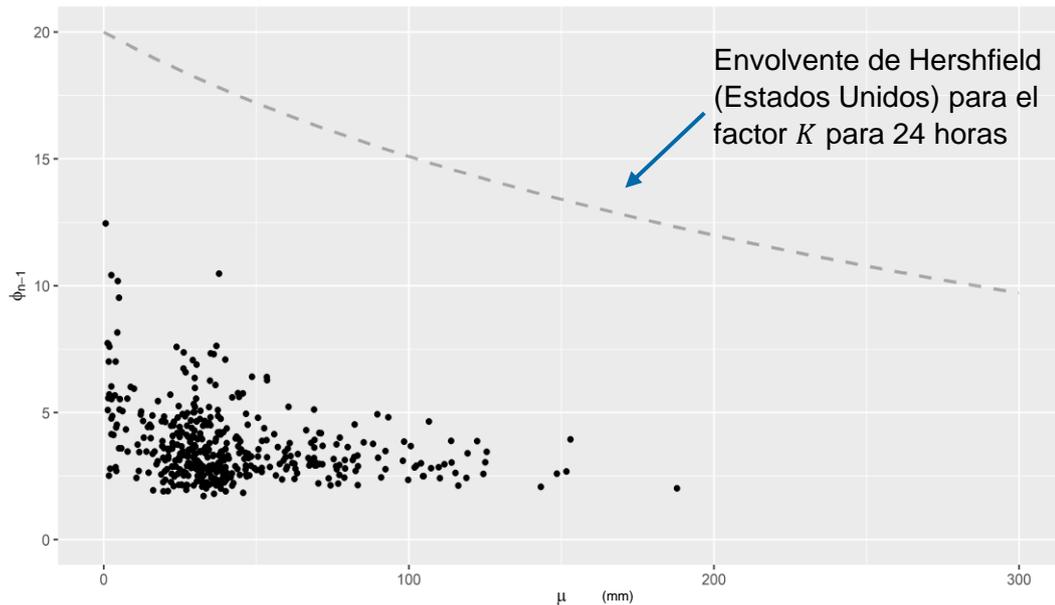
Fuente: Elaboración propia.

Figura 28 *$K_{muestra}$ versus μ de máximos anuales para 24 horas, antes de análisis de outliers*



Fuente: Elaboración propia.

Figura 29 $K_{muestral}$ versus μ de máximos anuales para 24 horas, luego de análisis de outliers



Fuente: Elaboración propia.

El análisis de *outliers* que se hizo mediante el análisis de las 12 gráficas mostradas anteriormente (desde la Figura 18 hasta la Figura 29) demostró que en la data recopilada había muchas estaciones que presentaban datos fuera de escala física creíble, los casos de sospecha encontrados los podemos clasificar en:

- Caso 1: Estaciones con máximos muy elevados;
- Caso 2: Estaciones con un valor de $K_{muestral}$ muy elevado;
- Caso 3: Estaciones con máximos muy elevados en comparación con su media.

En la Tabla 2 se muestran las 68 estaciones sospechosas que se observaron.

Tabla 2 Estaciones sospechosas de tener datos que fueron errores de registro o fuera de escala física creíble

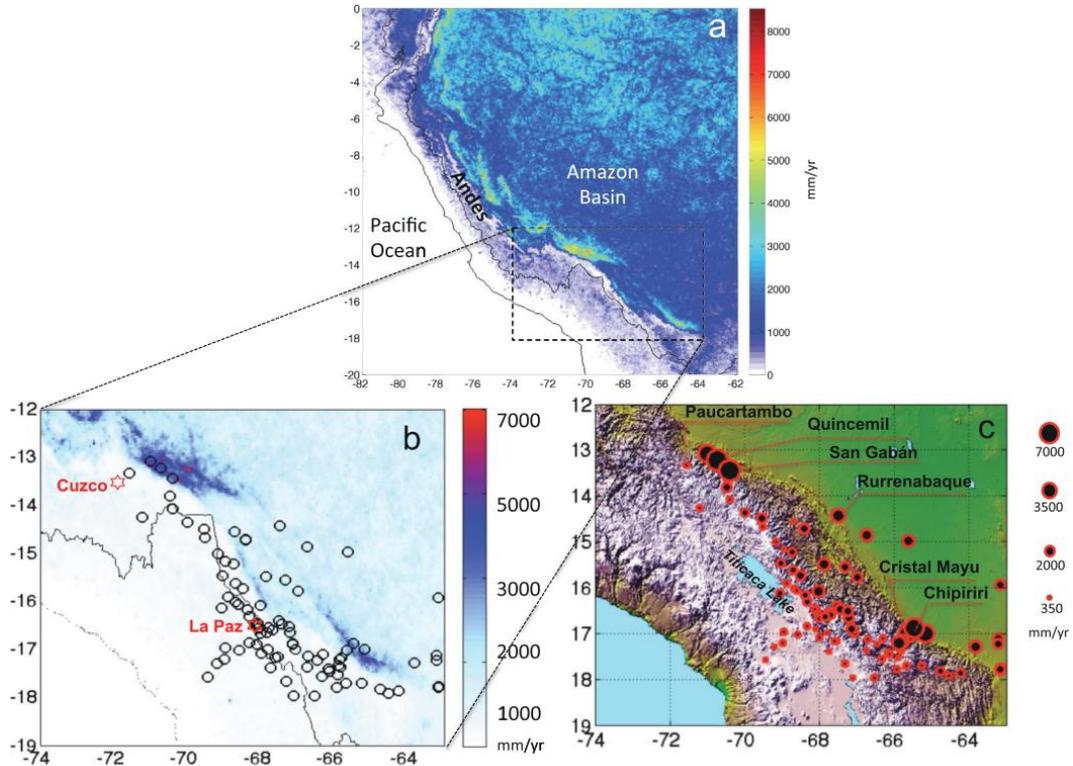
Código ANA	Nombre de la estación	Caso de sospecha	Código ANA	Nombre de la estación	Caso de sospecha
876	Candarave	1	158317	Susapaya	2
154108	Cajamarquilla	1	157312	Chachas	2
241	El Limón	1	157311	Orcopampa	2
281	Genaro Herrera	1	158309	Pampa Umalzo (Titijones)	2
552	San Rafael	1	994	Pomacocha	2

Código ANA	Nombre de la estación	Caso de sospecha	Código ANA	Nombre de la estación	Caso de sospecha
157326	Pullhuay Ayahuasi	1	150003	Laguna Ramón	2
153110	Udima	1	153223	Cachachi	3
6205	San Pedro de Cachi	1	4450	Dos de Mayo	3
180	Pebas	1	152147	Barrios	3
152156	Ania Cabuyal	1	152111	Hacienda Bigote	3
152401	Santa Rita de Castilla	1	172	Tamshiyacu	3
152107	Virrey	1	594	San Juan de Jarpa	3
135	El Salto	1	3317	Pongo de Caynarachi	3
134	Papayal	1	153315	Shanusi	3
571	Satipo	1	151100	El Tigre	3
152140	Pico de Loro	1	3304	Campanilla	3
6401	San Gaban	1	130	Rica Playa	3
853	Locumba	1	478	Puerto Inca	3
155446	Casapalca	1	234	La Tina	3
155204	Pachangara	2	390	Juancito	3
157325	Tisco	2	208	Mallares	3
154114	Dildibuyo	2	152103	Lancones	3
745	Pauza	2	152409	Santa María de Nanay	3
839	La Pampilla	2	153319	Chazuta	3
873	Carumas	2	153311	Cuñumbuque	3
158205	Socabaya	2	299	Pucala	3
795	Cabanaconde	2	310	El Porvenir	3
549	Marcapomacocha	2	321	Namora	3
156110	Huangascar	2	331	Ferreñafe	3
731	Palpa	2	303	Chota	3
156101	San Damian	2	150112	Matapalo	3
157310	Andahua	2	150206	Santa Cruz	3
749	Cotahuasi	2	150901	Malvas	3
155200	Paccho	2	152123	Huar Huar Salala	3
216	Partidor	3			

Fuente: Elaboración propia.

Se consideraron como excepciones a las estaciones Quincemil (código ANA: 6401), San Gabán (código ANA: 693) y Aguaytía (código ANA: 462), ya que estas son estaciones han sido bien controladas y que han registrado las mayores lluvias en el país, tal que son consideradas como *hotspots* de lluvia en el Perú (Chavez & Takahashi, 2017; Espinoza et al., 2015), como se puede apreciar en la Figura 30. Dado ello se consideró crucial su inclusión en el análisis.

Figura 30 (a) Precipitación media anual (1998–2012) estimada con TRMM Pa^a 2A25. (b) Estaciones meteorológicas y precipitación media anual. (c) Precipitación media anual por estación (1975–2005).



Fuente: Espinoza et al. (2015)

Nota: La topografía se representa usando SRTM a una resolución de 1 km.

Los valores del K muestral obtenidos previo a realizarse el análisis de *outliers* eran bastante altos en algunas estaciones pluviométricas debido a que las series de datos de dichas estaciones contenían como dato máximo un registro de lluvia muy elevado (presuntamente un dato fuera de escala física creíble), tras pasar por la prueba de Grubbs con una significancia de 0.5% repetidas veces, las series de datos de las estaciones sospechosas se logró identificar y remover los datos incongruentes de las series.

Posterior al descarte de datos atípicos se filtraron aquellas estaciones que tuvieran al menos 30 años de información, ya que para calcular la PMP con el método de Hershfield se necesita un registro amplio de datos en cada estación pluviométrica que forma parte del análisis para definir bien la media y desviación estándar, buscando que los sucesos máximos atípicos no sean distorsionantes de la

muestra, del conjunto de estaciones con el que se estaba trabajando, quedaron 634 estaciones para las duraciones de 24; 48 y 72 horas.

3.5.3 Análisis de estacionariedad o tendencia

En el análisis de estacionariedad o de tendencia, se aplicó a las series de máximas el comando `zyp.yuepilon` del lenguaje R, el cual calcula en un vector de datos una serie pre-blanqueada sin alterar la tendencia presente en la serie mediante el método Trend Free Pre-Whitening (TFPW, por sus siglas en inglés), eliminando de la serie original autocorrelaciones no deseadas de datos. De forma general se realiza el procedimiento mencionado antes de un análisis de tendencia. La función mencionada pertenece al paquete `zyp` de R, el cual permite utilizar el método de Zhang o el método de Yue Pilon para calcular tendencias preblanqueadas no lineales. Para la presente tesis se utilizó el método de Yue Pilon.

Se identificaron estaciones con tendencia revisando el valor P de Kendall calculado para la serie temporal sin tendencia, si este valor es mayor o igual a 0.05 se puede afirmar que efectivamente la serie temporal no presenta tendencia.

Para el caso del presente análisis se decidió tomar como umbral mínimo un valor P de Kendall mayor o igual a 0.01 (lo cual corresponde a una significancia de 1%) para dar una mayor tolerancia a que la serie pueda incluir datos atípicos, que es lo que se busca.

Luego de analizar las series de datos de las estaciones las siguientes cantidades de estaciones no cumplieron los requisitos planteados para el presente análisis por tener series de datos con tendencia:

- 77 estaciones para el caso de 24 horas;
- 72 estaciones para el caso de 48 horas;
- 71 estaciones para el caso de 72 horas.

3.5.4 Análisis de homogeneidad

Para el análisis de homogeneidad, se aplicó a las series de máximas el comando `pettitt.test`, el cual realiza una prueba no paramétrica Pettitt para probar un cambio en la tendencia central de una serie de tiempo. La hipótesis nula H_0 , sin cambio, se contrasta con la hipótesis H_a , cambio. Estas hipótesis se pueden contrastar luego de analizar el valor P de Pettitt, el cual nos indica que la serie es homogénea si su valor es mayor o igual a 0.05.

Para el caso del presente análisis se decidió tomar como umbral mínimo un valor P de Pettitt mayor o igual a 0.01 (lo cual corresponde a una significancia de 1%) para dar una mayor tolerancia a que la serie pueda incluir datos atípicos, que es lo que se busca.

Luego de analizar las series de datos de las estaciones las siguientes cantidades de estaciones no cumplieron los requisitos planteados para el presente análisis por tener series de datos no homogéneas:

- 104 estaciones para el caso de 24 horas;
- 106 estaciones para el caso de 48 horas;
- 107 estaciones para el caso de 72 horas.

3.5.5 Análisis de independencia serial

Para el análisis de independencia serial, aplicó a las series de máximas el comando LjungBoxTest. Esta prueba estadística determina si un grupo cualquiera de autocorrelaciones de una serie de tiempo son diferentes de cero. El comando genera un vector de los 10 valores P de jungBox, se determina que una estación no presenta autocorrelación si menos de 3 estos valores P son menores que 0.01 (lo cual corresponde a una significancia de 1%).

Luego de analizar las series de datos de las estaciones las siguientes cantidades de estaciones no cumplieron los requisitos planteados para el presente análisis por tener series de presentaban autocorrelación:

- 57 estaciones para el caso de 24 horas;
- 57 estaciones para el caso de 48 horas;
- 69 estaciones para el caso de 72 horas.

3.5.6 Resultados de los análisis de consistencia

Luego de analizar las series de datos de las estaciones se compiló en la Tabla 3 las cantidades de estaciones que cumplieron los requisitos de los análisis planteados segregadas por cantidad de años de información.

Tabla 3 Cantidad de estaciones que cumplen con los análisis de consistencia

Duraciones de estudio	26 - 39 años de información	40 - 49 años de información	50 < años de información	Total
24 horas	129	109	246	484
48 horas	124	111	249	484

Duraciones de estudio	26 - 39 años de información	40 - 49 años de información	50 < años de información	Total
72 horas	132	110	236	478

Fuente: Elaboración propia.

Como parte final del presente capítulo, se aplicaron los factores de corrección de la media y la desviación estándar estimados a partir de la Figura 2, Figura 3 y Figura 4.

Capítulo IV: Clasificación regional y determinación de envolventes

Dada la amplia variedad climática y de mecanismos de producción de precipitaciones máximas, al igual que otras investigaciones previas sobre precipitaciones máximas en el Perú tales como “Estudio de la Hidrología del Perú” (Instituto Italo - Latino Americano et al., 1982), “Regionalización de Precipitaciones Máximas en el Perú” (SENAMHI, 2017), “Desarrollo de curvas pluviométricas Intensidad-Duración-Frecuencia (IDF) en Perú” (Montesinos et al., 2023) se buscaron relaciones válidas de alcances regionales entre las establecidas en dichos estudios y las presentadas en la presente tesis.

El *clustering* (agrupamiento) de estaciones se efectuó en base a estadísticas de las series de máximas anuales (media, máximo, coeficiente de variación) e información de la ubicación espacial de las estaciones (latitud, longitud y altitud), para determinar el número ideal de *clusters*, o regiones, y las estaciones pluviométricas que los conforman.

A continuación, se llevaron a cabo análisis exploratorios para la determinación de la envolvente del factor K de cada *cluster*, los análisis mencionados se enfocaron principalmente en graficar envolventes del factor K de otros estudios realizados en distintas partes del mundo en conjunto con los datos peruanos con el propósito de observar si su comportamiento era similar, de lo cual se pudo determinar que no lo es del todo.

Además, se llevó a cabo una evaluación similar a la realizada por Stowhas (1983) en su estudio de las precipitaciones máximas en Chile que consistió en determinar envolventes para el parámetro K basadas en la amplificación de la envolvente de la precipitación máxima adimensional mediante la variación del coeficiente de variación, ello para verificar si era posible replicar apropiadamente dicha evaluación con los datos peruanos y averiguar si se podrían llegar a conclusiones semejantes.

Finalmente, se determinaron las envolventes del factor K para cada *cluster* y se identificaron envolventes adicionales que podían usarse para un conjunto de *clusters*.

4.1 Clasificación regional de las estaciones pluviométricas por el método de agrupamiento (*clustering*)

El número de *clusters* óptimo se definió aplicando dos métodos. El primero consiste en evaluar 30 índices estadísticos, que individualmente cada uno define el número óptimo de *clusters* de un conjunto de elementos, habiendo diversos índices estadísticos que se basan en distintas variables y características de los datos considerados, tales como las propiedades geométricas o estadísticas de las series de datos, el número de objetos de datos y las mediciones de disimilitud o similitud (Charrad et al., 2014). El segundo método que se aplicó fue el Método del Codo que también señala el número óptimo de *clusters* para un conjunto a partir de la optimización de la WSS (Total Within Sum of Square).

Se inició evaluando los 30 índices estadísticos, los cuales se listan en la Tabla 4, para su evaluación se consideró un número indicador que referencia a la región de lluvia a la que pertenece cada estación, la media, el máximo y el coeficiente de variación de la serie de máximos anuales, así como la latitud, longitud y altitud de cada estación.

Tabla 4 Índices estadísticos empleados para estimar el mejor número de *clusters*

Índice	Índice	Índice
CH	Frey	McClain
Duda	Hartigan	Rubin
Pseudot2	Tau	KL
C	Ratkowsky	Silhouette
Gamma	Scott	Gap
Beale	Marriot	Dindex
CCC	Ball	Dunn
Ptbiserial	Trcovw	Hubert
Gplus	Tracew	SD
DB	Friedman	SDbw

Fuente: Elaboración propia.

Los números óptimos de *clusters* para el conjunto de estaciones de 24 horas (el cual se tomó como base para las otras duraciones) indicados por los 30 índices fueron resumidos y agrupados de acuerdo con la cantidad de *clusters* que estos propusieron:

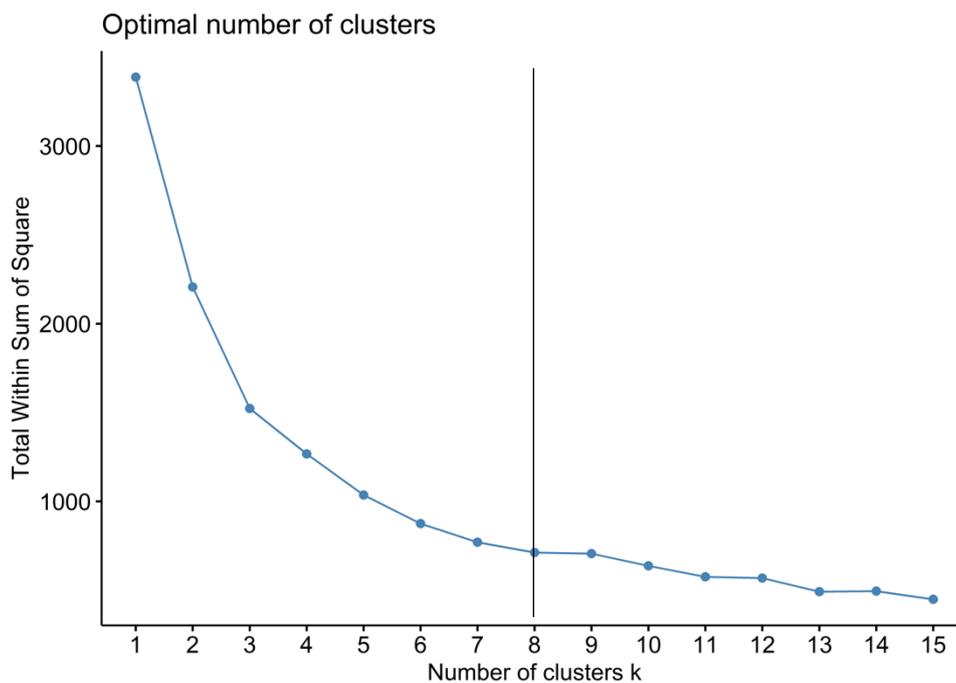
- 7 índices propusieron 5 como el mejor número de *clusters*
- 4 índices propusieron 6 como el mejor número de *clusters*

- 5 índices propusieron 7 como el mejor número de *clusters*
- 1 índice propuso 10 como el mejor número de *clusters*
- 1 índice propuso 12 como el mejor número de *clusters*
- 4 índices propusieron 13 como el mejor número de *clusters*
- 2 índices propusieron 15 como el mejor número de *clusters*

De lo cual, el presente método sugiere que el número óptimo de *clusters* es 5, por ser el valor más frecuente.

Por otro lado, el Método del Codo cuya gráfica es mostrada en la Figura 31, sugiere que el número óptimo de *clusters* es 8, puesto que a partir de ese valor la reducción en la suma total de varianza intra-*cluster* deja de ser sustancial.

Figura 31 Número óptimo de *clusters* según el Método del Codo



Fuente: Elaboración propia.

Con base en los dos métodos se decidió que la opción óptima era considerar 6 *clusters*.

Sin embargo, con base en la experiencia y el análisis de las series anuales de los máximos diarios de las estaciones ubicadas en la zona muy próxima al litoral (menos de 15 km de distancia al mar), se definieron 2 *clusters* adicionales de manera forzada, tomando en cuenta el comportamiento particular de las precipitaciones máximas en la zona mencionada, el cual corresponde a un muy

bajo promedio del máximo diario y un alto coeficiente de variación, donde las series de máximos diarios muestran varios años sin precipitación alguna o con una máxima diaria menor a 5 o 10 mm, y esporádica o muy esporádicamente valores altos. Los *clusters*, aunque difieren entre sí, son muy sensiblemente afectados por el fenómeno El Niño y la corriente de Humboldt, dado que ambas producen precipitaciones atípicamente altas o no, comportamiento que es complicado de advertirse por interpolación espacial y análisis de agrupamientos.

El primero de los *clusters* adicionales intuitivamente abarca desde el litoral de Piura hasta el litoral de Ancash, mientras que el segundo abarca desde el litoral de Lima hasta el litoral de Tacna, resultando así, en un total de 8 *clusters*.

En la Tabla 5 se listan las estaciones que pertenecen a los *Clusters* 7 y 8, adicionalmente en la Figura 32 se pueden apreciar ubicaciones de las estaciones que conforman estos *clusters* adicionales.

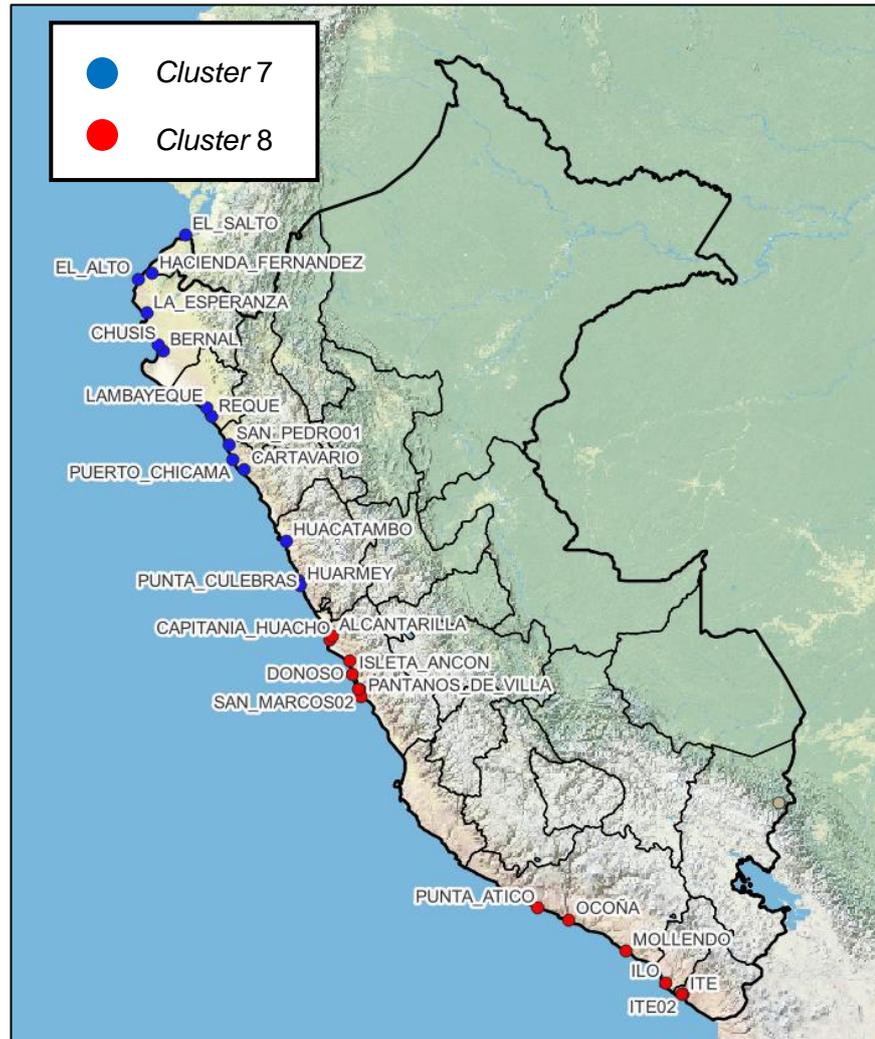
Tabla 5 Estaciones que conforman los *Clusters* 7 y 8

Código ANA	Nombre de la estación	Cluster
000413	Punta Culebras	7
000530	Huarmey	7
000418	Huacatambo	7
000332	Reque	7
000301	Lambayeque	7
000326	San Pedro	7
152100	Bernal	7
000231	Chusis	7
000230	La Esperanza	7
000345	Cartavio	7
000329	Puerto Chicama	7
000209	El Alto	7
150108	Hacienda Fernandez	7
000135	El Salto	7
000132	Puerto Pizarro	7
000602	Campo de Marte CGFAP	8
006617	Campo de Marte	8
000830	Punta Ático	8
000858	Ite	8
000840	Ilo	8
158300	Ite	8
000800	Mollendo	8
000624	Pantanos de Villa	8
000619	San Marcos	8
000612	Rímac	8
000599	Isleta Ancón	8

Código ANA	Nombre de la estación	Cluster
000546	Donoso	8
000486	Capitanía Donoso	8
000501	Alcantarilla	8
000831	Ocoña	8
000899	La Yarada	8

Fuente: Elaboración propia.

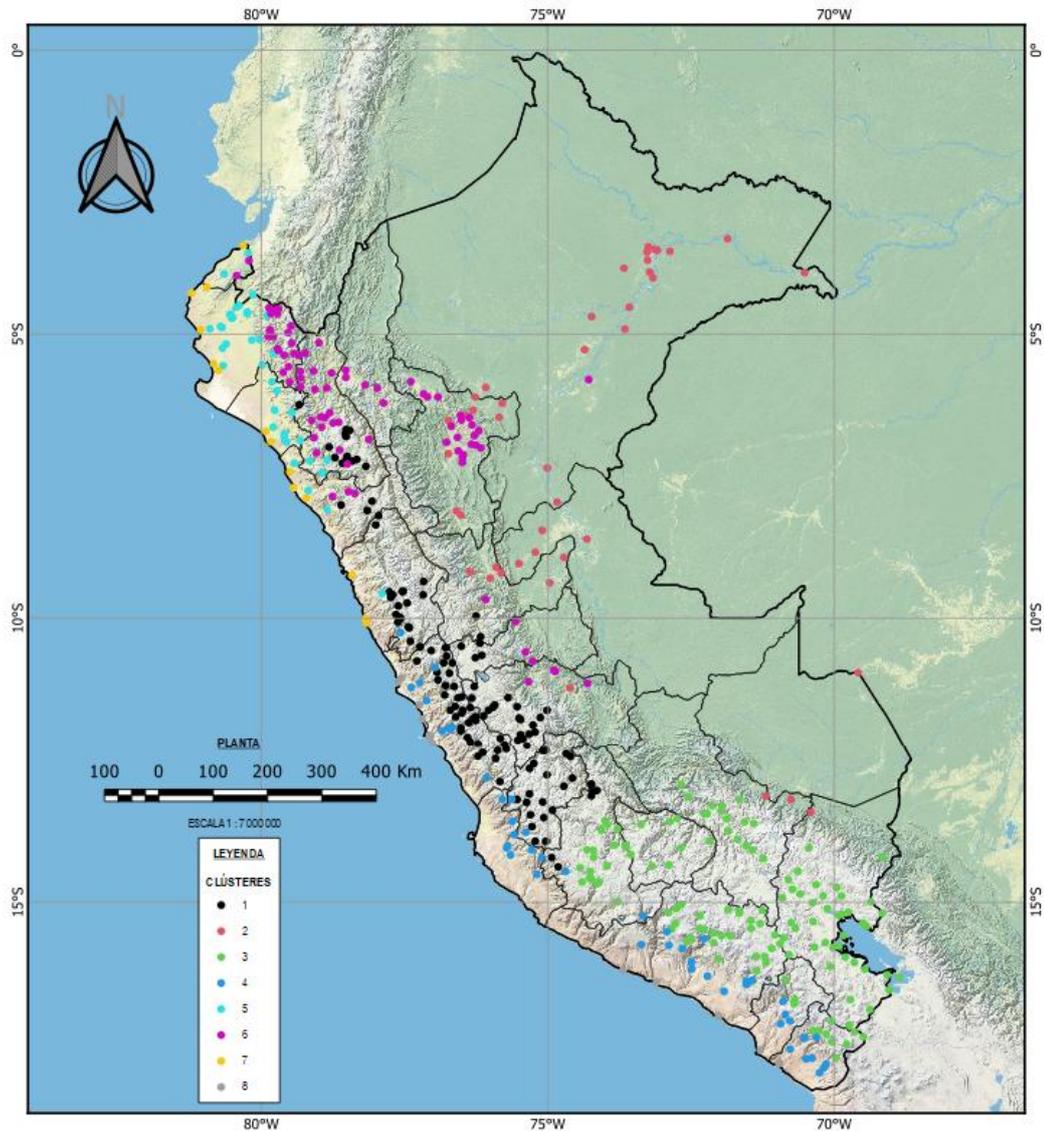
Figura 32 Ubicaciones de las estaciones de los Clusters 7 y 8



Fuente: Elaboración propia.

Definido el número de *clusters* se procedió a generar los agrupamientos con el método de *clustering* mediante mínimos cuadrados, obteniendo las conformaciones de *clusters* correspondientes a las duraciones de estudio, se muestra en la Figura 33 la distribución de estaciones de los 8 *clusters* en el caso de 24 horas, en los Mapas D.1, E.1 y F.1 pueden apreciarse de manera más detallada los *clusters* de cada duración.

Figura 33 Clusters para el caso de 24 horas



Fuente: Elaboración propia.

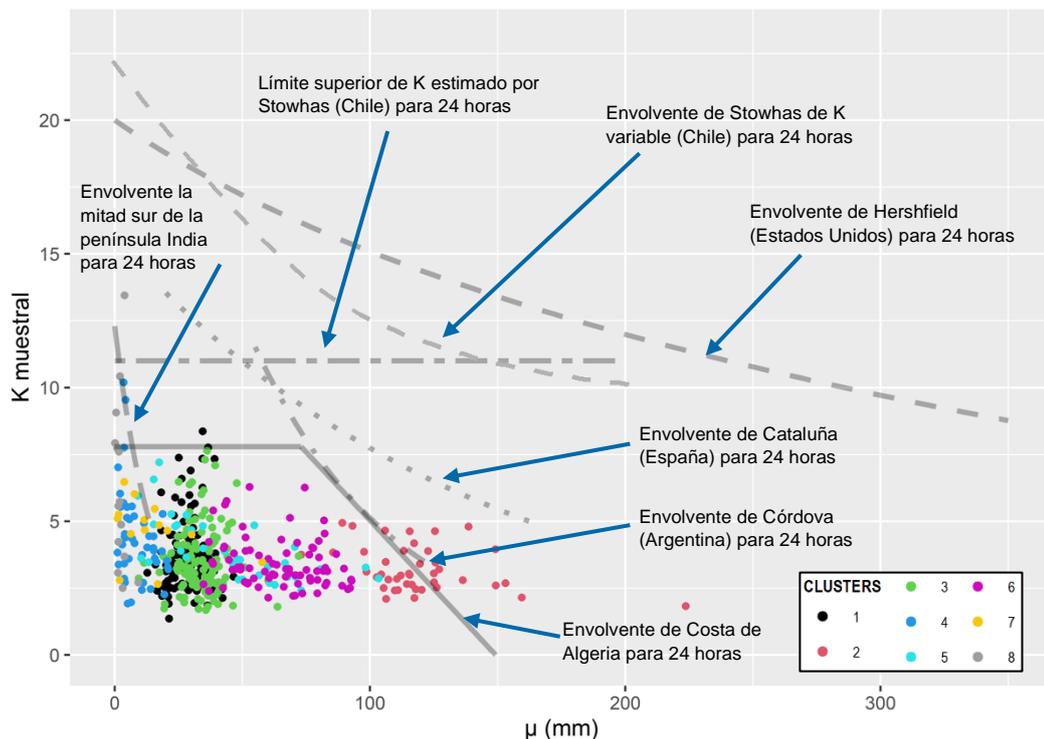
4.2 Análisis exploratorios para la determinación de envolventes

Los análisis exploratorios se realizaron para buscar estructuras relacionales entre el factor K de Hershfield, y la media (μ), el coeficiente de variación (Cv), el valor máximo absoluto (P_{max}) y el valor adimensional de los máximos anuales $\left(\frac{P_{max}-\mu}{\mu}\right)$, que ayuden a determinar la forma de las envolventes del factor K de Hershfield para cada región o grupo de regiones. Los análisis exploratorios se efectuaron principalmente con la variable de máximas anuales en 24 horas porque es mayor

el empleo de análisis con esta variable en los estudios previos, en el caso de 48 y 72 horas se procedió del mismo modo.

El gráfico fundamental del método de Hershfield es la relación entre el factor de maximización K y la media de los máximos anuales en 24 horas (μ). Con los datos de las series de máximas anuales se evaluaron los factores K observados los que se denominaron $K_{muestral}$. Se puede apreciar en la Figura 34 la nube de los $K_{muestral}$ evaluados para el Perú en el plano K vs μ contrastados con la envolvente de Hershfield con pretensiones globales, y otras envolventes particulares desarrolladas para otras partes del mundo: Chile, Córdoba en Argentina, Cataluña en España, mitad sur de la península India y la costa de Algeria (Batout et al., 2022; Casas et al., 2010; Desa M. & Rakhecha, 2007; Dhar et al., 1981; Hershfield, 1961, 1965; Stowhas, 1983).

Figura 34 $K_{muestral}$ versus μ de máximos anuales y envolventes de otros estudios para 24 horas



Fuente: Elaboración propia.

En los estudios anteriores de la PMP que se citaron anteriormente, el factor K llegó como máximo a 20 en el caso de la curva de Hershfield, de la cual se comenta que en muchos casos tiende a sobreestimar el valor de K y en la mayoría de los casos, dicho valor no supera a 15, en el caso de Perú, $K_{muestral}$ varía entre 0 y 15, por lo

que se afirma que los valores se encuentran dentro de un rango coherente a nivel mundial.

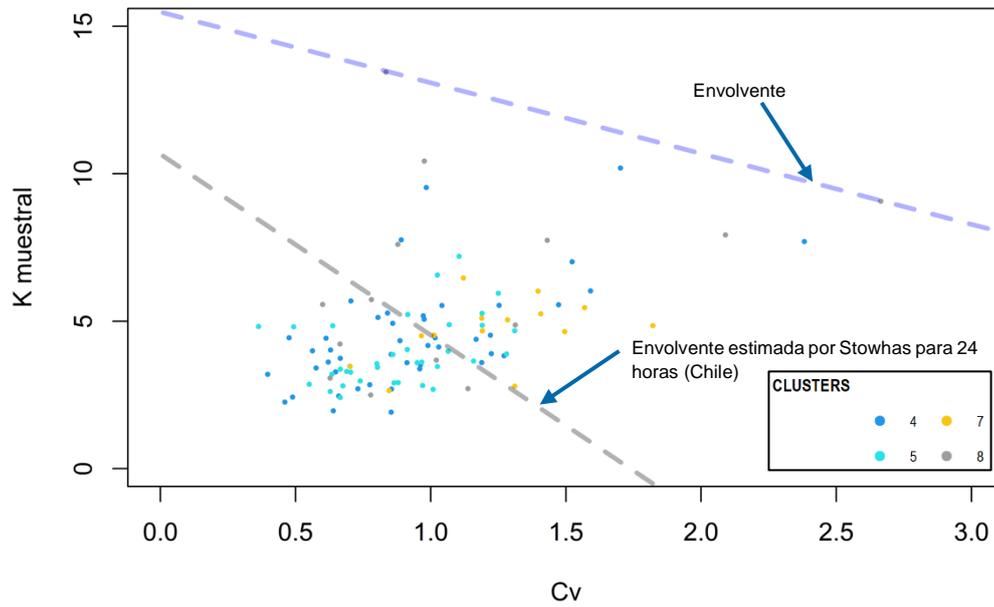
Otra relación importante es la que existe entre el factor K y el coeficiente de variación de los máximos anuales. Tal como realizó Stowhas (1983) en su estudio para Chile, en el caso peruano se buscó establecer una envolvente superior que defina una frontera extrema en la relación entre el $K_{muestral}$ y el coeficiente de variación (C_v).

Para buscar uniformizar las relaciones entre los distintos parámetros se consideró agrupar determinados *clusters* basándose en la ubicación y comportamiento esperado, también se consideró que las estaciones de los *clusters* debían ser espacialmente cercanas y que los *clusters* preferentemente debían estar en la vecindad los uno de los otros, ya que no sería lógico establecer que distintos *clusters* que se encuentren en ubicaciones alejadas o regiones marcadamente distintas compartan una curva envolvente.

En tal sentido, se decidió agrupar los *clusters* 4; 5; 7 y 8 en representación de la costa, los *clusters* 1 y 3 en representación de la sierra central y sur, el *cluster* 2 en representación de la selva y el *cluster* 6 en representación de la sierra norte y parte de la selva.

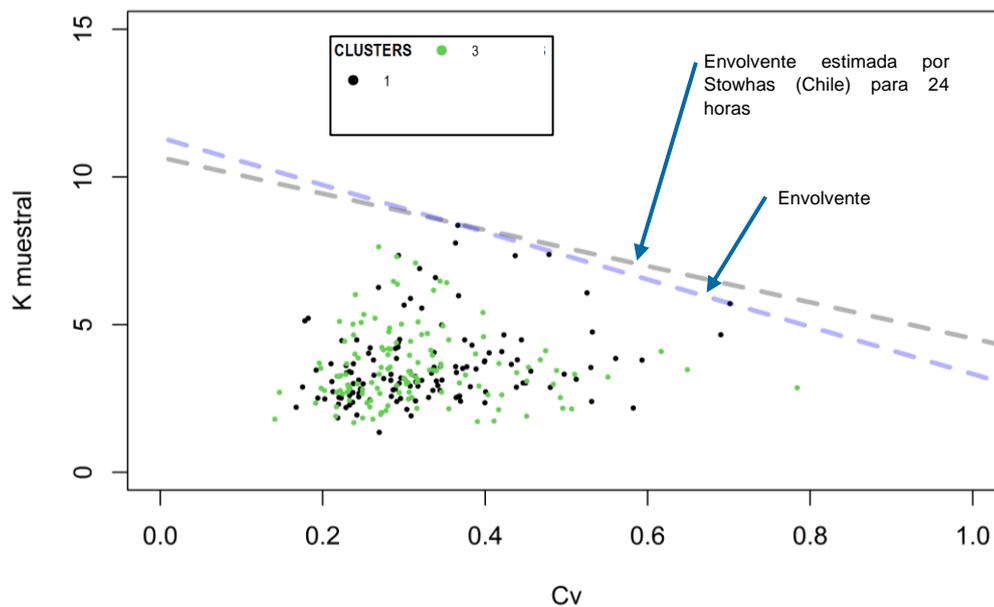
Inicialmente, siguiendo el procedimiento de Stowhas, se estableció una envolvente superior lineal en la gráfica de los parámetros de $K_{muestral}$ y el coeficiente de variación (C_v), en las Figura 35, Figura 36, Figura 37 y Figura 38 se puede apreciar la representación de los parámetros mencionados con las envolventes para los grupos de *clusters* establecidos para la duración de 24 horas.

Figura 35 $K_{muestral}$ versus Cv de los clusters 4; 5; 7 y 8 con su envolvente, para 24 horas



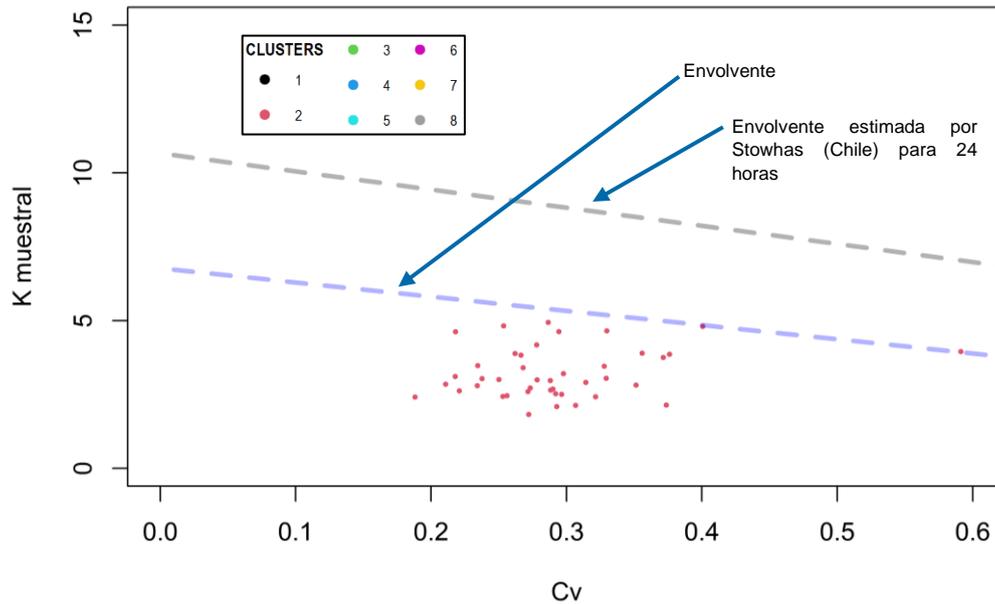
Fuente: Elaboración propia.

Figura 36 $K_{muestral}$ versus Cv de los clusters 1 y 3 con su envolvente, para 24 horas



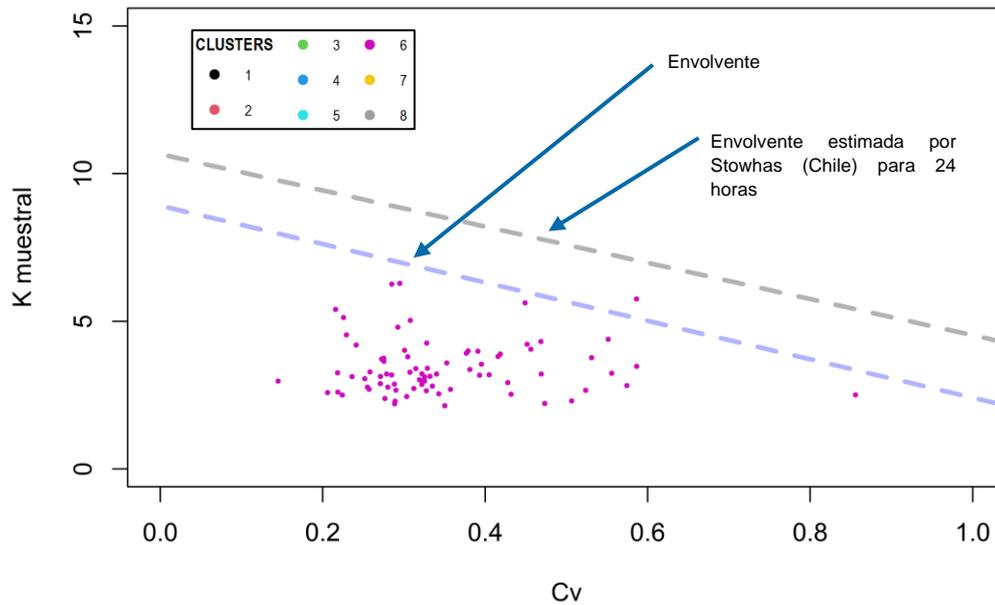
Fuente: Elaboración propia.

Figura 37 $K_{muestral}$ versus Cv del cluster 2 con su envolvente, para 24 horas



Fuente: Elaboración propia.

Figura 38 $K_{muestral}$ versus Cv del cluster 6 con su envolvente, para 24 horas

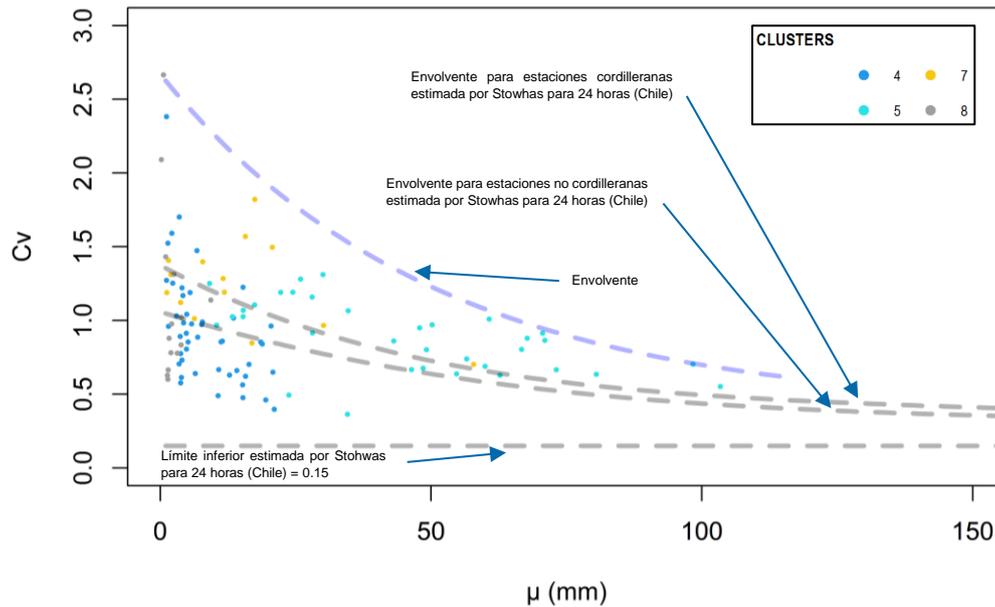


Fuente: Elaboración propia.

Es visible que los *clusters* 1; 2; 3 y 6 se encuentran bajo la envolvente de Stowhas, contrario a lo que se podría esperar (dado que Chile es un país principalmente costero), dichos *clusters* no se encuentran cercanos al litoral peruano, la razón de ello se debe al bajo coeficiente de variación que presentan sus estaciones tal como sucede en Chile.

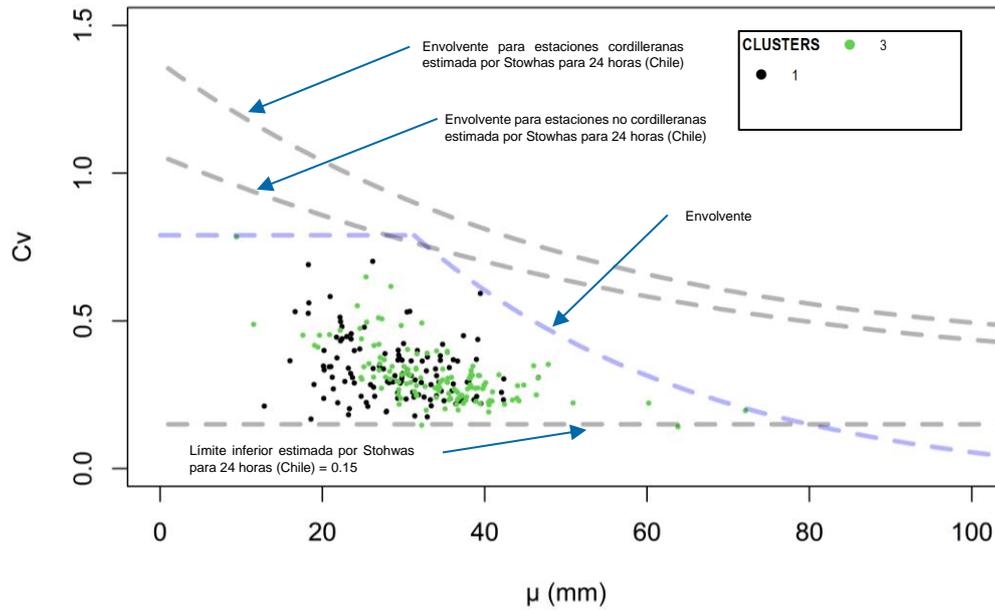
Continuando con el procedimiento seguido por Stowhas, se establecieron curvas envolventes superiores, en las gráficas del coeficiente de variación (C_v) en función de la media (μ), en las Figura 39, Figura 40, Figura 41 y Figura 42 se puede apreciar la representación de los parámetros mencionados con las envolventes para los grupos de *clusters* establecidos para la duración de 24 horas.

Figura 39 C_v versus μ de los clusters 4; 5; 7 y 8 con su envolvente, para 24 horas



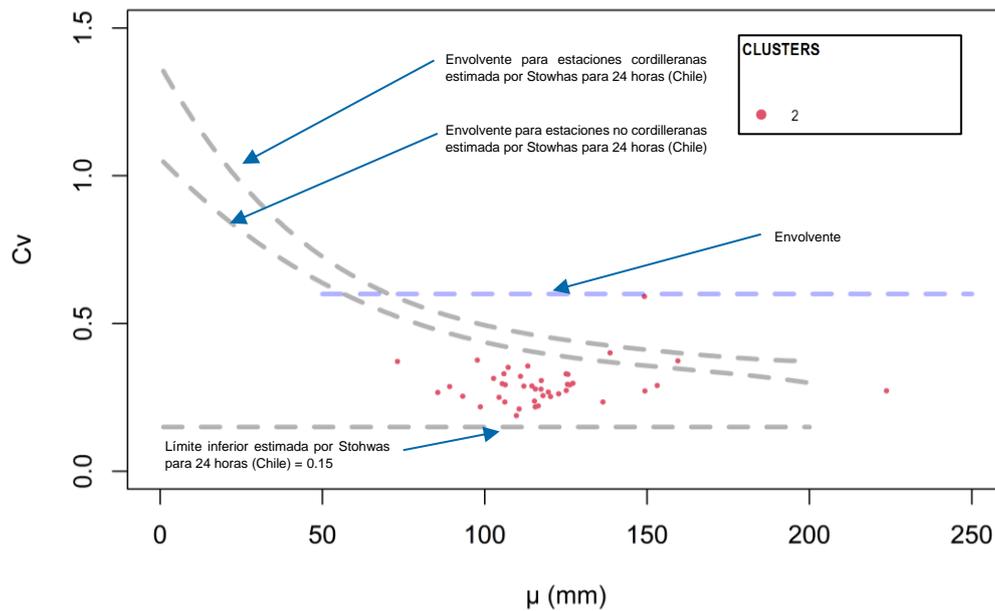
Fuente: Elaboración propia.

Figura 40 C_v versus μ de los clusters 1 y 3 con su envolvente, para 24 horas



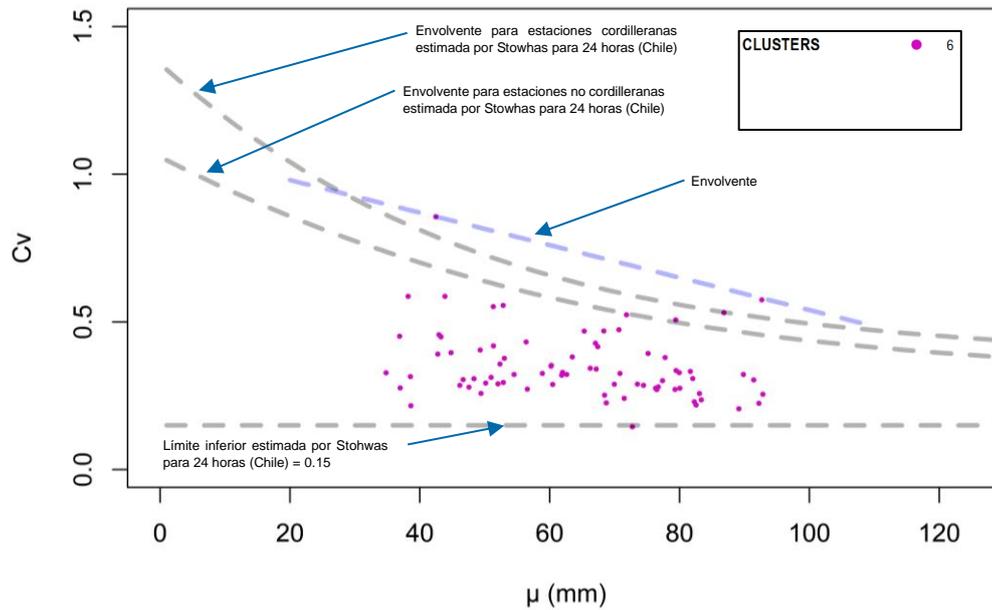
Fuente: Elaboración propia.

Figura 41 C_v versus μ del cluster 2 con su envolvente, para 24 horas



Fuente: Elaboración propia.

Figura 42 C_v versus μ del cluster 6 con su envolvente, para 24 horas



Fuente: Elaboración propia.

De las gráficas mencionadas también se identifica un valor mínimo para el C_v en cada caso, el cual servirá para calcular mediante la envolvente de K versus C_v un nuevo valor de K , el cual correspondería a un valor extremo que podría llegar a generarse en el grupo de *clusters* evaluado. En las Tabla 6, Tabla 7 y Tabla 8 se muestra los valores identificados de C_v , los valores de K calculados mediante la envolvente mencionada y su valor redondeado (siendo conservador).

Tabla 6 Identificación del valor extremo de K para 24 horas

Parámetros para 24 horas	Clusters 4; 5; 7 y 8	Clusters 1 y 3	Clister 2	Clusters 6
C_v mínimo	0,36	0,14	0,18	0,15
K calculado con la envolvente de K versus C_v	14,84	10,93	6,34	8,53
K redondeado	15	11	7	9

Fuente: Elaboración propia.

Tabla 7 Identificación del valor extremo de K para 48 horas

Parámetros para 48 horas	Clusters 4; 5; 7 y 8	Clusters 1 y 3	Clusters 2	Clusters 6
Cv mínimo	0,40	0,17	0,19	0,19
K calculado con la envolvente de K versus Cv	15,66	11,24	8,30	9,31
K redondeado	16	11,5	9	10

Fuente: Elaboración propia.

Tabla 8 Identificación del valor extremo de K para 72 horas

Parámetros para 72 horas	Clusters 4; 5; 7 y 8	Clusters 1 y 3	Clusters 2	Clusters 6
Cv mínimo	0,37	0,18	0,18	0,19
K calculado con la envolvente de K versus Cv	14,54	8,28	9,92	8,66
K redondeado	15	13	10	9

Fuente: Elaboración propia.

Tomando en consideración los valores de K identificados para cada duración y grupo de *clusters*, considerados extremos, se definieron los valores de $K_{max\ probable}$, tomando en cuenta que dichos valores no deberían descender a medida que aumente la duración de análisis, ya que esto podría generar valores incongruentes de PMP, en caso el análisis sugiera que el valor de $K_{max\ probable}$ de un grupo de *clusters* es menor en una duración superior podría deberse a la reducida información analizada, más no ser completamente acertado. Finalmente, en la Tabla 9, se muestran los valores de $K_{max\ probable}$ determinados.

Tabla 9 $K_{max\ probable}$ determinado para 24; 48 y 72 horas

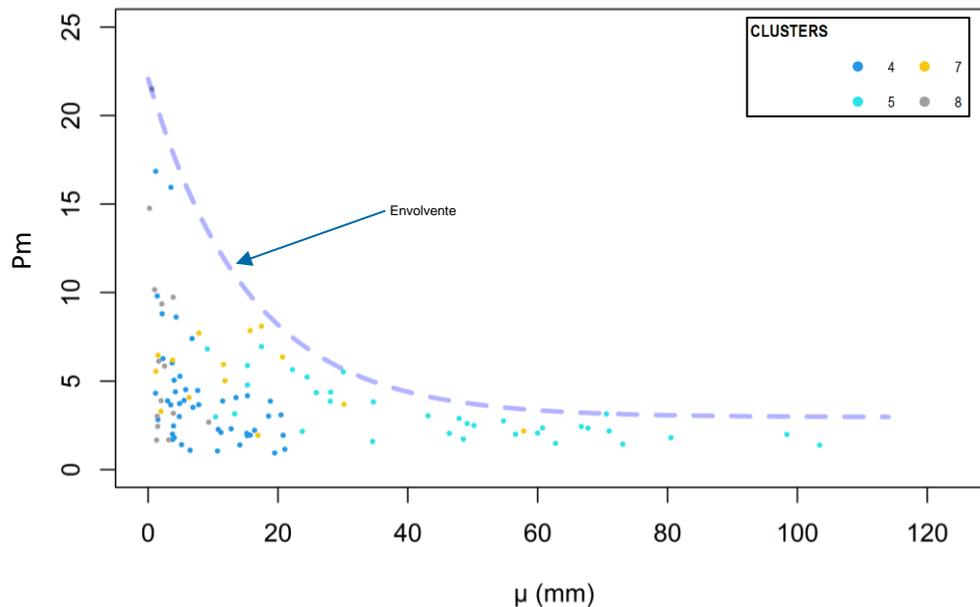
Duración	Clusters 4; 5; 7 y 8	Clusters 1 y 3	Clusters 2	Clusters 6
24 horas	15	11	7	8,5
48 horas	16	11,5	9	10
72 horas	16	13	10	10

Fuente: Elaboración propia.

4.3 Determinación de las envolventes regionales de K

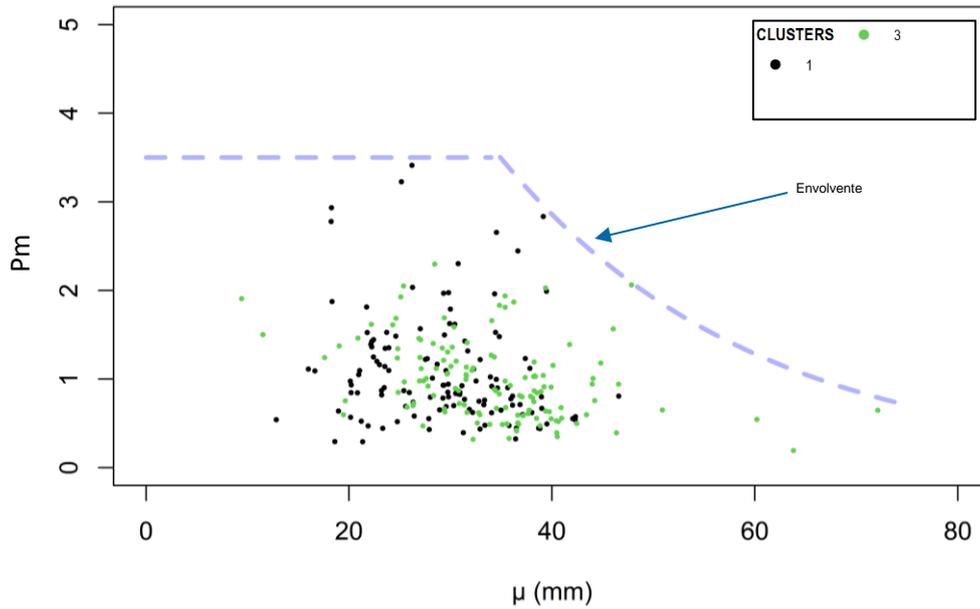
A continuación, de acuerdo con el procedimiento de Stowhas, se establecieron curvas envolventes superiores en las gráficas del máximo adimensional ($P_m = \frac{P_{max} - \mu}{\mu}$) en función de la media (μ), partiendo de la forma de una exponencial negativa (acotada) (Caamaño Nelli & García, 1999), la cual fue considerada dado su buen ajuste a la frontera superior de la nube de puntos de cada una de estas gráficas, en las Figura 43, Figura 44, Figura 45 y Figura 46 se puede apreciar la representación de los parámetros mencionados con las envolventes para los grupos de *clusters* establecidos para el caso de 24 horas.

Figura 43 P_m versus μ de los clusters 4; 5; 7 y 8 con su envolvente, para 24 horas



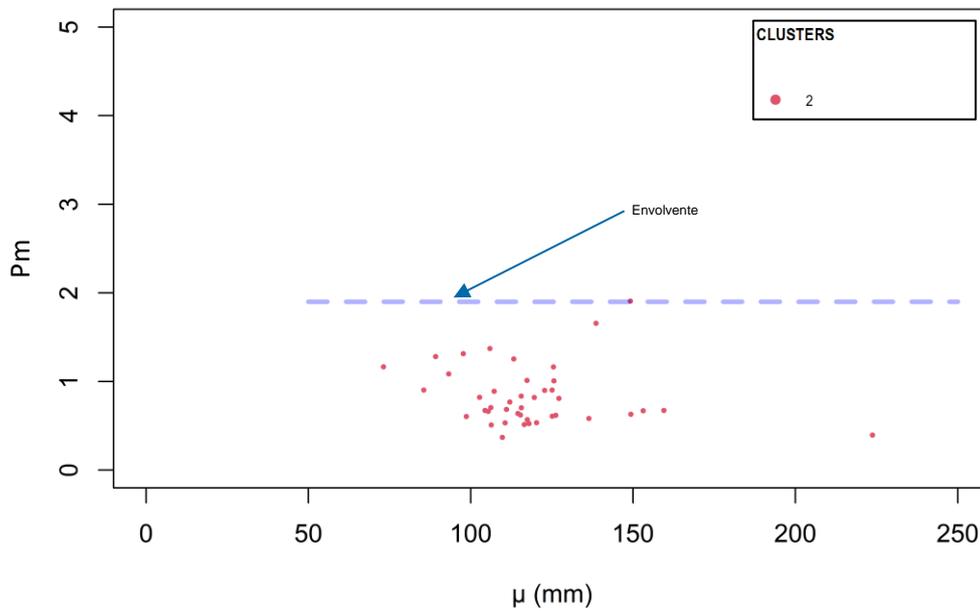
Fuente: Elaboración propia.

Figura 44 P_m versus μ de los clusters 1 y 3 con su envolvente, para 24 horas



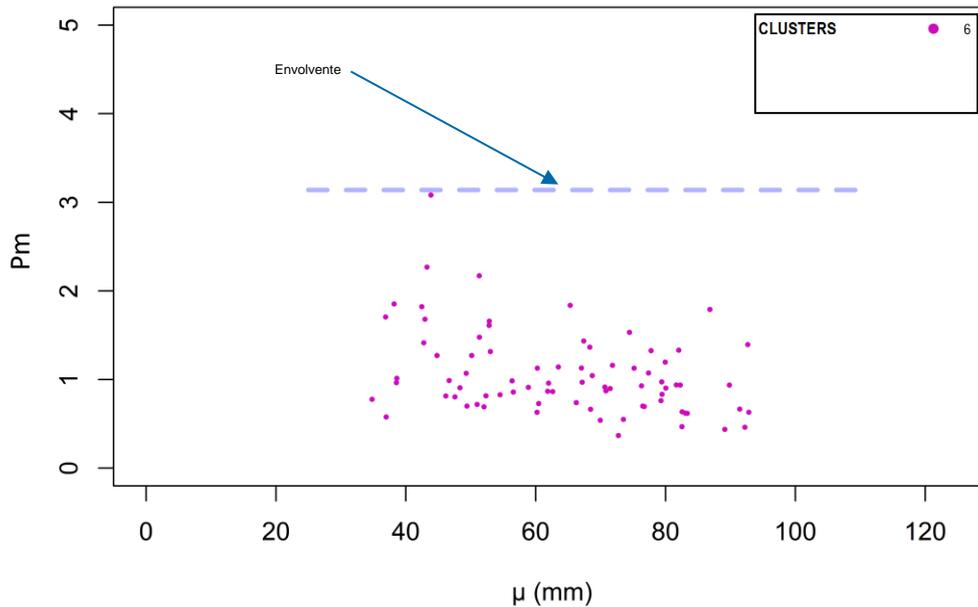
Fuente: Elaboración propia.

Figura 45 P_m versus μ del cluster 2 con su envolvente, para 24 horas



Fuente: Elaboración propia.

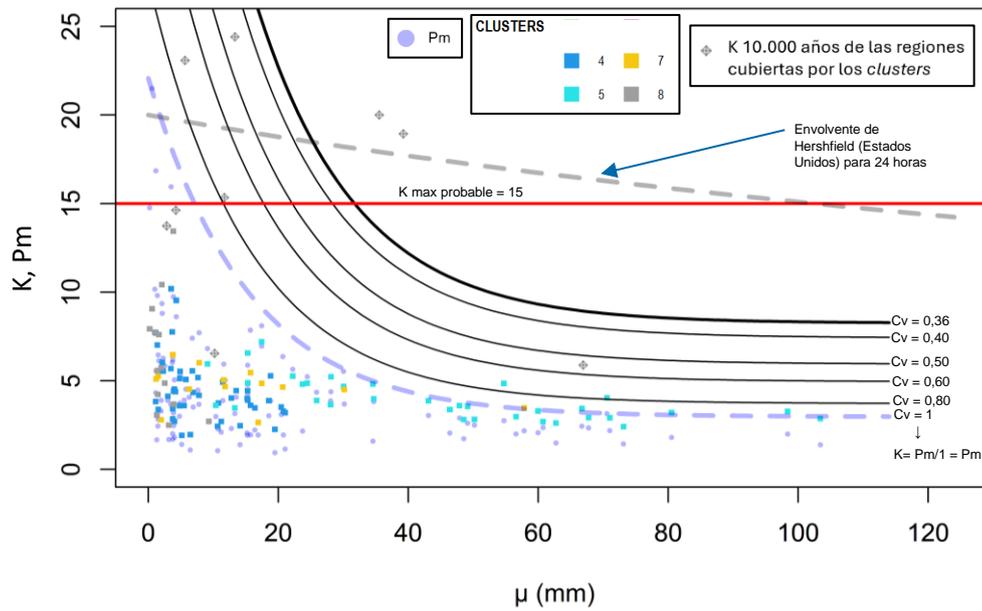
Figura 46 P_m versus μ del cluster 6 con su envolvente, para 24 horas



Fuente: Elaboración propia.

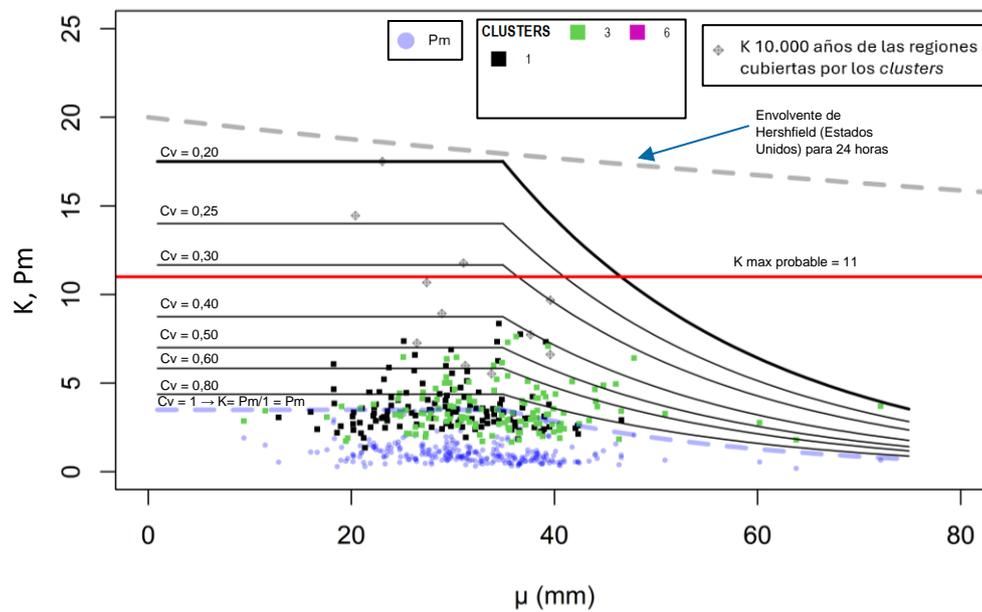
Tras haber trazado las envolventes de la precipitación máxima adimensional, se amplificaron de acuerdo con la relación despejada por Stowhas (1983), la cual corresponde a la ecuación (2.7) que nos dice que $K = \frac{P_m}{C_v}$, para el caso de 24 horas, en las Figura 47, Figura 48, Figura 49 y Figura 50 se muestran de manera simultánea P_m y K en función de μ , así como las curvas envolventes de P_m amplificadas a partir de múltiples valores de coeficientes de variación, los cuales se fueron probando de manera progresiva descendente partiendo de 1 hasta encontrar el más adecuado, considerando también los valores mínimos de C_v . Para la presente evaluación también se tomó en cuenta los valores de $K_{max\ probable}$ identificados previamente para cada grupo de *clusters*.

Figura 47 Amplificación de envoltura de los clusters 4; 5; 7 y 8, para 24 horas



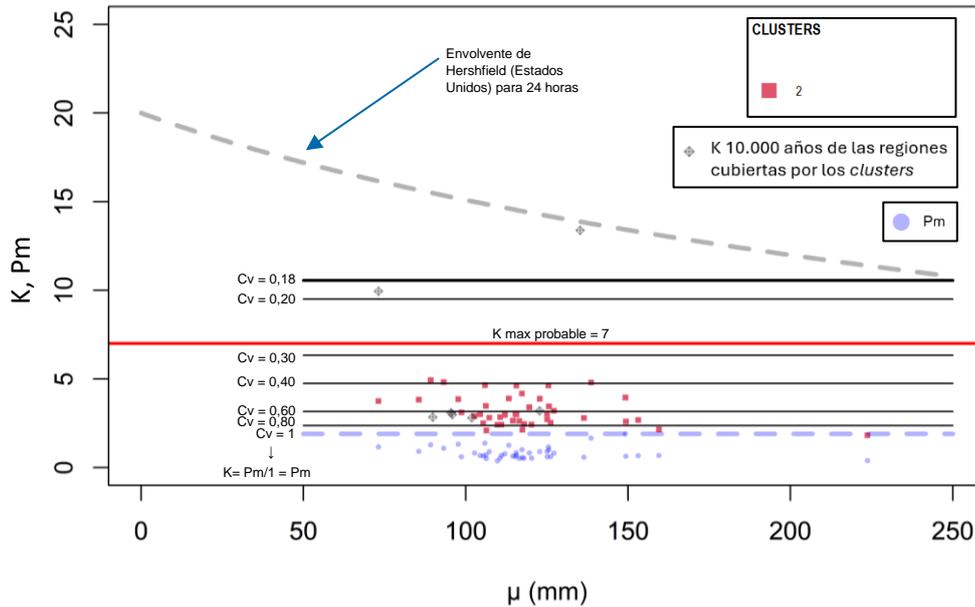
Fuente: Elaboración propia.

Figura 48 Amplificación de envoltura de los clusters 1 y 3, para 24 horas



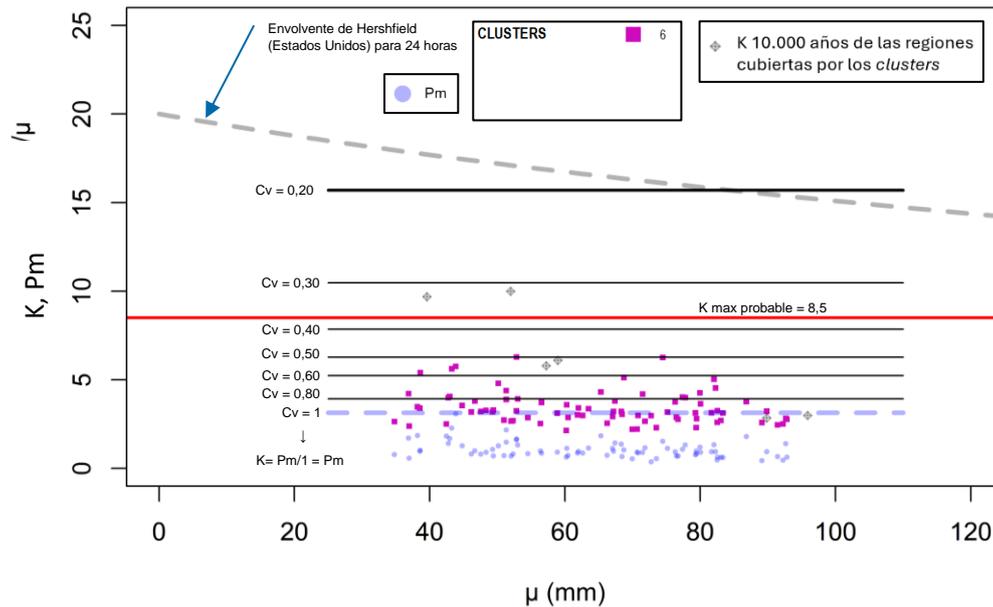
Fuente: Elaboración propia.

Figura 49 Amplificación de envolvente del cluster 2, para 24 horas



Fuente: Elaboración propia.

Figura 50 Amplificación de envolvente del cluster 6, para 24 horas



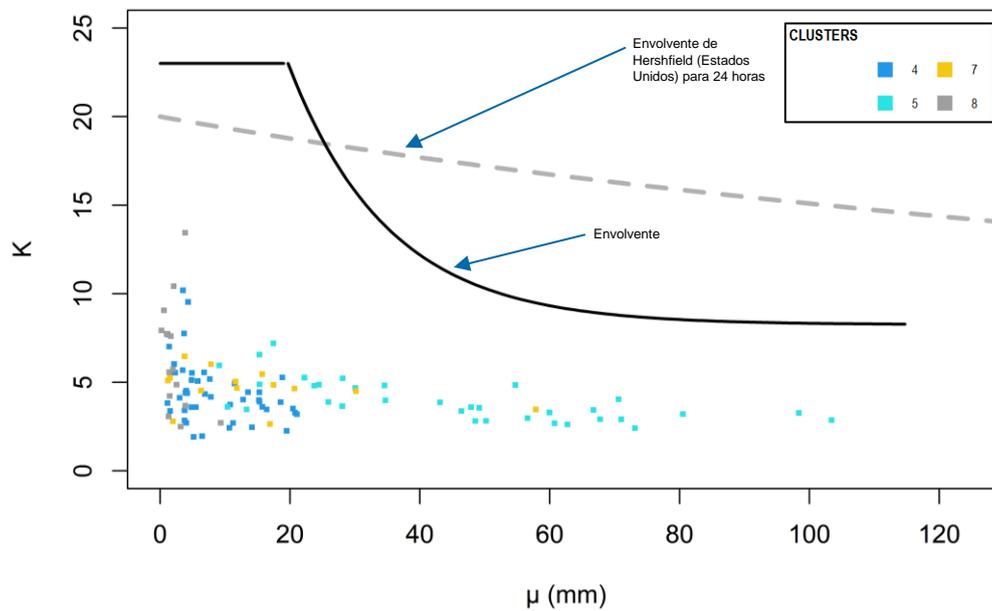
Fuente: Elaboración propia.

Los gráficos mencionados tienen como objetivo ayudar a seleccionar un valor de C_v que amplifique la envolvente de P_m de manera que la envolvente generada sea capaz de cubrir los valores graficados de K para llegar a la doble maximización buscada, dado que P_m se maximiza doblemente con $K = \frac{P_m}{C_v}$, es decir con la

selección de un valor de C_v lo más desfavorable posible, adicionalmente, justamente la relación $\frac{P_m}{C_v}$ determina la función de relación entre K y μ .

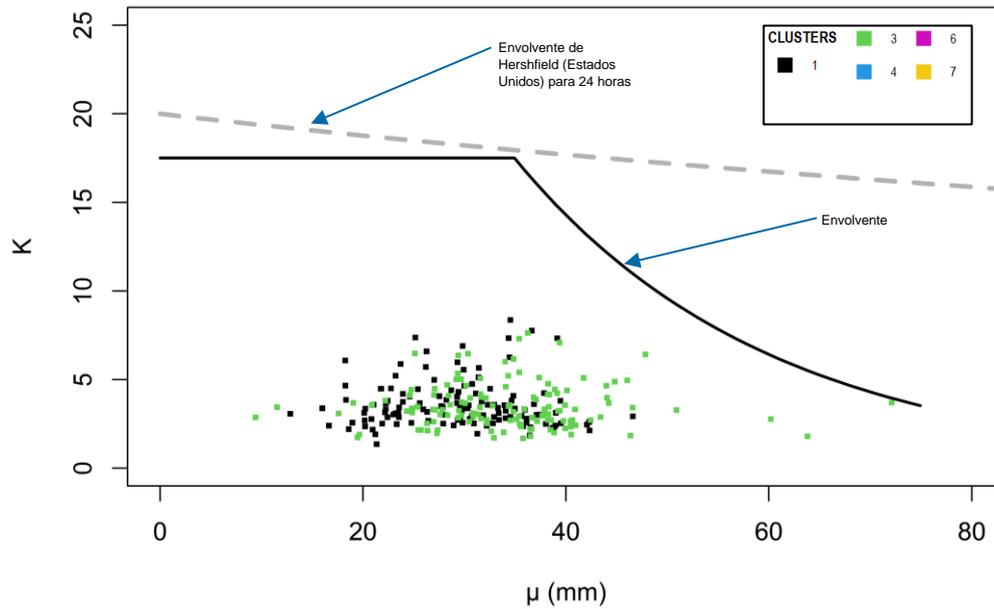
En las Figura 51, Figura 52, Figura 53 y Figura 54 se muestran las envolventes identificadas para cada grupo de *clusters* para el caso de 24 horas, las cuales se establecieron ya sea mediante combinaciones entre la envolvente amplificada de P_m con la recta horizontal definida por $K_{max\ probable}$ o solo considerando dicha recta. Las gráficas de las envolventes de cada grupo de *clusters* para las diferentes duraciones se pueden encontrar en el ANEXO C.

Figura 51 Envolvente de K versus μ de los *clusters* 4; 5; 7 y 8, para 24 horas



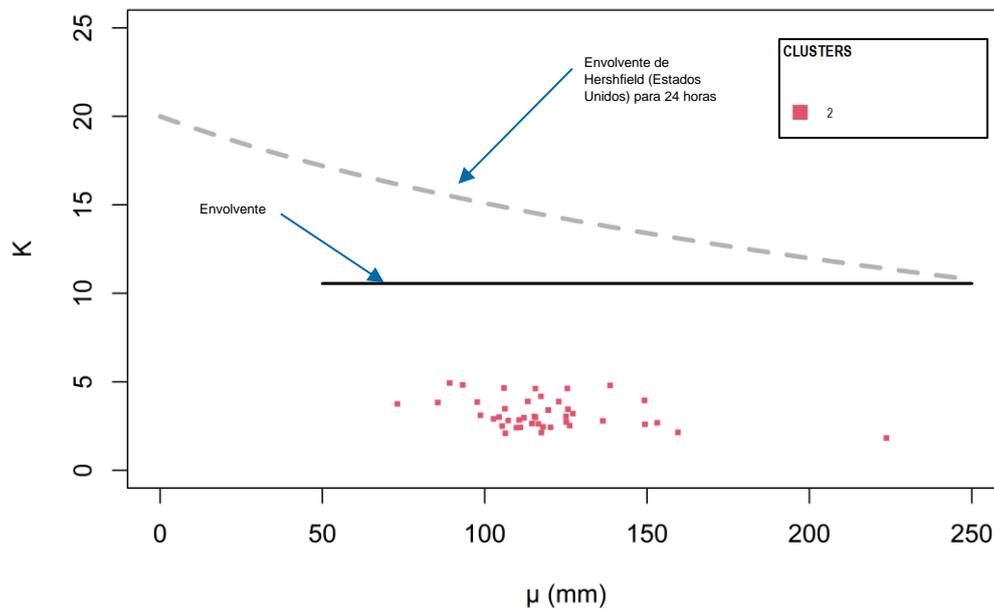
Fuente: Elaboración propia.

Figura 52 Envoltente de K versus μ de los clusters 1 y 3, para 24 horas



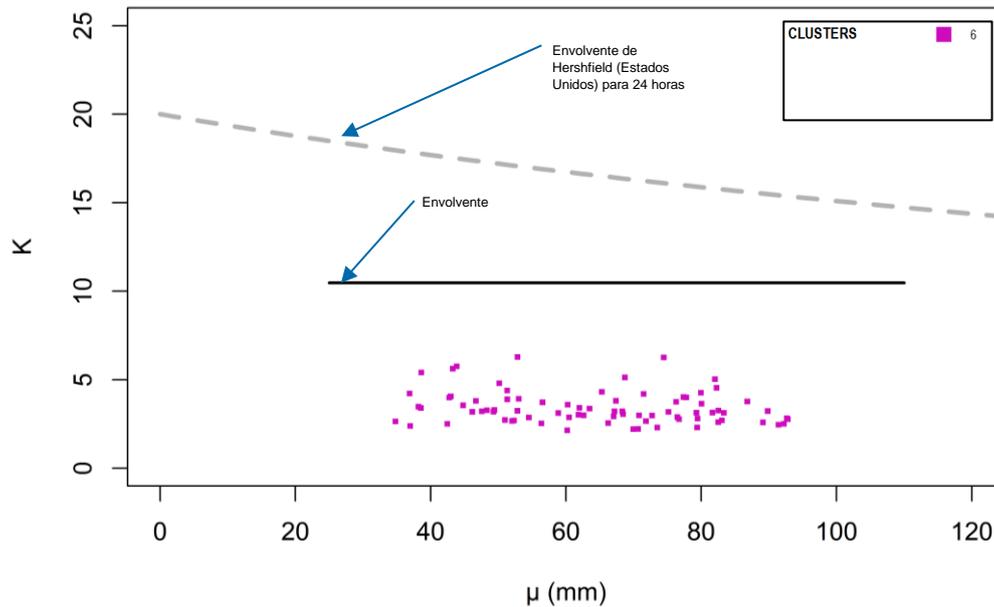
Fuente: Elaboración propia.

Figura 53 Envoltente de K versus μ del cluster 2, para 24 horas



Fuente: Elaboración propia.

Figura 54 Envoltente de K versus μ del cluster 6, para 24 horas

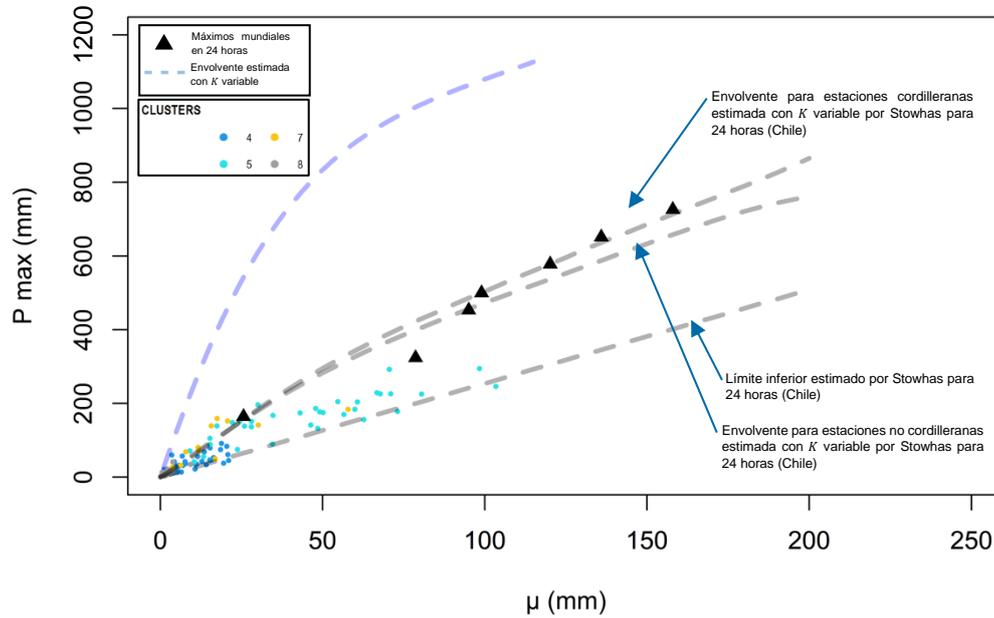


Fuente: Elaboración propia.

4.4 Determinación de las envolventes de máximos y PMP

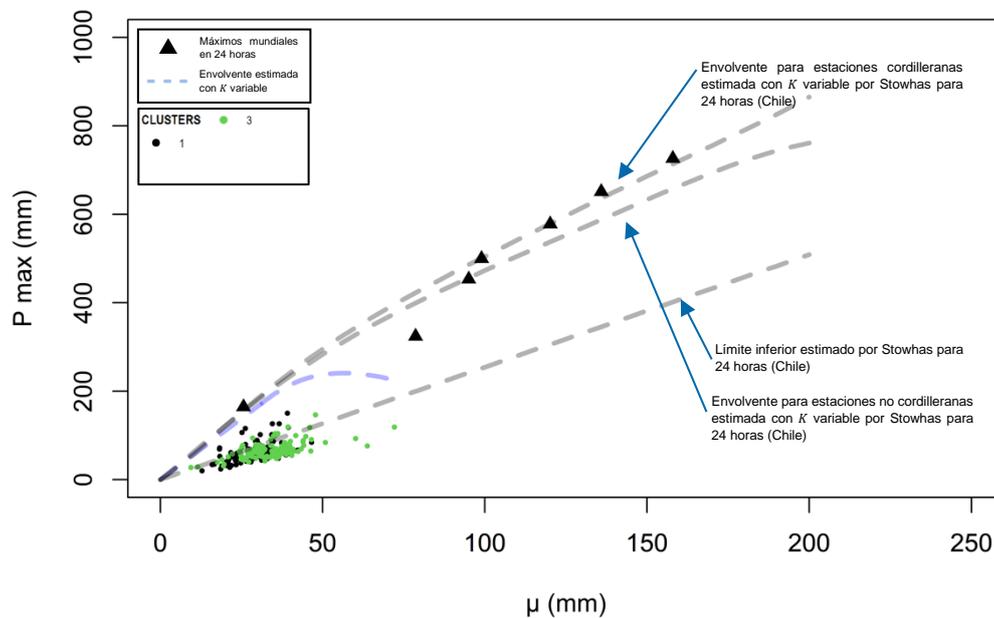
Adicional a lo visto en el subcapítulo anterior, se graficó las precipitaciones máximas (P_{max}) de cada estación en función de μ y las curvas generadas al combinar las ecuaciones de las envolventes de K versus Cv con las envolventes de Cv versus μ de cada grupo de *clusters* al reemplazar la ecuación de la segunda en la primera a fin de generar la ecuación de una envoltente de K versus μ que represente los casos más extremos posibles de K , se reemplazó la ecuación de la envoltente obtenida en la ecuación (2.8) determinando así un límite superior al que podrían llegar las precipitaciones máximas, en las Figura 55, Figura 56, Figura 57 y Figura 58 se muestran los casos para 24 horas.

Figura 55 P_{max} versus μ y envoltente generada de los clusters 4; 5; 7 y 8, para 24 horas



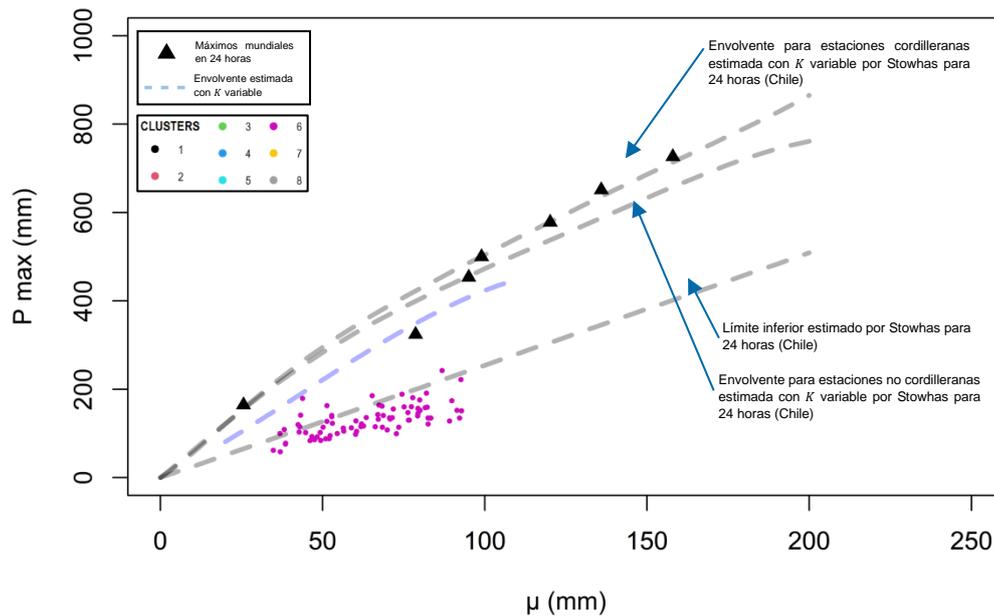
Fuente: Elaboración propia.

Figura 56 P_{max} versus μ y envoltente generada de los clusters 1 y 3, para 24 horas



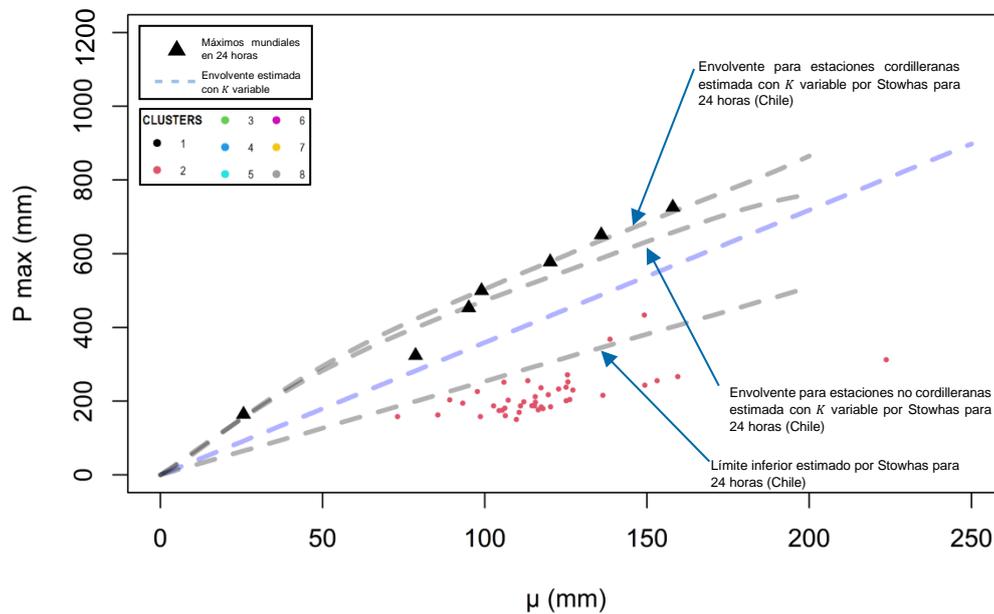
Fuente: Elaboración propia.

Figura 57 P_{max} versus μ y envolvente generada del cluster 6, para 24 horas



Fuente: Elaboración propia.

Figura 58 P_{max} versus μ y envolvente generada del cluster 2, para 24 horas



Fuente: Elaboración propia.

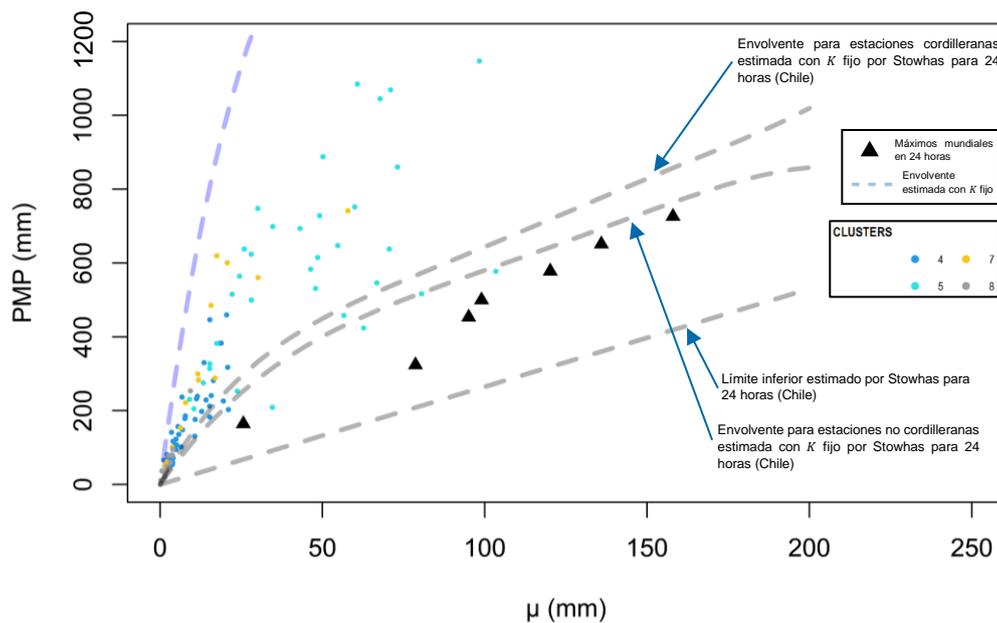
Se puede apreciar que la envolvente de los clusters 4; 5; 7 y 8 sobrepasa considerablemente a las envolventes definidas por Stowhas, debido a que el comportamiento de los datos de las estaciones de la costa norte del Perú se encuentra influenciados por el Fenómeno El Niño y las estaciones que se

encuentran cercanas al litoral de los *clusters* 7 y 8 (menos de 15 km de distancia al mar) tienen un comportamiento particular, el cual corresponde a un C_v alto y un μ muy bajo.

Por otro lado, las envolventes de los demás grupos de *clusters* mantienen congruencia con las envolventes establecidas por Stowhas y los registros máximos mundiales graficados, teniendo en cuenta que las envolventes son establecidas específicamente para los *clusters* mencionados a diferencia de las envolventes de Stowhas que fueron establecidas para todo Chile.

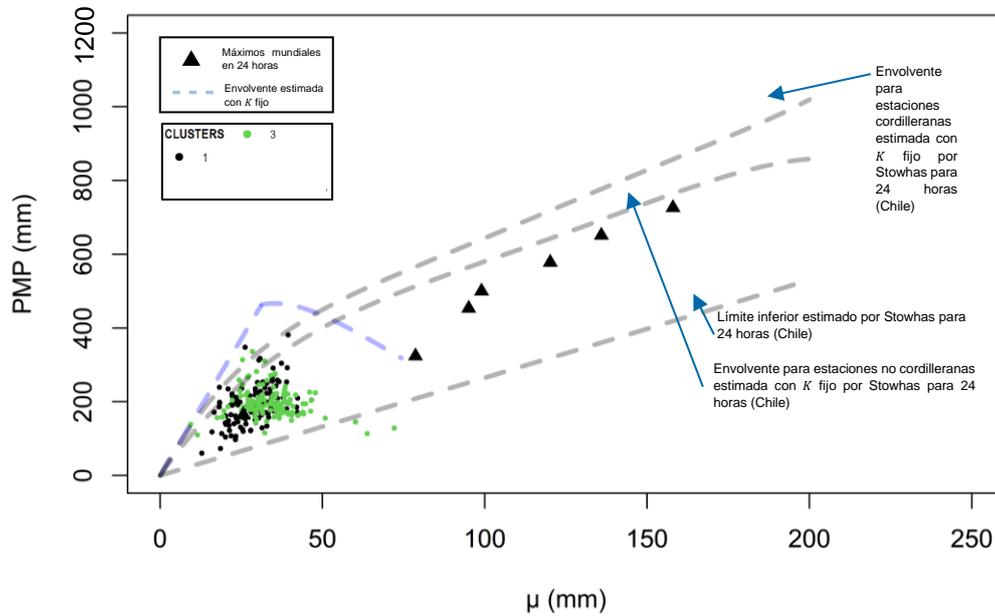
Adicional a lo anterior, tomando el criterio mencionado de maximizar K y C_v pero colocando directamente el valor del K_{max} en la ecuación (2.8), se logró obtener una envolvente para la PMP en las Figura 59, Figura 60, Figura 61 y Figura 62 se muestran los casos para 24 horas.

Figura 59 PMP versus μ y envolventes generadas de los *clusters* 4; 5; 7 y 8, para 24 horas



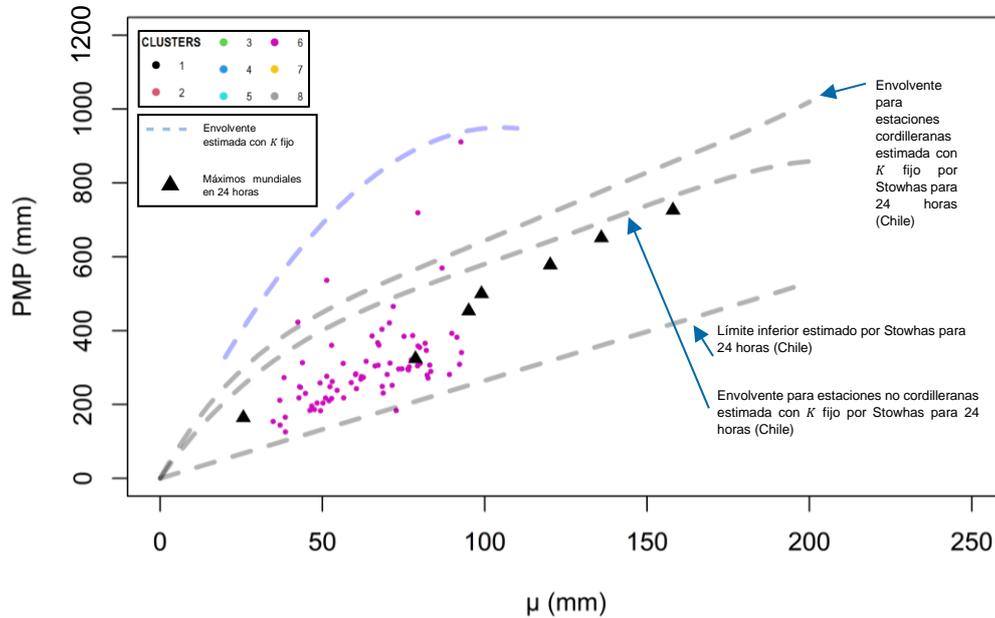
Fuente: Elaboración propia.

Figura 60 PMP versus μ y envolventes generadas de los clusters 1 y 3, para 24 horas



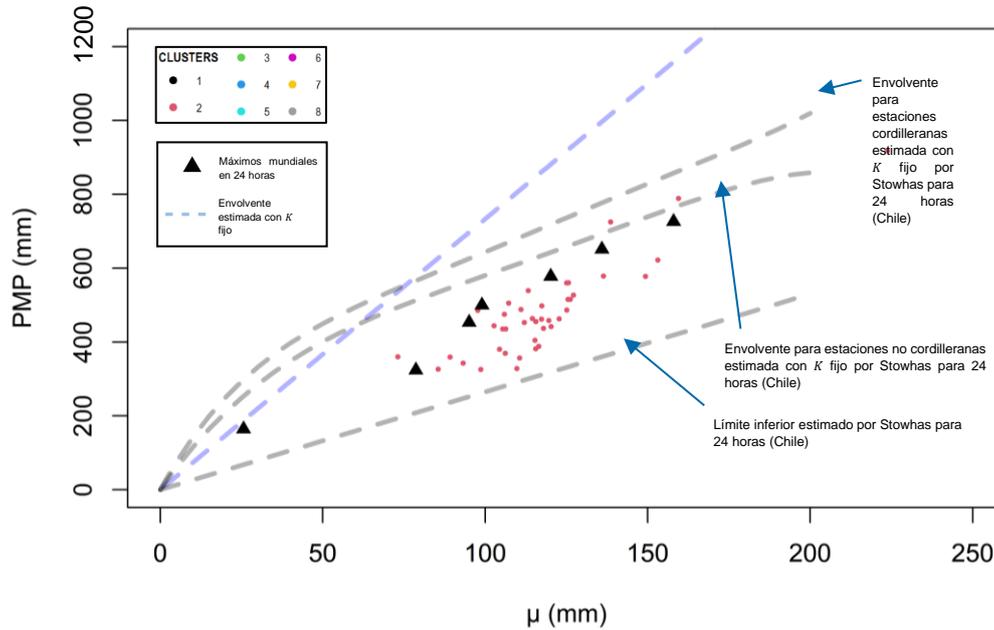
Fuente: Elaboración propia.

Figura 61 PMP versus μ y envolventes generadas del cluster 6, para 24 horas



Fuente: Elaboración propia.

Figura 62 PMP versus μ y envolventes generadas del cluster 2, para 24 horas



Fuente: Elaboración propia.

Se puede apreciar en este caso también que la envolvente de los *clusters* 4; 5; 7 y 8 sobrepasa considerablemente a las envolventes definidas por Stowhas, debido a los motivos antes mencionados, el comportamiento de los datos de las estaciones de la costa norte del Perú se encuentra influenciados por el Fenómeno El Niño y las estaciones que se encuentran cercanas al litoral de los *clusters* 7 y 8 (menos de 15 km de distancia al mar) tienen un comportamiento particular, el cual corresponde a un Cv alto y un μ muy bajo.

Por otro lado, tal como en el caso anterior, las envolventes de los demás grupos de *clusters* mantienen congruencia con las envolventes establecidas por Stowhas y los registros máximos mundiales graficados, teniendo en cuenta que las envolventes son establecidas específicamente para los *clusters* mencionados a diferencia de las envolventes de Stowhas que fueron establecidas para todo Chile.

Capítulo V: Cálculo de la PMP

En el presente capítulo se describe la evaluación de la PMP a partir de las envolventes propuestas en el capítulo anterior, así como el procedimiento y consideraciones tomados para la generación de mapas de la PMP mediante el uso de una regresión espacial.

Hablando de la ubicación espacial, los *clusters* no permiten definir fronteras explícitas entre ellos, sino que hay zonas donde se encuentran estaciones de diferentes *clusters*, es decir, existe un solapamiento entre ellos, por ello se eligió que, como producto final, luego de hacer el cálculo de la PMP en cada estación se generase un *raster* de para las localizaciones intermedias, en vez de delimitar fronteras entre *clusters* y envolventes para dichas fronteras.

5.1 Cálculo de la PMP para cada estación

Para cada estación se calculó la media y la desviación estándar, cada uno de ellos se calculó incluyendo y excluyendo el máximo, así como los cocientes $\frac{\mu_{n-1}}{\mu}$ y $\frac{\sigma_{n-1}}{\sigma}$ y la cantidad de años de información disponible para cada estación, con dicha información se corrigieron los valor de la media y desviación standard de acuerdo al procedimiento propuesto por Herschfield (1961).

Con las curvas envolventes propuestas en el capítulo anterior se pudo estimar el $K_{cluster}$ de cada estación pluviométrica para las duraciones de 24; 48 y 72 horas, tras haber evaluado el valor de la media en la función de la curva envolvente del parámetro K correspondiente al *cluster* al que pertenece cada estación.

A partir de los valores de $K_{cluster}$ obtenidos, se realizó el cálculo de la PMP mediante la fórmula de Hershfield (1961), mostrada en la ecuación (2.2), para cada estación y para cada duración de estudio, obteniendo de esa manera una bases de datos de valores puntuales de PMP para cada una de las distintas duraciones; dicha base de datos final englobaría a aquellas estaciones que cumplieron con los diferentes análisis estadísticos y pasaron el filtro de cantidad mínima de 30 años de información.

5.1.1 Verificación y corrección de valores de PMP

Dado que existen determinados casos particulares de estaciones las cuales la diferencia de su media entre ciertas duraciones de lluvia es de una gran magnitud

o a veces casi nula y esto puede desembocar en un valor de K que origine valores de PMP incongruentes (por ejemplo, una PMP de 48 horas ligeramente menor que la PMP de 24 horas) se realizó una verificación de toda la base de datos de PMP, para identificar los siguientes posibles errores:

- Error 1: La PMP de 48 horas sea mayor que la PMP de 24 horas;
- Error 2: La PMP de 72 horas sea mayor que la PMP de 48 horas;
- Error 3: La PMP de 72 horas sea mayor que la PMP de 24 horas.

En la Tabla 10 se muestran los errores identificados con los nombres de las estaciones respectivas y un valor de diferencia entre los valores de PMP que generan cada tipo de error.

Tabla 10 Errores identificados en los valores de PMP

Estación	Error 1	Δ PMP error 1 (mm)	Error 2	Δ PMP error 2 (mm)	Error 3	Δ PMP error 3 (mm)
La Esperanza	X	6.075		-		-
Pucala	X	4.086		-		-
Santiago						
Antunez de Mayolo	X	3.415		-		-
Iñapari	X	12.190		-		-
Tingo	X	5.217		-		-
Paico	X	38.746		-		-
Huariaca		-	X	0.887		-

Fuente: Elaboración propia.

Dentro de los pocos errores encontrados, en la mayoría de los casos son errores menores, siendo el más crítico el caso de la estación Paico, el cual tiene su explicación en el comportamiento particular de sus series de precipitación, teniendo una diferencia muy notoria entre la media de 24 horas (35,4 mm) y la de 48 horas (50.8 mm), lo cual llevaría a que el K de 48 horas acabe siendo menor al del correspondiente de 24 horas y finalmente estimando una PMP de 48 horas menor a la de 24 horas (ver gráficas de envolventes en el ANEXO C).

Por tanto, como medida de remediación para los errores presentes en 7 estaciones (1.25% de un total de 559 estaciones en estudio) siendo una de ellas un error por un caso particular (0.18% de un total de 559 estaciones en estudio) se decidió por reemplazar la PMP con un valor incongruente con la PMP superior en cada caso de error.

5.2 Regresión espacial y generación de imágenes rasterizadas de la PMP

Para obtener la extensión de la evaluación de la PMP a la mayor parte posible del territorio nacional, se aplicó un análisis y regresión espacial al coeficiente de variación de los máximos anuales, el cual es una de las medidas estadísticas que participan en la formulación de Hershfield, asimismo se realizó ello para los parámetros de ubicación (longitud, latitud y altitud), obteniendo así, *rasters* de dichas variables que sirvieron como insumo para finalmente realizar la regresión espacial de la PMP.

La cantidad de píxeles que conforman los *rasters* elaborados en la presente tesis tienen una resolución de $0,01^\circ \times 0,01^\circ$ se estima adecuada para generar una resolución óptima que deje apreciar a detalle el comportamiento de las variables y a su vez es lo suficientemente baja para evitar que el *raster* se convierta en un archivo que muy pesado y que se vuelva demasiado complejo de procesar

Para tomar la decisión de cuál sería el mejor método para la regresión espacial, se probaron diversos métodos: Support Vector Machines (SVM) de tipo radial, Random Forest (RF), Generalized Lineal Models (GLM), Generalized K Nearest Neighbors (GKNN), gcvEarth, entre otros.

Tras realizar la regresión espacial empleando los diversos métodos y observando el error de raíz cuadrada media (RMSE por sus siglas en inglés), coeficiente de determinación (R^2) y error absoluto medio (MAE por sus siglas en inglés), se eligió al método al SVM de tipo radial porque puede representar una distribución espacial con mayor flexibilidad, debido a que tiene 2 parámetros ajustables que permiten la manipulación eficiente de los resultados a fin de obtener errores menores en las simulaciones:

- σ : El parámetro σ de la función de "kernel" gaussiana controla la anchura de la "campana". Si se utiliza un valor grande para σ los valores de los puntos de las características conocidas se extenderán más, lo que

resultará en un cambio más suave en sus valores. Por otro lado, si se utiliza un valor pequeño para σ , los valores de los puntos de nuestras características conocidas tendrán un dominio de valores más reducido y el cambio entre sus valores será más brusco.

- C : El parámetro de costo, también conocido como el costo de la clasificación errónea, determina cuánta flexibilidad se debe permitir con los datos. Si se utiliza un valor bajo para C , se obtiene una superficie de decisión suave, mientras que un valor alto para C permite clasificar puntos correctamente.

Se tomaron como variables características para la interpolación espacial a la latitud, la longitud, la altitud y el coeficiente de variación. Al código de programación generado en lenguaje R se le ingresó una base de datos organizada con los datos de las estaciones consistente en las variables mencionadas más la PMP calculada, de modo que, a partir de la información rasterizada de elevaciones, latitudes, longitudes, y de los coeficientes de variación, se pudiera estimar el valor de la PMP correspondiente a cada celda del *raster* generado, considerando un 95% de los datos como datos de calibración y un 5 % de los mismos como datos de validación de las simulaciones, y estableciendo una misma semilla en todas las evaluaciones a fin de garantizar obtener un mismo resultado cada vez que se ejecute el código con los mismos datos.

El procedimiento se aplicó para cada una de las duraciones 24; 48 y 72 horas y requirió el desarrollo previo de información rasterizada de las variables características.

En la Tabla 11 muestra los parámetros del método SVM radial utilizados para la interpolación espacial del Cv . Asimismo, la Tabla 12 y Tabla 13 muestran las métricas de ajuste de lo simulado en el procedimiento de calibración y validación en el caso del Cv .

Tabla 11 Parámetros para la regresión del Cv

Duración	σ	C
24 horas	0,50	9
48 horas	0,60	9
72 horas	0,80	10

Fuente: Elaboración propia.

Tabla 12 Métricas del proceso de calibración del C_v

Duración	$RMSE$	R^2	MAE
24 horas	0,1801	0,7310	0,1037
48 horas	0,1899	0,7444	0,1023
72 horas	0,1866	0,7372	0,1020

Fuente: Elaboración propia.

Tabla 13 Métricas del proceso de validación del C_v

Duración	$RMSE$	R^2	MAE
24 horas	0,1497	0,8949	0,0837
48 horas	0,1782	0,7193	0,1301
72 horas	0,1760	0,8158	0,1119

Fuente: Elaboración propia.

A partir de un *raster* de elevaciones correspondiente a Shuttle Radar Topography Mission (NGA & NASA, 2013), se generó un *raster* de latitudes y otro de longitudes, teniendo así, 3 *rasters* con los parámetros de ubicación.

Po otro lado, también se trabajaron *rasters* de medias, en la Tabla 14 se muestra los parámetros del método SVM radial utilizados para la interpolación espacial del C_v . Asimismo, la Tabla 15 y Tabla 16 muestran las métricas de ajuste de lo simulado en el procedimiento de calibración y validación en el caso del C_v .

Tabla 14 Parámetros para la regresión de μ

Duración	σ	C
24 horas	0,9	3
48 horas	1,4	6
72 horas	2,2	4

Fuente: Elaboración propia.

Tabla 15 Métricas del proceso de calibración de μ

Duración	$RMSE$	R^2	MAE
24 horas	13,1454	0,8286	7,7792
48 horas	16,0586	0,8129	10,006
72 horas	19,3352	0,7932	12,1821

Fuente: Elaboración propia.

Tabla 16 Métricas del proceso de validación de μ

Duración	RMSE	R^2	MAE
24 horas	9,7199	0,9208	6,8804
48 horas	20,2081	0,7539	10,3472
72 horas	8,5257	0,9576	6,7985

Fuente: Elaboración propia.

Los *rasters* generados de la media no han sido considerados para la elaboración de los *rasters* finales de la PMP dado que se observó que los resultados de las métricas evaluadas y en general el comportamiento del *raster* de la PMP era más congruente con la realidad cuando no se consideraba el *raster* mencionado entre las variables de input para la simulación. Otro sustento para ello se encuentra en el hecho que existen estaciones con medias de magnitud pequeña pero que poseen una PMP superior a la de estaciones con medias más altas (ver.Figura 59, Figura 60, Figura 61 y Figura 62)

Finalmente, con el conjunto de *rasters* de las variables características se llevó a cabo la interpolación espacial de la PMP. En la Tabla 17 se pueden apreciar los parámetros utilizados para la generación de la interpolación espacial de la PMP usando el método SVM radial, por otro lado, en las Tabla 18 y 19 se muestran las métricas de error asociadas al proceso de calibración y validación de la generación de los *rasters*.

Tabla 17 Parámetros para las regresiones de la PMP

Duración	σ	C
24 horas	1	6
48 horas	0,20	5
72 horas	0,90	6

Fuente: Elaboración propia.

Tabla 18 Métricas del proceso de calibración de la PMP

Duración	RMSE	R^2	MAE
24 horas	104.5533	0.6471	54.5781
48 horas	118.9540	0.7013	62.6296
72 horas	146.7101	0.6830	88.4829

Fuente: Elaboración propia.

Tabla 19 Métricas del proceso de validación de la PMP

Duración	RMSE	R ²	MAE
24 horas	112.3122	0.8000	58.5115
48 horas	174.6982	0.4894	77.5433
72 horas	182.2990	0.7987	92.8786

Fuente: Elaboración propia.

En los *rasters* formados a partir de la regresión espacial elegido se obtuvieron valores un poco distintos a los valores puntuales, puesto que este método suaviza los saltos abruptos en los valores para dar una representación más congruente. Sin embargo, existen dos condiciones que no son óptimas para realizar el análisis espacial, las cuales son el comportamiento de los valores de precipitación son bastante cambiante en cortas distancias y de la falta de estaciones con amplios registros de datos en el territorio de la selva peruana.

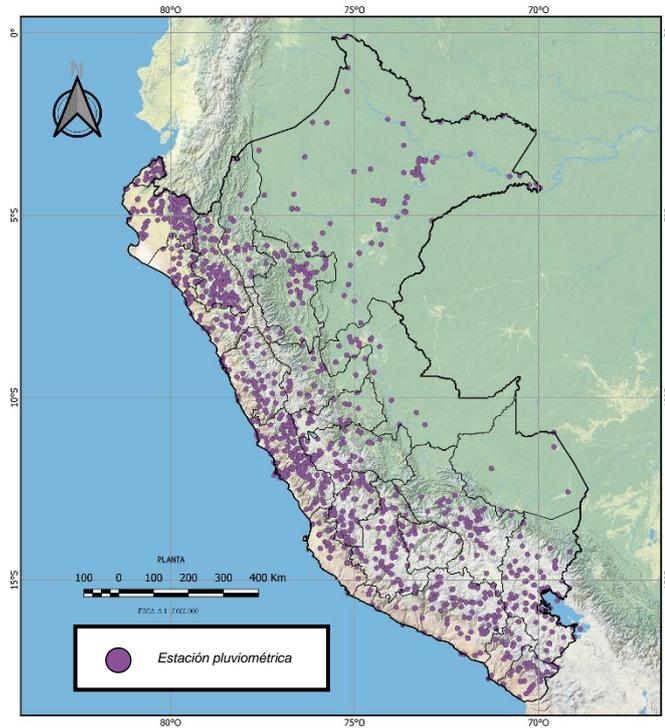
Cabe resaltar que, las precipitaciones pueden variar significativamente en distancias cortas en las lluvias más fuertes, los datos de los pluviómetros son mediciones puntuales muy específicas de una ubicación y no hay suficientes pluviómetros con un amplio registro de datos que faciliten reflejar dicho comportamiento.

5.3 Consideraciones en la elaboración de mapas debido a las zonas con información escasa

Actualmente en el territorio peruano existe una escasez de estaciones pluviométricas en la selva, como se puede apreciar en la Figura 63, la diferencia en densidad de estaciones a comparación de la costa y la sierra es fácilmente notable; la escasez es aún más notoria en los departamentos de Madre de Dios y Ucayali. Hay múltiples motivos por los cuales se presenta esa escasez, algunos son:

- La selva peruana tiene una gran variedad de climas y múltiples ecosistemas bastante diversos, lo que dificulta la instalación y el mantenimiento de nuevas estaciones pluviométricas.
- La selva peruana es un área remota y de difícil acceso, lo que genera altos costos logísticos y operativos para los pluviómetros.
- La selva peruana es una zona de conflicto social y ambiental que podría amenazar la seguridad e integridad de los pluviómetros.

Figura 63 Ubicación de todas las estaciones pluviométricas a nivel nacional

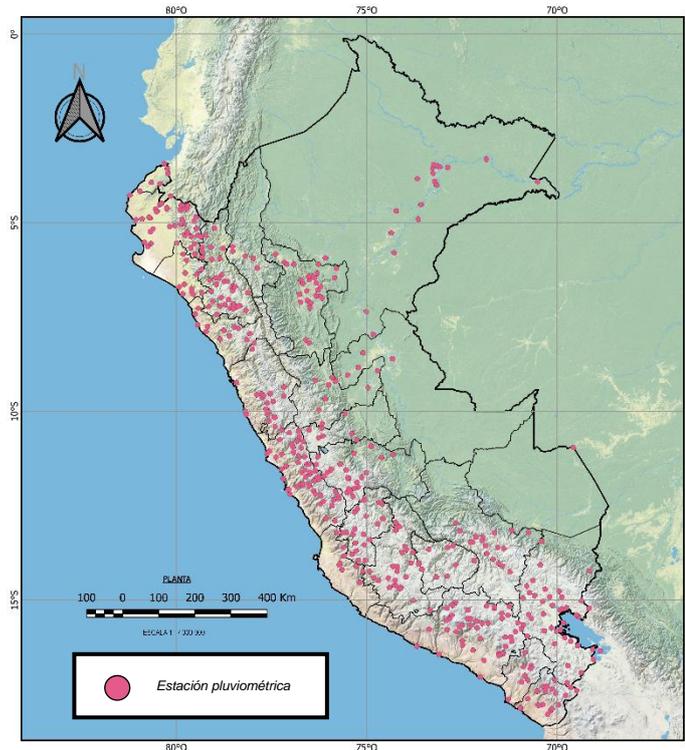


Fuente: Elaboración propia.

Asimismo, a medida que nos acercamos al territorio que abarca la sierra norte y la selva peruana en general, se puede apreciar una disminución considerable en la densidad de estaciones por unidad de área a diferencia de la costa, donde dicha densidad es muy superior, la distribución actual de estaciones a lo largo y ancho del Perú no es ideal para realizar una buena representación de las medidas estadísticas de las series de lluvias.

Aun existiendo la peculiaridad de la baja densidad de estaciones en determinadas partes del Perú, la cantidad de estaciones en todo el país disminuye a casi la mitad luego de realizar la depuración de estaciones con escasos registros pluviométricos y aquellas estaciones que no cumplieron los análisis estadísticos realizados, quedando así una densidad sumamente baja de estaciones en la selva peruana sobre todo, la Figura 64 muestra el escenario mencionado.

Figura 64 Ubicación de las estaciones pluviométricas restantes para el caso de 24 horas



Fuente: Elaboración propia.

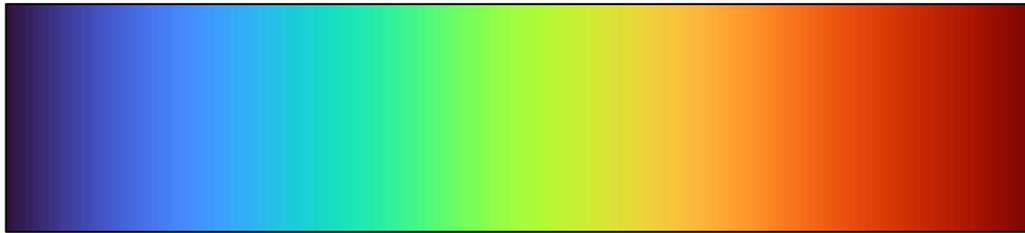
A pesar de ello, se vio que es posible estimar la PMP y otras variables en el Perú en ubicaciones donde no hay una alta densidad de estaciones pluviométricas mediante métodos de interpolación espacial, sin embargo, la mayoría de ellos tienden a realizar una estimación no del todo correcta en ubicaciones donde la escasez de datos es demasiado alta y no se tiene una condición de frontera (límites del área geográfica del Perú) que ayude al método a estimar valores con una transición suave en las ubicaciones mencionadas.

Para que se pueda representar la PMP en zonas donde haya una baja densidad de estaciones, se decidió asignar un estilo de bandas discretizadas a los *rasters* generados previamente con el método SVM radial. Se cambió el estilo de representación de los *rasters* (lo cual no afecta a los valores de los *raster*, es únicamente estético), de manera que se puedan visualizar los valores de los *raster* mediante una variedad de colores más amplia de la que se presenta por defecto (escala de blancos y negros) y que sea posible visualizar de manera más evidente el comportamiento de los valores, mostrando como estos aumentan o disminuyen

y la velocidad del cambio a través de bandas diferenciadas de rangos de valores, similares a curvas a nivel.

El estilo utilizado está basado en la rampa de color llamada Turbo, el cual se puede apreciar en la Figura 65, ya que posee una amplia variedad de colores que van cambiando de manera gradual y con una transición suave que ayuda al visualizador a diferenciar fácilmente el valor representado comparando el color con el de la leyenda.

Figura 65 *Rampa de color Turbo*



Fuente: Elaboración propia.

Adicionalmente, se optó por censurar u ocultar determinadas zonas de la selva peruana en la representación del mapa en los cuales la densidad de estaciones extremadamente baja o no hay estaciones en absoluto, de manera que solo se mostró lo resaltado en color azul en la Figura 66, puesto que, si bien, el método de regresión puede calcular datos en las ubicaciones mencionadas, dichos datos no llegarían a ser precisos, puesto que no estarían respaldados por información puntual de PMP de una estación en concreto.

Figura 66 Censura adoptada en el mapa del Perú



Fuente: Elaboración propia.

Establecidas las consideraciones y los modos de representación mencionados, se generaron mapas que representan las principales medidas estadísticas de las series de precipitaciones, es decir, la media de los máximos anuales (ver Mapa D.3, Mapa E.3 y Mapa F.3), el coeficiente de variación de los máximos anuales (ver Mapa D.5, Mapa E.5 y Mapa F.5) y la propia PMP (ver Mapa D.7, Mapa E.7 y Mapa F.7).

Capítulo VI: Análisis y discusión de resultados

En el presente capítulo se evalúan y se discuten nuevamente determinados apartados previamente mencionados en la tesis con el propósito de responder determinadas dudas y aportar con aclaraciones sobre posibles diferencias entre los resultados obtenidos en la investigación respecto a otros que se obtuvieron en investigaciones previas que se utilizaron como referencias bibliográficas.

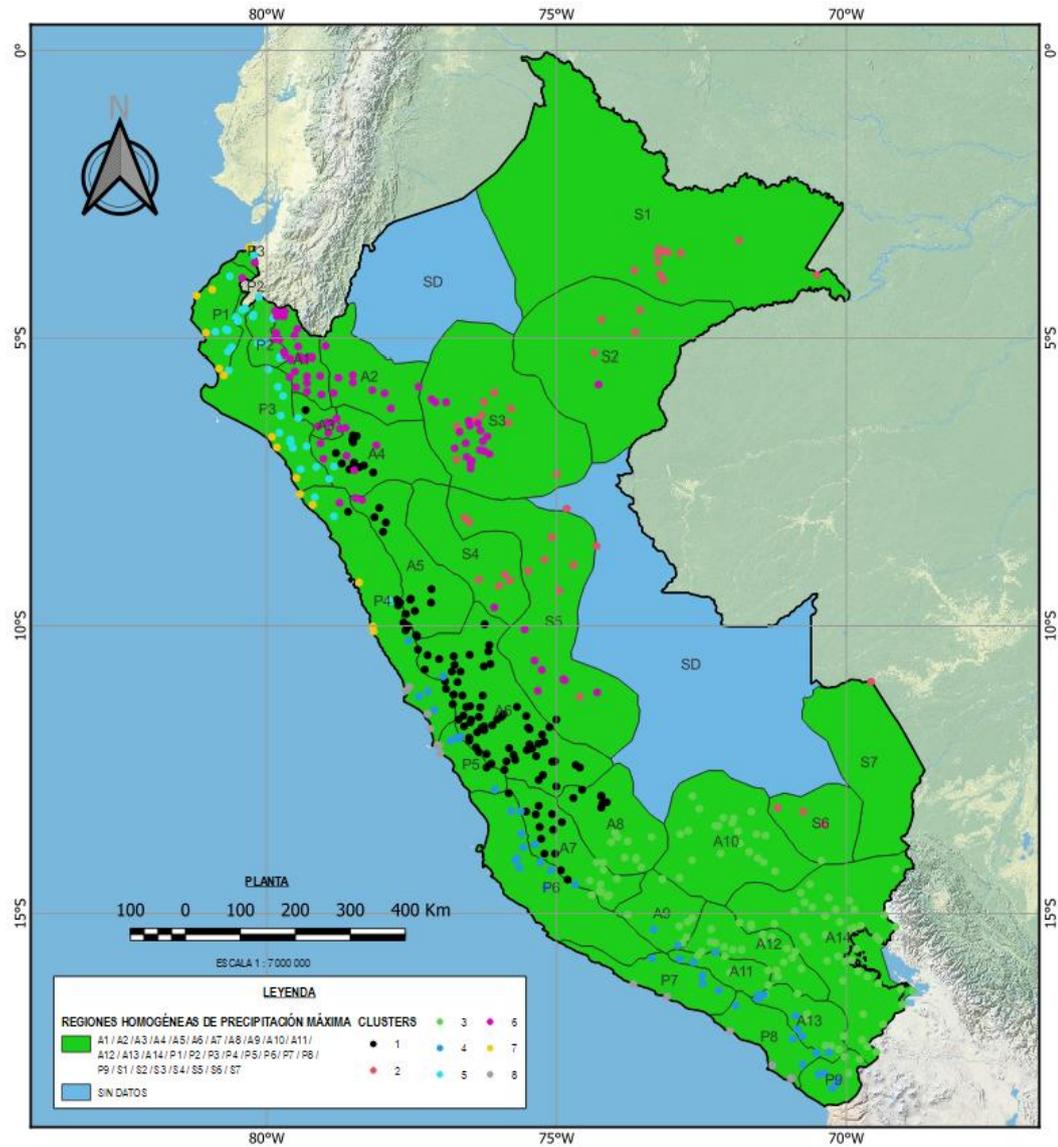
6.1 Clusters

Para verificar la calidad de los agrupamientos formados se tomó en cuenta si la distribución espacial de las estaciones pertenecientes a cada *cluster* es acorde a la delimitación territorial del Perú mostrada en el Mapa de Clasificación Climática Nacional (SENAMHI, 2020a), en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú (SENAMHI, 2017) y en las zonas pluviométricas identificadas en el “Estudio de la Hidrología del Perú” (Instituto Italo - Latino Americano et al., 1982).

Tal como se muestra en la Figura 67, Figura 68 y Figura 69, al superponer los *clusters* y las regiones establecidas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú, mostrado en el Mapa A.1, se puede apreciar que existen algunas relaciones próximas.

El *cluster* 1 pertenecen a la sierra central, abarca las regiones A4, A5, A6, mitad de A7 y A8; el *cluster* 2 pertenecen a la selva, abarca las regiones S1, S2, S3, S4, S5, S6 y S7; el *cluster* 3 pertenece a la sierra sur, abarca las regiones A9, A10, A12, A14, mitad de A7, A8, A11 y A13; el *cluster* 4 pertenece a la costa sur, abarca las regiones P5, P6, P7, P8, P9 y mitad de P4; el *cluster* 5 pertenece a la costa norte, abarca las regiones P1, P2, P3 y mitad de P4; el *cluster* 6 pertenece a la sierra norte, abarca las regiones A1, A2, A3, A4, mitad de S3 y S5.

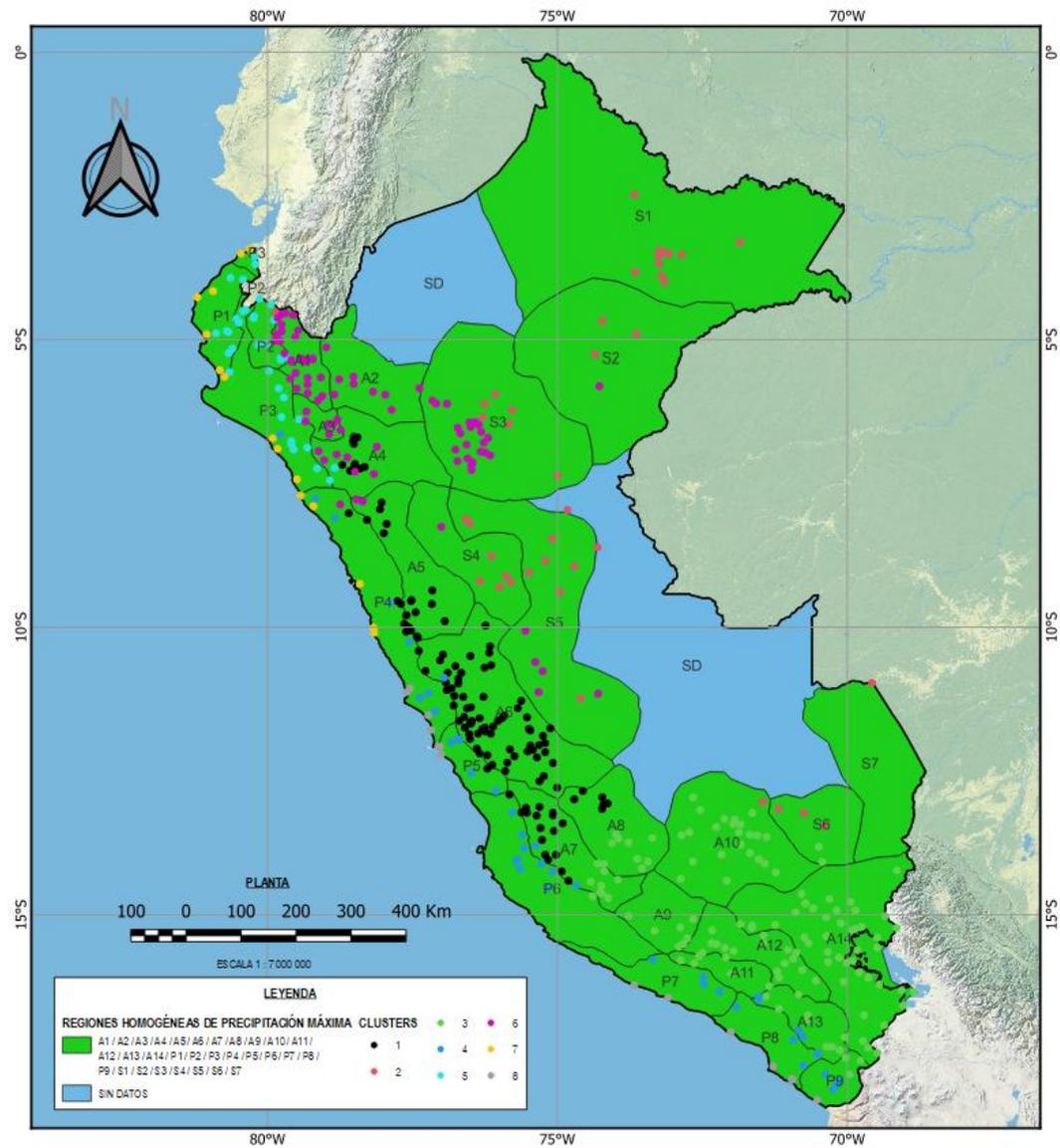
Figura 67 Clusters de estaciones para 24 horas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú



Fuente: Elaboración propia.

Nota: Las Regionalización de Precipitaciones Máximas en el Perú fueron adaptadas de Senamhi (2017)

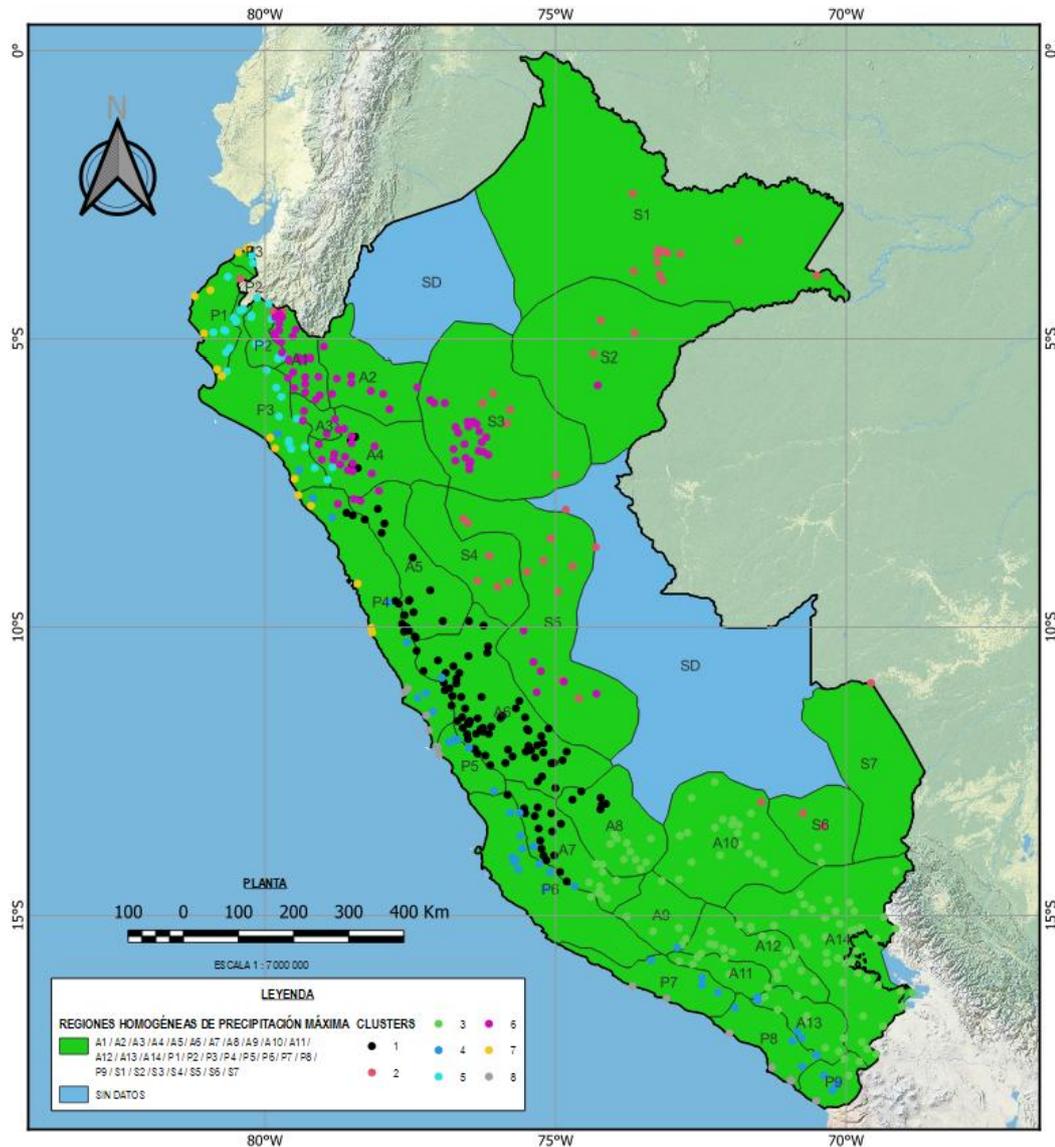
Figura 68 Clusters de estaciones para 48 horas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú



Fuente: Elaboración propia.

Nota: Las Regionalización de Precipitaciones Máximas en el Perú fueron adaptadas de Senamhi (2017)

Figura 69 Clusters de estaciones para 72 horas en el Mapa de la Regionalización de Precipitaciones Máximas en el Perú



Fuente: Elaboración propia.

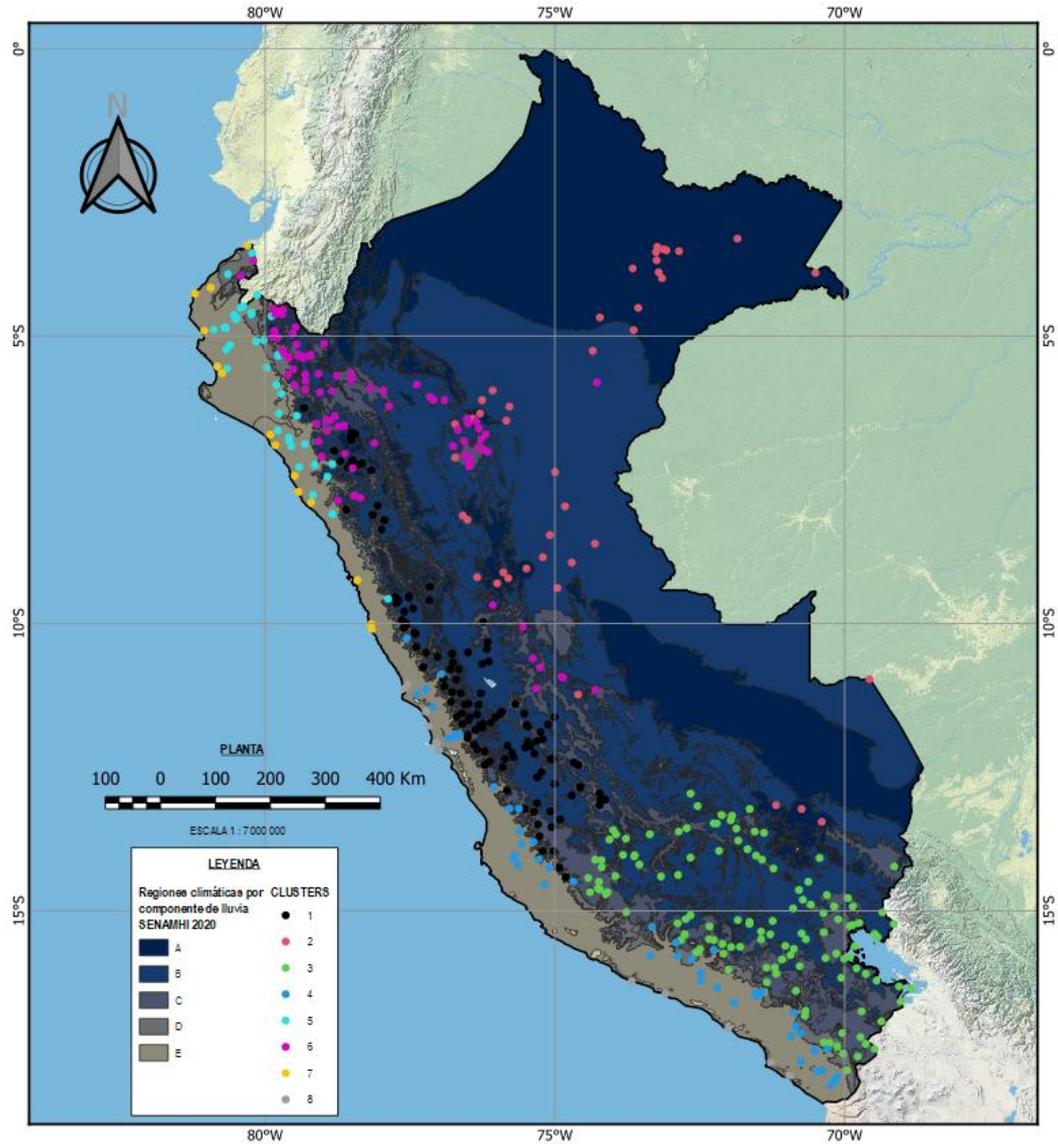
Nota: Las Regionalización de Precipitaciones Máximas en el Perú fueron adaptadas de Senamhi (2017)

Tal como se muestra en Figura 70, Figura 71 y Figura 72, al superponer los *clusters* y el Mapa de Clasificación Climática Nacional basado en su componente de lluvias, mostrado en el Mapa A.3, se puede apreciar que existen unas relaciones próximas.

Los *clusters* 1, 3 y 6 pertenecen a la sierra, se encuentra entre las regiones de lluvias B y C; el *cluster* 2 pertenece a la selva (laderas orientales de los Andes),

se encuentra entre las regiones de lluvias A y B; los *clusters* 4, 5, 7 y 8 pertenecen a la costa, se encuentran dentro de la región de lluvias E.

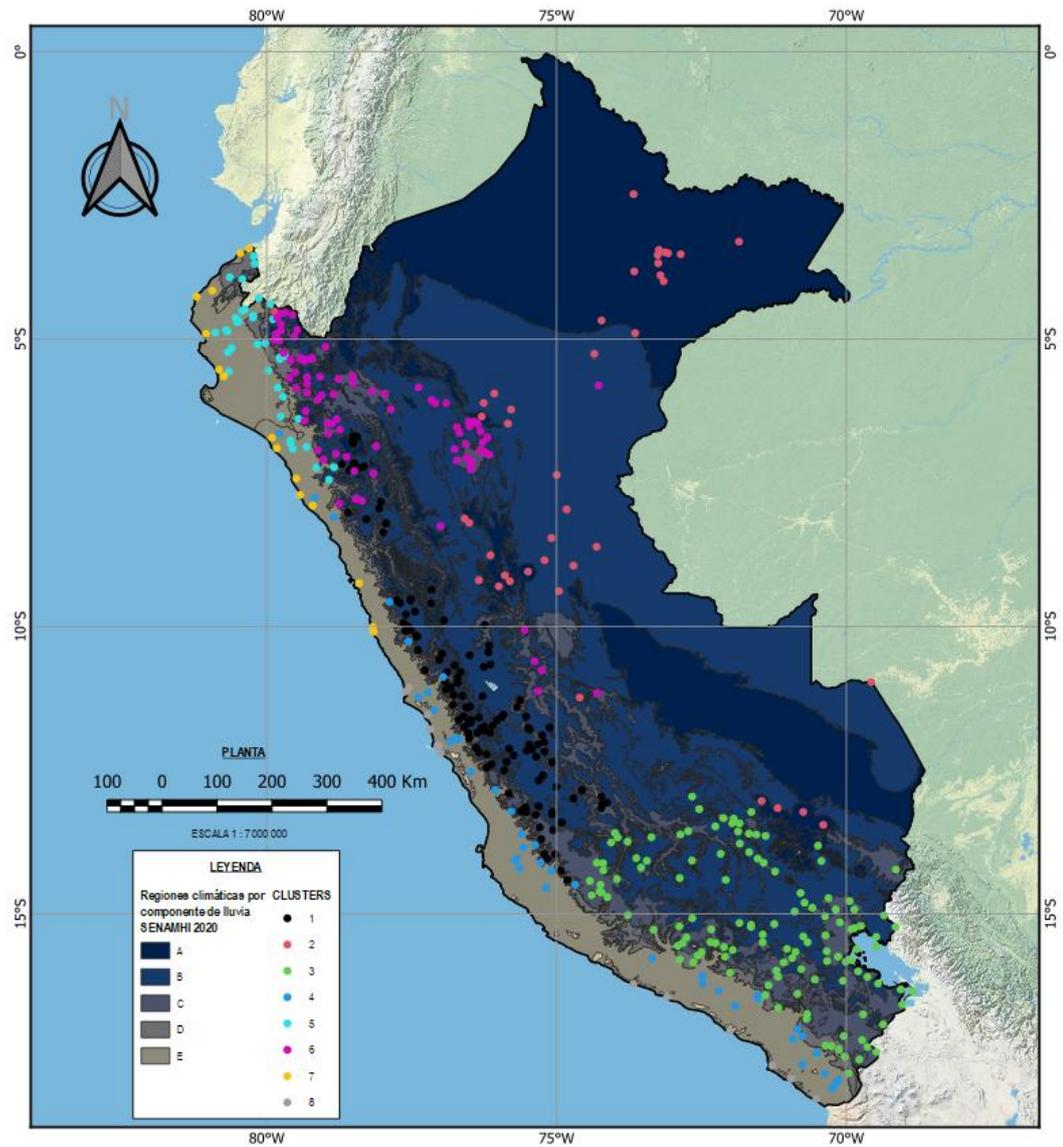
Figura 70 Clusters de estaciones para 24 horas en el Mapa de Clasificación Climática del Perú (2020) componente de llluvias



Fuente: Elaboración propia.

Nota: Las Clasificación Climática del Perú por componente de lluvia fueron adaptadas de Senamhi (2020a)

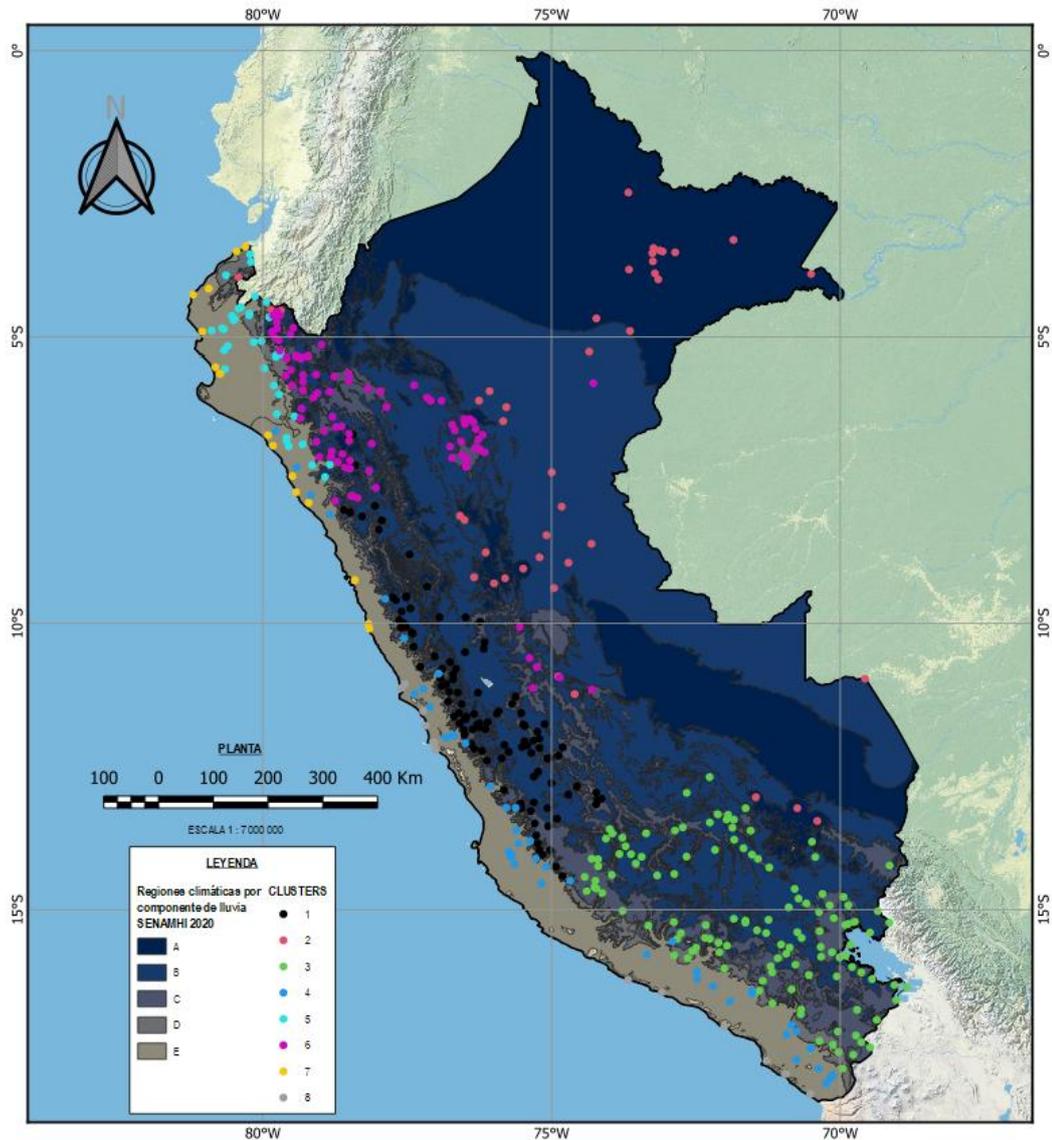
Figura 71 Clusters de estaciones para 48 horas en el Mapa de Clasificación Climática del Perú (2020) componente de llluvias



Fuente: Elaboración propia.

Nota: Las Clasificación Climática del Perú por componente de lluvia fueron adaptadas de Senamhi (2020a)

Figura 72 Clusters de estaciones para 72 horas en el Mapa de Clasificación Climática del Perú (2020) componente de llluvias



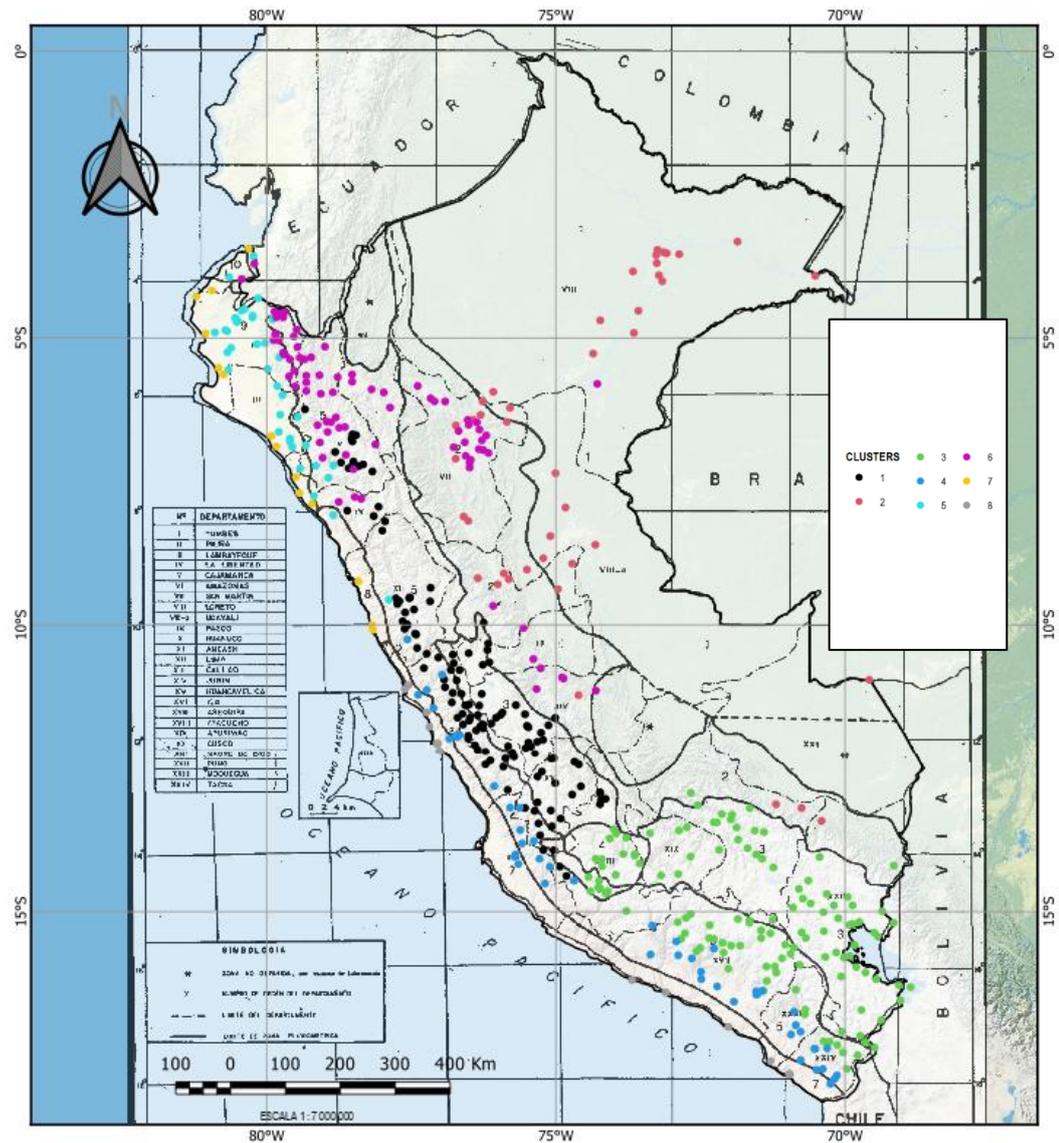
Fuente: Elaboración propia.

Nota: Las Clasificación Climática del Perú por componente de lluvia fueron adaptadas de Senamhi (2020a).

Tal como se muestra en la Figura 73, Figura 74 y Figura 75, al superponer los *clusters* y las zonas pluviométricas del Perú identificadas en “Estudio de la Hidrología del Perú”, mostradas en el mapa del ANEXO A.4, se puede apreciar que existen similitudes en la forma de los *clusters* y las regiones pluviométricas establecidas, incluso algunas coincidencias, tales como que los *clusters* 1 y 5 pertenecientes a la costa norte se encuentran dentro de las zonas 9 y 10; el

clusters 8 pertenecientes a la costa sur se encuentra en la zona 7. De manera general, también se aprecia que la forma de los clusters y las zonas pluviométricas son alargadas y se encuentran siguiendo la dirección de la línea del litoral y de la Cordillera de los Andes.

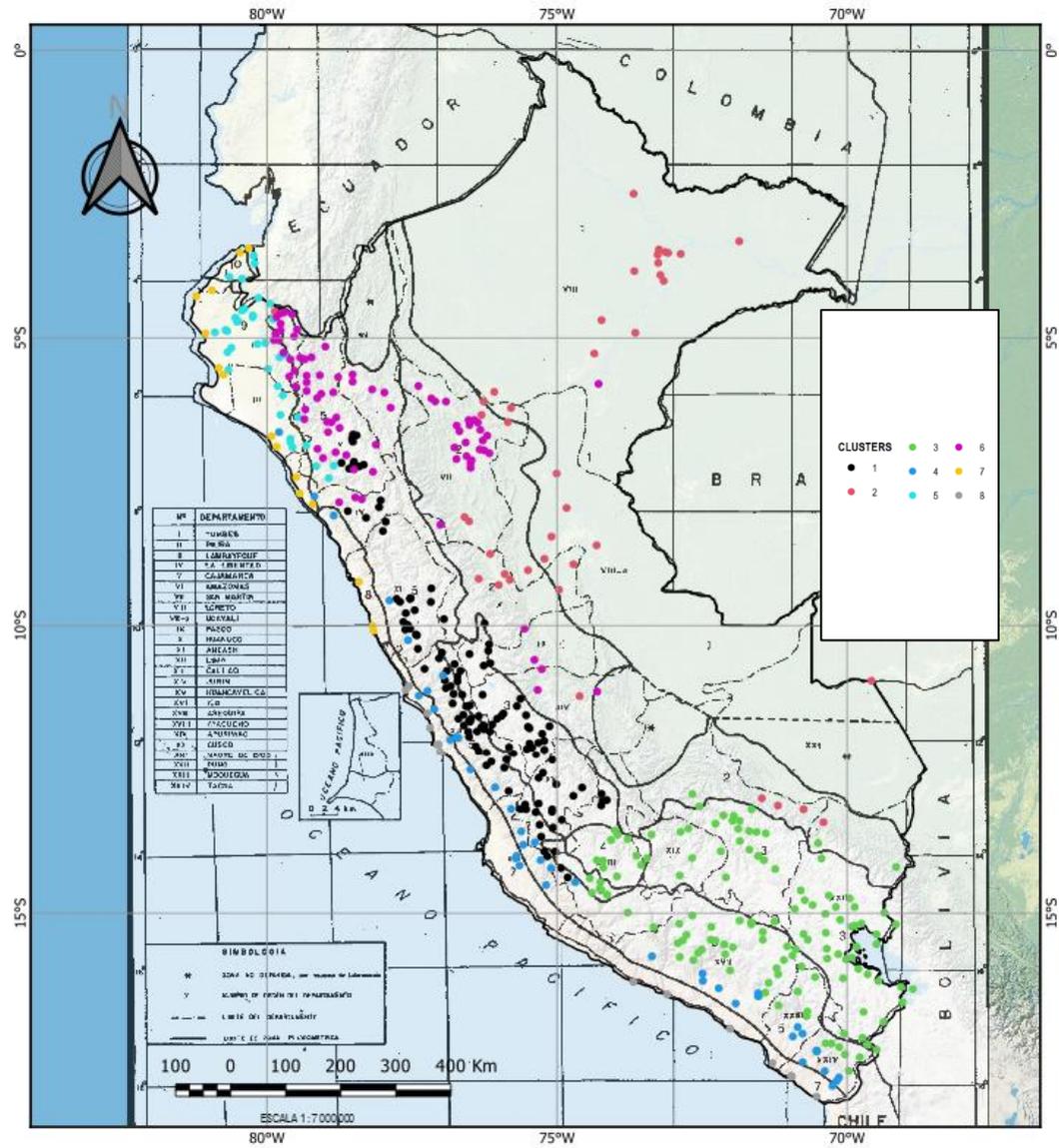
Figura 73 Clusters de estaciones para 24 horas en las zonas pluviométricas del Perú



Fuente: Elaboración propia.

Nota: Las zonas pluviométricas del Perú fueron adaptadas de Instituto Italo - Latino Americano et al., 1982

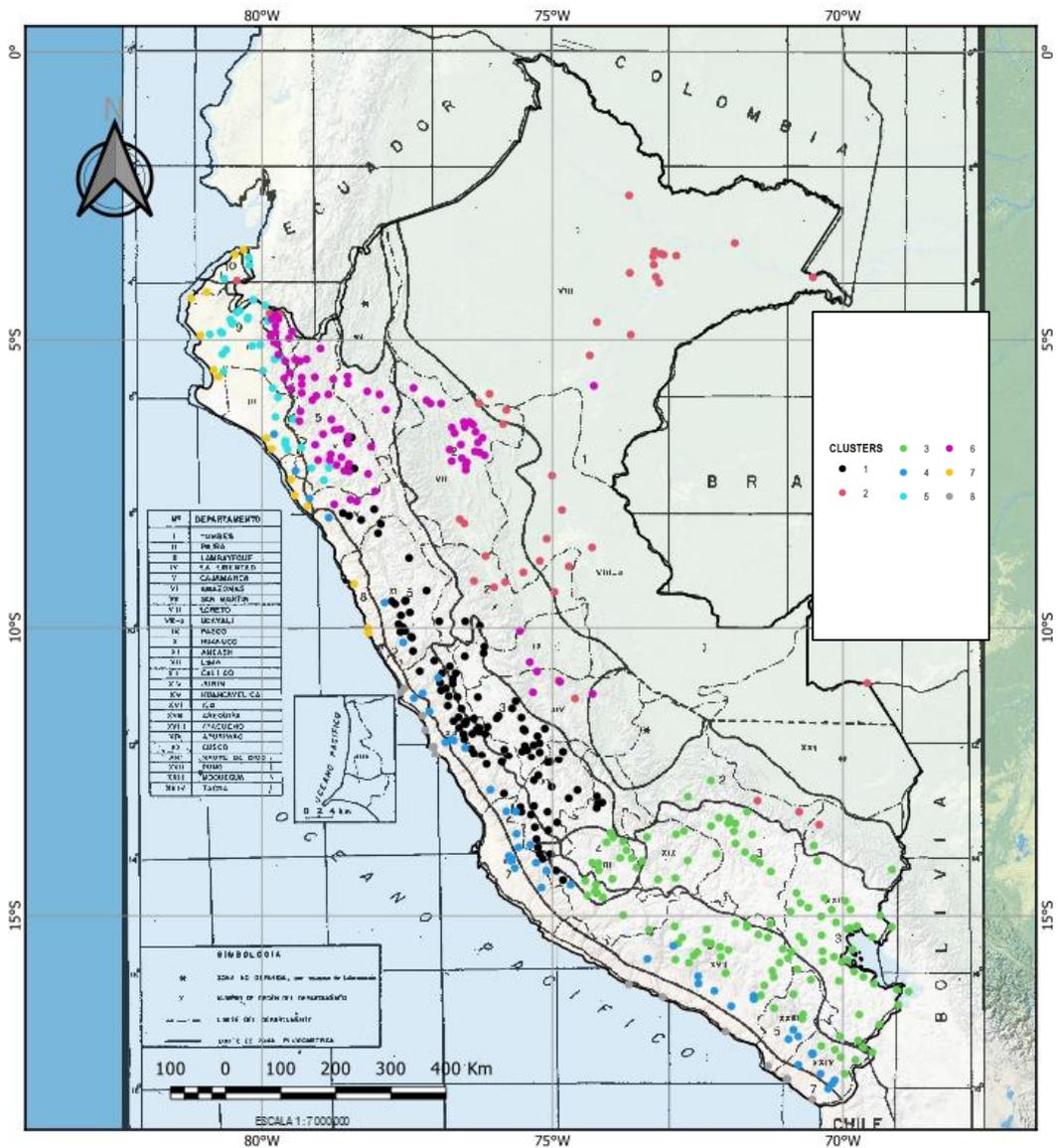
Figura 74 Clusters de estaciones para 48 horas en las zonas pluviométricas del Perú



Fuente: Elaboración propia.

Nota: Las zonas pluviométricas del Perú fueron adaptadas de Instituto Italo - Latino Americano et al., 1982

Figura 75 Clusters de estaciones para 72 horas en las zonas pluviométricas del Perú



Fuente: Elaboración propia.

Nota: Las zonas pluviométricas del Perú fueron adaptadas de Instituto Italo - Latino Americano et al., 1982

6.2 Envoltentes

Para emitir un juicio de comparación entre las envoltentes del Perú propuestas en la presente tesis y las curvas envoltentes determinadas en estudios realizados en otros países (Caamaño et al., s.f.; Casas et al., 2008, 2010; Desa M. & Rakhecha, 2007; Dhar et al., 1981; Hershfield, 1961, 1965; Stowhas, 1983), así como

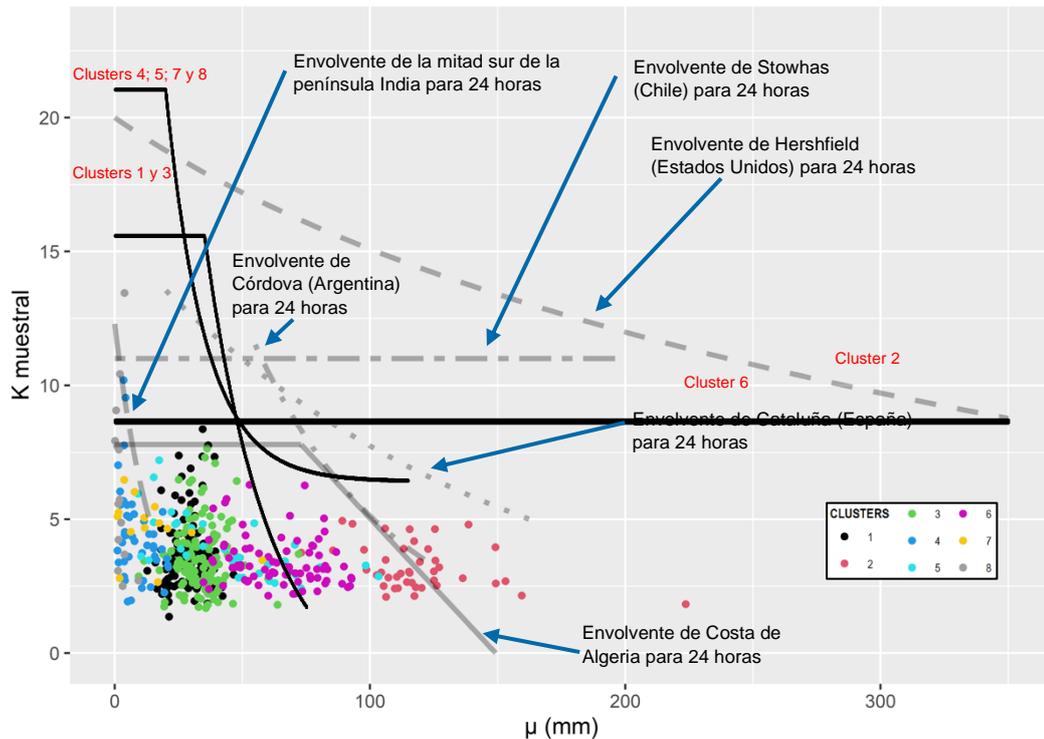
envolventes estimadas a partir de la ecuación de Koutsoyiannis & Xanthopoulos (1999), la cuál es la siguiente:

$$K = 20 - 8.6 * \ln \left(1 + \frac{\mu}{130 \text{ mm}} \right) * \left(\frac{24 \text{ h}}{d} \right)^{0.4} \quad (3)$$

Nota: Donde d representa la duración en horas de la lluvia y μ la media de máximos anuales.

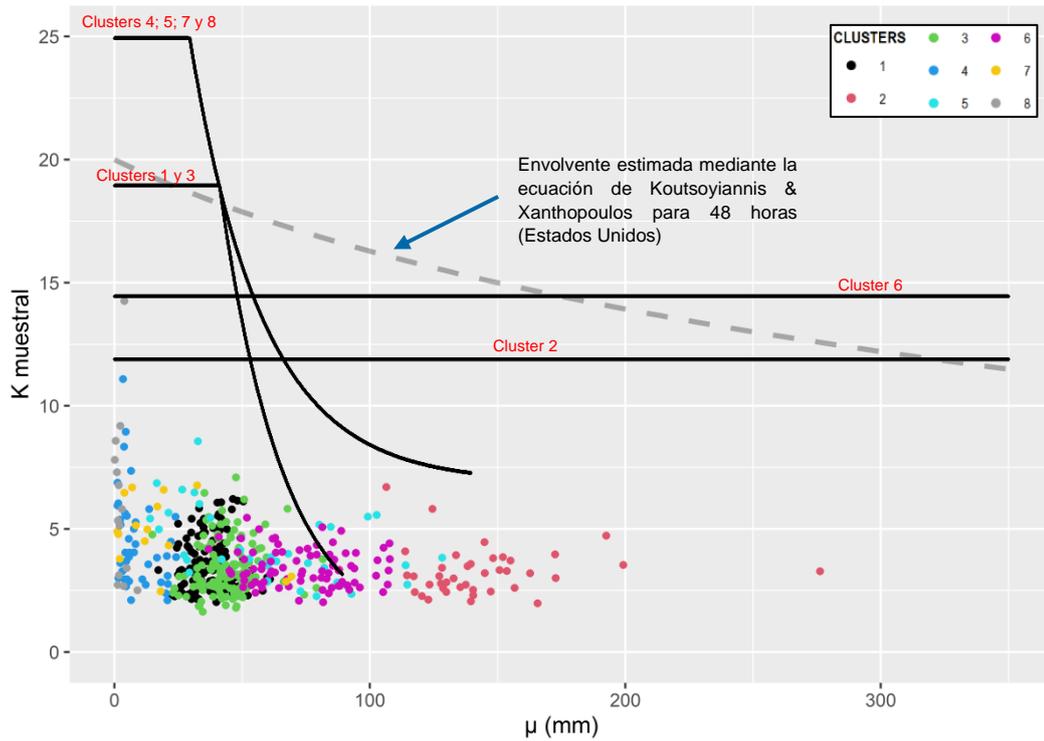
Se elaboraron la Figura 76, Figura 77 y Figura 78, las cuales muestran múltiples curvas envolventes en gráficos de las duraciones de estudio, con ello es posible apreciar que las curvas graficadas presentan una forma cóncava hacia arriba y una terminación asintótica en la dirección de y , propia de una exponencial negativa tal como propusieron Caamaño Nelli & García (1999)

Figura 76 Curvas envolventes de los clusters para 24 horas



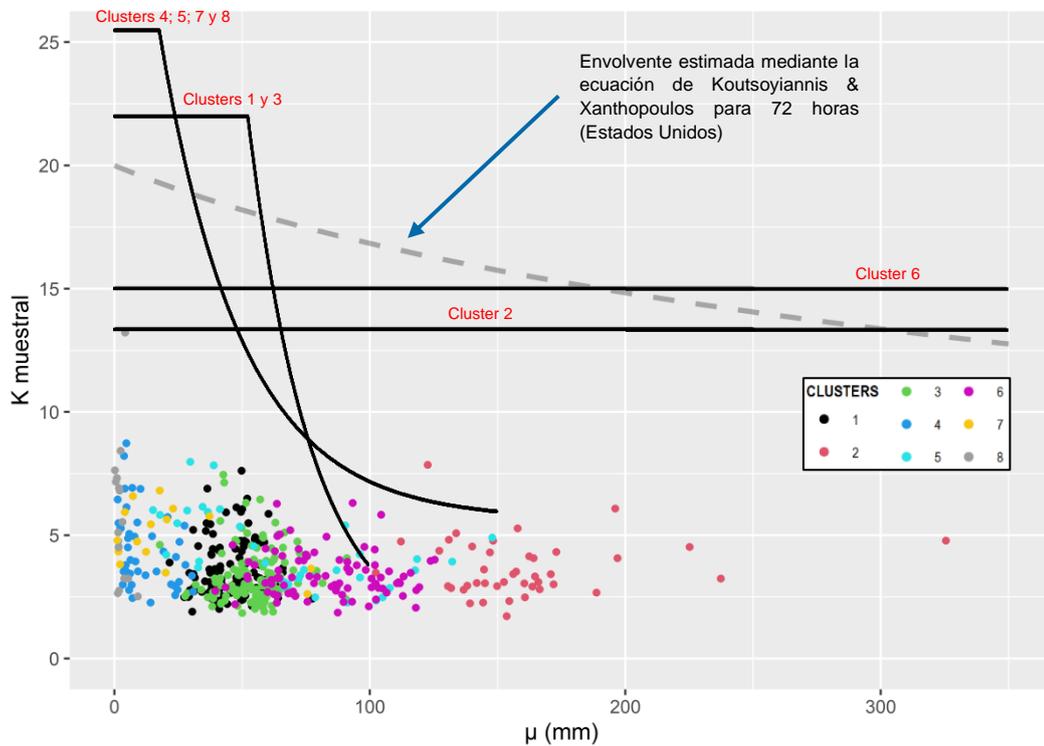
Fuente: Elaboración propia.

Figura 77 Curvas envolventes de los clusters para 48 horas



Fuente: Elaboración propia.

Figura 78 Curvas envolventes de los clusters para 72 horas



Fuente: Elaboración propia.

Se puede apreciar que todas las envolventes peruanas se ubican bastante cerca de la región inferior izquierda del plano representado, tal como es el caso de las envolventes de 24 horas de Córdoba (Argentina), Cataluña (España) e India, los cuales son estudios más recientes a diferencia de los realizados en Estados Unidos (Hershfield, 1961) y en Chile (Stowhas, 1983), esto llevaría a teorizar que en los estudios iniciales de la PMP, se llegó a sobreestimar la misma al ser demasiado poco conservadores en la regulación del valor de K , tal como se especula entre los profesionales de la hidrología.

En los casos de 48 y 72 horas no hay muchas comparaciones para realizar con otras curvas, pero en general se aprecia que las envolventes en conjunto dibujan la silueta de las envolventes estimada mediante la ecuación de Koutsoyiannis & Xanthopoulos, sin embargo, las curvas peruanas al encontrarse diferenciadas en 4 curvas particulares, nos llevaría a una estimación más precisa de la PMP en estas duraciones a diferencia de recurrir a las de Koutsoyiannis & Xanthopoulos.

6.3 Valores puntuales de la PMP en comparación con precipitaciones máximas observadas en el mundo

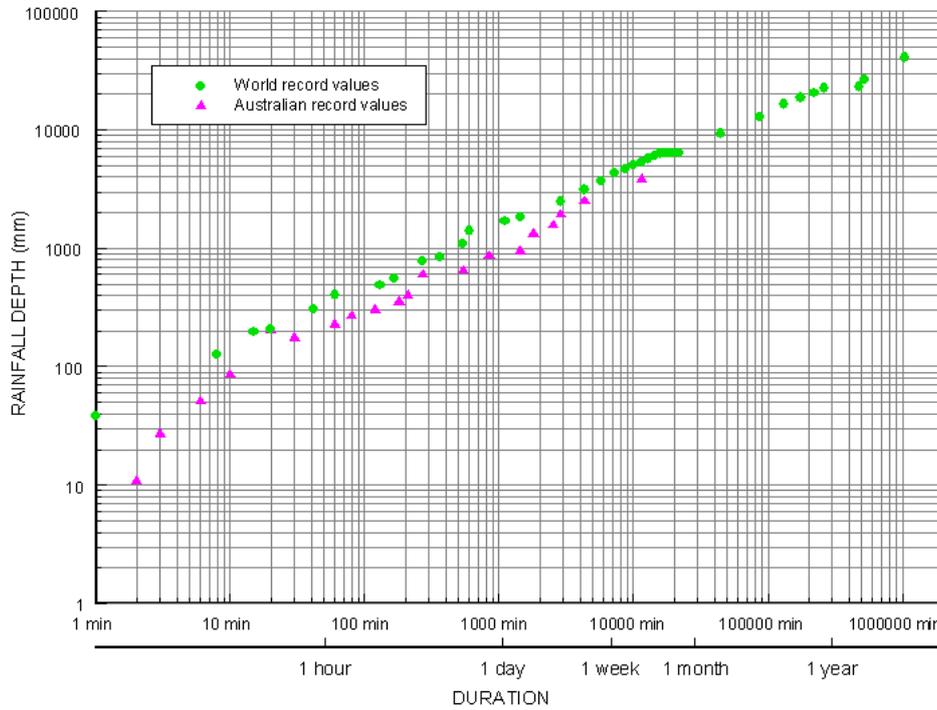
Según el registro mundial de precipitaciones récord (ver Tabla 20 y Figura 79) la precipitación máxima registrada en 24 horas, 2 y 3 días son 1825 mm, 2467 mm y 3030 mm respectivamente, todas en la isla La Reunión; en delimitadas zonas del mundo son menores, como por ejemplo en Australia son 960 mm, 1990 mm y 2610 mm, respectivamente.

Tabla 20 *Lluvias puntuales de mayor intensidad observadas en el mundo*

Duración	Precipitación (mm)	Ubicación	Fecha
24 h	1825	Foc Foc, La Réunion	7-8 enero 1966
2 días	2467	Aurere, La Réunion	7-9 abril 1958
3 días	3130	Aurere, La Réunion	6-9 abril 1958

Fuente: WMO (1994)

Figura 79 Comparación de episodios de precipitaciones puntuales notables en Australia con los más importantes observados en el mundo



Fuente: Australian Government Bureau of Meteorology (s.f.)

Los valores de PMP máximos para el Perú evaluados en la presente tesis son:

- 1333,5 mm en la duración de 24 horas;
- 1645,3 mm en la duración de 48 horas;
- 1934.5 mm en la duración de 72 horas.

Comparando, se afirma que no son valores exagerados y tampoco se encuentran fuera de una escala física creíble.

Por cada *cluster*, en las tablas Tabla 21, Tabla 22 y Tabla 23 se muestran los valores máximos de la PMP para los casos de 24; 48 y 72 horas respectivamente. Los máximos para las 3 duraciones se producen en el *Cluster 2* (ver Mapa D.6, Mapa E.6 y Mapa F.6) en la estación San Gabán (código ANA: 693). En el *Cluster 2* se encuentran también las estaciones Quincemil (código ANA: 6401) y Aguaytía (código ANA: 462), estaciones que han registrado las mayores lluvias en el país, tal que son consideradas como *hotspots* de lluvia (Chavez & Takahashi, 2017; Espinoza et al., 2015).

Por otro lado, respecto a los *clusters* del litoral costero (*clusters* 7 y 8) se puede observar que en el *cluster* 7 las estaciones con mayor PMP son El Salto (código ANA: 135) para el caso de 24 horas y Puerto Pizarro (código ANA: 132) para los casos de 48 y 72 horas, las cuales se encuentran ubicadas en la parte norte de dicho *cluster* donde más afecta el fenómeno El Niño. En el *Cluster* 8 la estación con mayor PMP es Isleta Ancón (código ANA: 599), para las 3 duraciones de análisis, la cual se encuentra en el extremo norte del *cluster*, por esa posición también algo afectada por el fenómeno El Niño. El fenómeno El Niño produce una mayor variabilidad en las series de máximas anuales, lo cual se manifiesta en una mayor PMP.

En el caso del *cluster* 6 se aprecia que en el caso de 24 horas, el valor máximo de PMP es superior a valor máximo en el caso de 48 horas dado que la estación Cabo Inga (código ANA: 139) pertenece al *cluster* 6 en 24 horas pero en 48 horas forma parte del *cluster* 5, lo cual vendría a ser un caso particular ya que si bien la distribución espacial de los *clusters* es muy similar entre las duraciones, algunas estaciones cambian de *clusters* (ver Mapa D.1, Mapa E.1 y F.1) de acuerdo a lo establecido por proceso de *clustering* efectuado.

Tabla 21 Valores máximos de PMP por *cluster* para el caso de 24 horas

<i>Cluster</i>	Estación	PMP en 24 horas (mm)
1	SAN RAFAEL 02	381,07
2	SAN GABAN	1333,49
3	CHUQUIBAMBA	335,43
4	SOCABAYA	459,28
5	ALAMOR	1147,15
6	CABO INGA	910,83
7	EL SALTO	741,70
8	ISLETA ANCON	253,39

Fuente: Elaboración propia.

Tabla 22 Valores máximos de PMP por cluster para el caso de 48 horas

Cluster	Estación	PMP en 48 horas (mm)
1	TAPACOCHA	648,15
2	SAN GABAN	1645,26
3	CHUQUIBAMBA	502,78
4	CHAMANA	639,91
5	PASABAR	1372,50
6	TOCMOCHE	831,87
7	PUERTO PIZARRO	1049,62
8	ISLETA ANCON	268,41

Fuente: Elaboración propia.

Tabla 23 Valores máximos de PMP por cluster para el caso de 72 horas

Cluster	Estación	PMP en 72 horas (mm)
1	PARARIN	841,54
2	SAN GABAN	1934,46
3	CHUQUIBAMBA	586,80
4	CHAMANA	737,93
5	ALAMOR	1691,23
6	TOCMOCHE	1028,87
7	PUERTO PIZARRO	1155,80
8	ISLETA ANCON	277,50

Fuente: Elaboración propia.

6.4 PMP y P24h de 10 000 años

Puesto que el periodo de retorno de una PMP debería encontrarse en el rango de 10^4 y 10^9 años (Nathan & Weinmann, 2019; National Research Council, 1994), dado que en la presente tesis no se han estimado periodos de retorno de las PMP estimadas, para evaluar la validez y tener una referencia de si efectivamente los

periodos de retorno se podrían encontrar en el rango mencionado, se realizó una comparación de los valores más altos de PMP del país calculados en la presente tesis y sus respectivas precipitaciones con periodo de retorno de 10 000 años estimadas a partir de la información del estudio “Desarrollo de curvas pluviométricas Intensidad-Duración-Frecuencia (IDF) en Perú” de SENAMHI (Montesinos et al., 2023) a fin de validar si los valores de PMP eran superiores.

Cabe mencionar que, si bien es cierto, no es lo mismo hablar de una PMP que de una precipitación de un determinado periodo de retorno que se calcula en base a una función de distribución determinada para una estación o región, existe el criterio relevante de que, en códigos de diseño estructurales, para casos exigentes, como los de diseño de depósitos de relaves reemplazaron PMP por la precipitación con 10 000 años de periodo de retorno como equivalente.

Para la comparativa, se logró hacer las estimaciones de precipitaciones de 10 000 años dado que en el estudio mencionado se describen las distribuciones de probabilidad de las diferentes regiones de precipitación máximas determinadas y sus parámetros estadísticos característicos, sin embargo, en el estudio se afirma que los resultados muestran que los RMSE (errores de raíz cuadrada media) son lo suficiente fiables como para permitir que las estimaciones de cuantiles se utilicen con confianza cuando los periodos de retorno son inferiores a 200 años, las estimaciones de periodos de retorno más altos (por ejemplo, 1000 años o más) requieren de registros históricos suficientes para mejorar la confiabilidad en la estimación de cuantiles.

Tabla 24 Estaciones con PMP de 24 horas más altas del Perú, comparación con precipitación de TR=10 000 años según SENAMHI

Latitud (°)	Longitud (°)	Elevación (msnm)	Código ANA	Nombre	Media (mm)	Máximo (mm)	PMP 24 horas (mm)	Precipitación TR=10 000 (SENAMHI, 2023) (mm)
-13.44053	-70.40464	820	6401	SAN GABAN	149.17	433.58	1333.49	713.81
-4.488689	-80.38700	150	152153	ALAMOR	98.37	294.38	1147.15	2074.94
-5.833611	-79.816944	124	262	PASABAR	60.74	203.74	1084.34	1281.20
-4.716944	-80.516944	103	150303	ALIVIADEROS	71.01	226.00	1068.88	1476.74
-4.642778	-80.547222	150	152103	LANCONES	67.76	226.00	1044.58	1429.28
-13.21694	-70.75028	651	693	QUINCÉMIL	223.73	312.11	917.47	1070.59
-3.966667	-80.433333	205	139	CABO INGA	92.66	221.82	910.83	2882.17
-5.533333	-79.983333	230	152107	VIRREY	50.19	175.26	887.55	1058.67

Latitud (°)	Longitud (°)	Elevación (msnm)	Código ANA	Nombre	Media (mm)	Máximo (mm)	PMP 24 horas (mm)	Precipitación TR=10 000 (SENAMHI, 2023) (mm)
-4.516667	-80.433333	150	152202	ARDILLA SOLANA BAJA	73.12	177.95	859.64	1544.02
-3.516944	-73.083611	105	150204	TIMICURILLO	159.50	266.68	788.71	434.63

Fuente: Elaboración propia.

De la Tabla 24 se aprecia que, en base a la experiencia se puede afirmar que en la mayoría de casos los valores de precipitación de 10 000 años de periodo de retorno, despejados a partir del trabajo de SENAMHI son bastante elevados y no guardan una congruencia con las medias y los máximos registrados, justamente estos casos corresponden a estaciones ubicadas en la costa norte del Perú en donde se sabe, afecta el fenómeno del niño, pero resulta difícil afirmar que las precipitaciones de 10 000 años en esta ubicación puedan ser tan elevadas y superar por tanto a las PMP estimadas.

En los casos de San Gaban y Timicurillo donde, los valores de PMP estimados en la presente tesis, superan a las precipitaciones de 10 000 años, asimismo, en el caso de Quincemil, se observa una relativa cercanía entre la PMP y la precipitación de 10 000 años, esto sucede justamente en las estaciones que como se mencionó anteriormente se encuentran en un *hotspot* de lluvias en el Perú.

Puesto que, como se mencionó previamente, lo trabajado en “Desarrollo de curvas pluviométricas Intensidad-Duración-Frecuencia (IDF) en Perú” no es información que deba utilizarse para estimar precipitaciones sobre los 200 años de periodo de retorno, por lo que lo estimado y analizado es únicamente con el fin de tener una referencia.

Dado que, las estaciones las cuales se sospechaba que tenían una PMP más elevada por su ubicación en el *hotspot* del Perú, cumplen con tener una PMP mayor a las precipitaciones de 10 000 años, se afirma que la metodología utilizada en general ha sido correcta y ha permitido llegar a estimar valores con justificación estadística fuerte como para considerarlos correctos.

Conclusiones

A la culminación de la presente tesis se concluye que:

Se generó una base de datos depurada de precipitaciones máximas anuales de 1; 2 y 3 días, así como de 24; 48 y 72 horas, la cual abarca toda la información disponible del SENAMHI, ANA y producto PISCO Prec v2p1stable daily.

Se minimizó la posibilidad de que el proceso de descarte de *outliers* elimine verdaderas precipitaciones máximas, debido a que, por una parte, se seleccionaron los datos mejor calificados por continuidad y completitud de cada fuente disponible de información pluviométrica en el Perú (Autoridad Nacional del Agua, s.f.; International Research Institute for Climate and Society, 2019; SENAMHI, s.f.); y por otro lado realizando los análisis de consistencia (visual, detección de *outliers*, estacionariedad, homogeneidad e independencia serial) empleando índices de significancia que van de 0,5% a 1%.

Respecto a los gráficos del $K_{muestral}$ en función de la media, para el caso de 24 horas (Figura 34), presentan en su frontera superior un comportamiento en general similar al de la envolventes de 24 horas de Estados Unidos, Cataluña (España) y Argentina (Caamaño et al., s.f.; Casas et al., 2008; Hershfield, 1961; Koutsoyiannis & Xanthopoulos, 1999); teniendo una forma cóncava hacia arriba y una terminación asintótica en la dirección de y . Sin embargo, no se puede notar una semejanza notable con la envolvente generada siguiendo el razonamiento de Stowhas. Cabe resaltar que, entre los casos mencionados, la frontera superior de la nube de puntos del caso peruano es una de las que se ubica más abajo y más a la izquierda en el primer cuadrante del plano cartesiano en comparación a los demás.

En el cálculo de la PMP en el Perú mediante la fórmula de Hershfield, el Cv representa una variable mucho más robusta que la media (μ), ya que en general se distribuye sin saltos abruptos en el territorio peruano, teniendo un valor en un rango de valores comprendido entre 0.8 y 1.7, en la franja costera de los 60 km más cercanos al mar y valores que van de 0 a 0.8 en el resto del territorio peruano, dando lugar a considerarla dentro de las variables utilizadas para la formación de *clusters* y para la regresión de valores de la PMP.

Se definieron en total 8 *clusters* hidrológicamente homogéneos para el análisis de la PMP en el Perú a partir de diversas medidas estadísticas de sus series de

máximas de precipitación, los cuales fueron: el coeficiente de variación, valor máximo y media; asimismo se consideró para la formación de los *clusters* la ubicación espacial de las estaciones analizadas (latitud, longitud, elevación) y la región de lluvia a la que pertenece cada una.

Se elaboraron envolventes del factor K para 24; 48 y 72 horas para cada *cluster* siguiendo el procedimiento de Stowhas (1983), descrito en su estudio de la PMP de Chile, basándose en un principio de doble maximización del factor K , es decir que con la selección de un valor de C_v lo más desfavorable posible, el valor del máximo adimensional (P_m) se llega a maximizar doblemente con $K = \frac{P_m}{C_v}$ y en consecuencia se optimiza lo más posible el valor de K .

Se evaluó el factor K de las estaciones pluviométricas seleccionadas para duraciones de 24; 48 y 72 horas utilizando las envolventes elaboradas.

Las envolventes de cada *cluster* (ver ANEXO C) son en general distintas, sin embargo, es observable que se pudo simplificar el conjunto de envolventes de los 8 *clusters* al agrupar determinados *clusters* por su comportamiento y ubicación y así definir envolventes que sean aplicables a más de un *cluster* a la vez en algunos casos (ver Figura 76, Figura 77 y Figura 78).

Existen regiones muy específicas del Perú donde el comportamiento de las lluvias es muy particular, siendo el primer caso a mencionar el territorio peruano muy próximo al litoral (menos de 15 km de distancia al mar), el cual muestra una alta variabilidad en los registros de precipitaciones máximas anuales con coeficientes de variación altos, los cuales son superiores a 0,8, pudiéndose encontrar años en los que se ha registrado 0 mm como máximo anual, es decir, años completamente secos. El otro caso corresponde al norte de Puno y este de Cusco, además del noreste de Ucayali, en donde se registran valores altísimos de lluvias máximas, pudiéndose evidenciar este comportamiento en las series de máximas precipitaciones anuales y valores de PMP de las estaciones Quincemil, Aguaytía y San Gabán, estas estaciones se ubican en los *hotspots* de lluvia del Perú.

El uso en el Perú de envolventes del factor K desarrolladas para otros países u otras zonas geográficas llevan a la sobreestimación de la PMP (p.e. la envolvente de Herhsfield (1965) para el caso de 24 horas, la cual es comúnmente usada y no es del todo parecida a la envolvente de Cataluña (España) para 24 horas), o a la subestimación de la misma (p.e. el uso de la envolvente de Córdoba (Argentina)

que implícitamente está acotada hasta 175 mm para 24 horas), lo que llevaría en consecuencia a cálculos erróneos en el dimensionamiento de estructuras de protección y control tales como presas o aliviaderos, las cuales son de suma importancia puesto que de ellas depende la seguridad de muchas de personas en ciertos casos.

Se evaluó la PMP de 24; 48 y 72 horas en el Perú de manera puntual en las ubicaciones de las estaciones que cumplieron con los análisis establecidos y se extendieron los resultados por medio de una regresión espacial con el método SVM de tipo radial obteniendo 3 *rasters* de resolución $0,01^\circ \times 0,01^\circ$ que abarcan gran parte del territorio nacional, para que el público usuario pueda extraer directamente los valores de PMP en las coordenadas y duración de su interés. No fue posible abarcar todo el territorio del Perú, como la selva sureste y selva norte por la falta de estaciones con suficientes registros que generen puntos con PMP evaluada, necesarios para la aplicación del método SVM de tipo radial.

Las estimaciones de la PMP no son definitivas puesto que el mismo análisis puede repetirse y mejorarse modificando los parámetros de la regresión SVM, incorporando a futuro más datos (actualizando el proyecto), definiendo de otra manera los agrupamientos de estaciones y delineando de otra forma las envolventes empleando nuevos criterios, haciendo ello que varíen los valores de la PMP. En consecuencia, se afirma que los resultados no son únicos.

Recomendaciones

En base a los resultados y conclusiones de la presente tesis, se comparten las siguientes recomendaciones para investigaciones posteriores que sean afines:

Al trabajar con data cruda de precipitación diaria proveniente de los repositorios del SENAMHI o la ANA o del sustento del producto PISCO es importante realizar previamente un análisis de atípicos (*outliers*) a fin de identificar y descartar datos que hayan sido registrados erróneamente. Se debe identificar los fuera de escala física creíble que pueden alterar las estadísticas de la serie a la que pertenecen, aunque haya riesgo de descartar registros atípicos correctos.

No se recomienda utilizar datos completados y depurados sin conocer bien el tratamiento que se les hizo, como el caso de la data diaria procesada para la elaboración del producto PISCO (Aybar et al., 2020), que abarca datos diarios de 441 estaciones del Perú entre 01/01/1981 y 31/07/2018, que fueron revisados y depurados de *outliers* con el criterio de retirar los de periodos de retorno superiores a los 200 años, ello los invalida para estudios de precipitaciones máximas y PMP, los cuales buscan precipitaciones con altos periodos de retorno.

Se puede emplear variables climatológicas adicionales y/o factores obtenidos a partir de la relación de varias de ellas al realizar una regresión SVM de tipo radial en investigaciones similares, ello para darle una mayor robustez a los resultados del método.

Es posible realizar investigaciones de PMP en un área de estudio más pequeña siguiendo la misma metodología propuesta en la presente tesis, en particular resultaría interesante que se realice en una zona de *hotspots* de precipitaciones con el fin de una evaluación a mayor detalle.

Se estima posible aplicar la Regresión con Vectores de Soporte para estimar valores extremos de distintas variables climatológicas y en otros casos de estudio de la PMP a nivel nacional o regional donde se cuente con información puntual y se requiera obtener una representación gráfica del comportamiento de la variable estudiada.

En cuanto a las envolventes del factor K , se puede realizar una delimitación especial de envolventes para las zonas más afectadas del fenómeno El Niño (Costa norte del Perú), con segregación bipoblacional, disgregando en años con

y sin efecto del fenómeno El Niño, desarrollando envolventes distintas para cada caso.

En el mismo sentido de la recomendación anterior, se puede realizar una delimitación especial de envolventes del factor K para estaciones de la costa peruana cercanas al litoral, fuera de la zona de afectación notable de El Niño, o zonas definidas como muy áridas que tengan un coeficientes de variación de precipitaciones elevado, de igual manera aplicando una segregación bipoblacional, disgregando en años con cero precipitación o precipitación despreciable y años con precipitación apreciable, desarrollando de envolventes distintas para cada caso.

Una línea de desarrollo para estudios posteriores es incluir data pluviométrica de países vecinos, en especial de Ecuador, Brasil y Colombia para compensar la falta de estaciones peruanas en la selva norte y selva sur del país, para que de esa manera mejorar o implementar la interpolación espacial de la PMP en esas zonas.

Otra línea de desarrollo para estudios posteriores es realizar la evaluación de la PMP pero con la consideración de distintas longitudes de registro mínimo por estación, para comparar los resultados y tendencias en la definición de envolventes para la evaluación de la PMP.

Se recomienda que la información generada se emplee como una herramienta de apoyo y consulta referencial para el diseño de obras hidráulicas de protección y control; de esta manera, se conseguirán diseños más seguros, y por tanto se minimizará la probabilidad de falla o excedencia de alguna de estas obras, lo que a su vez, reduciría la cantidad de inundaciones, las cuales tienen consecuencias nefastas, entre las más resaltantes son: la pérdida de numerosas vidas humanas, dañar otras obras civiles y dañar el medio ambiente.

Referencias bibliográficas

- Australian Government Bureau of Meteorology. (s.f.). *Australia's Record Rainfall*.
<http://www.bom.gov.au/water/designRainfalls/rainfallEvents/ausRecordRainfall.all.shtml>
- Autoridad Nacional del Agua. (s.f.). *Observatorio del Agua Nivel Nacional*.
Retrieved June 27, 2022, from <https://snirh.ana.gob.pe/observatorioSNIRH/>
- Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., & Felipe-Obando, O. (2020). Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, 65(5), 770–785.
<https://doi.org/10.1080/02626667.2019.1649411>
- Baker, F. B., & Hubert, L. J. (1975). Measuring the Power of Hierarchical Cluster Analysis. *Journal of the American Statistical Association*, 70(349), 31.
<https://doi.org/10.2307/2285371>
- Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification*.
- Ball, J., Babister, M., Retallick, M., & Weinmann, E. (2019). *Australian Rainfall and Runoff: A Guide to Flood Estimation*.
- Baraquet, M. M. (2018). Efectos de la variabilidad climática en la predicción de lluvias para diseño en la provincia de Córdoba [Tesis de maestría, Universidad Nacional De Córdoba]. Repositorio Digital UNC.
<http://hdl.handle.net/11086/11138>.
- Batout, S., Houichi, L., & Marouf, N. (2022). Influence of the envelope curve on the estimate of probable maximum precipitation (PMP) in the coastal region of Algeria. *Modeling Earth Systems and Environment*, 8(2), 2083–2093.
<https://doi.org/10.1007/s40808-021-01209-1>
- Bazzano, F. (2019). *Predicción de lluvias máximas para diseño hidrológico, desarrollo experimental en la provincia de Tucumán*.
- Beale, E. M. L. (1969). *Euclidean Cluster Analysis*. Scientific Control Systems Ltd.
- Benson, M. A. (1973). Thoughts on the design of design floods. *Floods and*

- droughts. *Proceedings of the 2nd International Symposium in Hydrology*, 27–33.
- Bezdek, J. C., & Pal, N. R. (1998). Some new indexes of cluster validity. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 28(3), 301–315. <https://doi.org/10.1109/3477.678624>
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). Training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual ACM Workshop on Computational Learning Theory*, 144–152. <https://doi.org/10.1145/130385.130401>
- Caamaño, G. E., Carlos, N., García, M., & Dasso, C. M. (s.f.). *Precipitación máxima probable de lapso variable para Córdoba, Argentina*.
- Caamaño Nelli, G., & García, C. (1999). Relación Intensidad-Duración-Recurrencia de Lluvias Máximas: Enfoque a través del Factor de Frecuencia, Caso Lognormal. *Ingeniería Hidráulica En México*, 37–44.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>
- Campos Aranda, D. (1998). *Procesos del ciclo hidrológico*. Universidad Autónoma de San Luis Potosí. <https://repositorioinstitucional.uaslp.mx/xmlui/handle/i/3331>
- Campos Aranda, D. (2014). Predicciones extremas de lluvia en 24 horas en el estado de Zacatecas, México. *Tecnología y Ciencias Del Agua*, 5.
- Casas, M. C., Rodríguez, R., Nieto, R., & Redaño, A. (2008). The estimation of probable maximum precipitation: The case of Catalonia. *Annals of the New York Academy of Sciences*, 1146, 291–302. <https://doi.org/10.1196/annals.1446.003>
- Casas, M. C., Rodríguez, R., Prohom, M., Gázquez, A., & Redaño, A. (2010). Estimation of the probable maximum precipitation in Barcelona (Spain). *International Journal of Climatology*. <https://doi.org/10.1002/joc.2149>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). Nbclust: An R package for determining the relevant number of clusters in a data set. *Journal of*

- Statistical Software*, 61(6), 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Chavez, S. P., & Takahashi, K. (2017). Orographic rainfall hot spots in the Andes-Amazon transition according to the TRMM precipitation radar and in situ data. *Journal of Geophysical Research: Atmospheres*, 122(11), 5870–5882. <https://doi.org/10.1002/2016JD026282>
- Chow, V. Te. (1951). A general formula for hydrologic frequency analysis. In *American Geophysical Union* (Vol. 32, Issue 2).
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Desa M., M. N., & Rakhecha, P. R. (2007). Probable maximum precipitation for 24-h duration over an equatorial region: Part 2-Johor, Malaysia. *Atmospheric Research*, 84(1), 84–90. <https://doi.org/10.1016/j.atmosres.2006.06.005>
- Dhar, O. N., Kulkarni, A. K., & Rakhechia, P. R. (1981). Probable maximum point rainfall estimation for the southern half of the Indian peninsula. In *Prec. Indian Acad. Sci. (Earth Planet. Sc.)* (Vol. 90).
- Duda, R. O., & Hart, P. E. (1973). *Pattern Classification and Scene Analysis*. John Wiley & Sons.
- Dunn†, J. C. (1974). Well-Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics*, 4(1), 95–104. <https://doi.org/10.1080/01969727408546059>
- Edwards, A. W. F., & Cavalli-Sforza, L. L. (1965). A Method for Cluster Analysis. *Biometrics*, 21(2), 362. <https://doi.org/10.2307/2528096>
- Endara, S. (2016). Ciclos Horarios De Precipitación En El Perú Utilizando Información Satelital Dirección. *Servicio Nacional de Meteorología e Hidrología*, 53. <http://repositorio.senamhi.gob.pe/handle/20.500.12542/112>
- Espinoza, J. C., Chavez, S., Ronchail, J., Junquas, C., Takahashi, K., & Lavado, W. (2015). Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric

- circulation. *Water Resources Research*, 51(5), 3459–3475.
<https://doi.org/10.1002/2014WR016273>
- Frey, T., & Van Groenewoud, H. (1972). A cluster analysis of the D2 matrix of white spruce stands in Saskatchewan based on the maximum-minimum principle. *The Journal of Ecology*, 873–886.
- Friedman, H. P., & Rubin, J. (1967). On Some Invariant Criteria for Grouping Data. *Journal of the American Statistical Association*, 62(320), 1159.
<https://doi.org/10.2307/2283767>
- Fukunaga, K., & Koontz, W. L. G. (1970). A Criterion and an Algorithm for Grouping Data. *IEEE Transactions on Computers*, C-19(10), 917–923.
<https://doi.org/10.1109/T-C.1970.222799>
- Gordon, A. (1999). *Classification, 2nd Edition* (Vol. 16). Chapman and Hall/CRC.
<https://doi.org/10.1201/9781584888536>
- Grubbs, F. E. (1950). Sample Criteria for Testing Outlying Observations. *The Annals of Mathematical Statistics*, 21(1), 27–58.
<https://doi.org/10.1214/aoms/1177729885>
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, 17(2), 107–145.
<https://doi.org/10.1023/A:1012801612483>
- Halkidi, M., & Vazirgiannis, M. (2001). Clustering validity assessment: finding the optimal partitioning of a data set. *Proceedings 2001 IEEE International Conference on Data Mining*, 187–194.
<https://doi.org/10.1109/ICDM.2001.989517>
- Hansen, E. M., Schreiner, L. C., & Miller, J. F. (1982). *Application of probable maximum precipitation estimates-United States East of 105th Meridian*.
- Hartigan, J. A. (1975). *Clustering Algorithms* (J. W. & Sons (ed.)). John Wiley & Sons. [https://people.inf.elte.hu/fekete/algorithmusok_msc/klaszterezes/John A. Hartigan-Clustering Algorithms-John Wiley & Sons \(1975\).pdf](https://people.inf.elte.hu/fekete/algorithmusok_msc/klaszterezes/John%20A.%20Hartigan-Clustering%20Algorithms-John%20Wiley%20&%20Sons%20(1975).pdf)
- Hershfield, D. M. (1961). Estimating the Probable Maximum Precipitation. *Journal of the Hydraulics Division*, 87(5), 99–116.
<https://doi.org/10.1061/JYCEAJ.0000651>

- Hershfield, D. M. (1962). Extreme Rainfall Relationships. *Journal of the Hydraulics Division*, 88(6), 73–92. <https://doi.org/10.1061/JYCEAJ.0000818>
- Hershfield, D. M. (1965). Method for Estimating Probable Maximum Rainfall. *American Water Works Association*, 57(8).
<http://www.jstor.org/stable/41264521>
- Hill, R. S. (1980). A Stopping Rule for Partitioning Dendrograms. *Botanical Gazette*, 141(3), 321–324. <http://www.jstor.org/stable/2474421>
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency Analysis*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511529443>
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218. <https://doi.org/10.1007/BF01908075>
- Hubert, L. J., & Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychological Bulletin*, 83(6), 1072–1080. <https://doi.org/10.1037/0033-2909.83.6.1072>
- International Council on Mining and Metals, United Nations Environment Programme, & Principles for Responsible Investment. (2020). *Global Industry Standard on Tailings Management*.
- Instituto Italo - Latino Americano, Servicio Nacional de Meteorología e Hidrología del Perú, & Universidad Nacional de Ingeniería. (1982). *Estudio de la Hidrología del Perú*. <https://hdl.handle.net/20.500.12542/2034>
- International Research Institute for Climate and Society. (2019). *SENAMHI HSR PISCO: Peruvian Interpolated data*.
<https://iridl.ldeo.columbia.edu/SOURCES/.SENAMHI/.HSR/.PISCO/index.html?Set-Language=es>
- Johnson, K. A., & Smithers, J. C. (2019). Methods for the estimation of extreme rainfall events. *Water SA*, 45 (3), 501–512. South African Water Research Commission. <https://doi.org/10.17159/wsa/2019.v45.i3.6747>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data*. En *Wiley series in probability and statistics*. <https://doi.org/10.1002/9780470316801>
- Koutsoyiannis, D., & Xanthopoulos, T. (1999). *Τεχνική Υδρολογία (Engineering Hydrology)* (pp. 161). <https://doi.org/10.13140/RG.2.1.4856.0888>

- Kraemer, H. C. (1982). Biserial Correlation. En *Encyclopaedia of Statistical Sciences*, 1, 276–279.
- Krzanowski, W. J., & Lai, Y. T. (1988). A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. *Biometrics*, 44(1), 23. <https://doi.org/10.2307/2531893>
- Lagos Zuñiga, M. Á. (2012). *Impacto del cambio climático en eventos extremos y análisis de la vulnerabilidad de algunas obras hidráulicas en Chile*.
- Lebart, L., Morineau, A., & Piron, M. (1995). *Statistique exploratoire multidimensionnelle*. Paris.
- Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. En *Biometrika*, 65(2), 297–303. <https://doi.org/10.1093/biomet/65.2.297>
- López, P., Seoane, R., Fornero, L. A., Rodríguez, J., & Nuñez, M. (2007). *Análisis de tendencia en series autocorrelacionadas*. 8. <https://imd.uncuyo.edu.ar/upload/07.doc.pdf>
- McClain, J. O., & Rao, V. R. (1975). CLUSTISZ: A Program to Test for the Quality of Clustering of a Set of Objects. *Journal of Marketing Research*, 12(4), 456–460. <http://www.jstor.org/stable/3151097>
- Mélice, L. C., & Reason, C. J. C. (2007). Return period of extreme rainfall at George, South Africa. *South African Journal of Science*, 103, 499–501. http://www.scielo.org.za/scielo.php?script=sci_arttext&pid=S0038-23532007000600014&nrm=iso
- Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45(3), 325–342. <https://doi.org/10.1007/BF02293907>
- Milligan, G. W. (1981). A monte carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika*, 46(2), 187–199. <https://doi.org/10.1007/BF02293899>
- Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159–179. <https://doi.org/10.1007/BF02294245>

- Montesinos, C., Lavado, W., Quijada, N., Gutiérrez, L., & Felipe, O. (2023). *Desarrollo de curvas pluviométricas Intensidad-Duración-Frecuencia (IDF) en Perú*. <https://repositorio.senamhi.gob.pe/handle/20.500.12542/2825>
- National Geospatial-Intelligence Agency & National Aeronautics and Space Administration. (2013). *Shuttle Radar Topography Mission (SRTM) Global. Distributed by OpenTopography*. <https://doi.org/10.5069/G9445JDF>
- Nathan, R., & Weinmann, P. E. (2019). Estimation of Very Rare to Extreme Floods (Book 8). In *Australian Rainfall and Runoff: A Guide to Flood Estimation*.
- National Research Council. (1994). *Estimating Bounds on Extreme Precipitation Events*. National Academies Press. <https://doi.org/10.17226/9195>
- Orloci, L. (1967). An Agglomerative Method for Classification of Plant Communities. *The Journal of Ecology*, 55(1), 193. <https://doi.org/10.2307/2257725>
- Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, 28(2), 126. <https://doi.org/10.2307/2346729>
- Ratkowsky, D. A., & Lance, G. N. (1978). Criterion for determining the number of groups in a classification. *Australian Computer Journal*, 10, 115–117.
- Rezacova, D., Pesice, P., & Sokol, Z. (2005). An estimation of the probable maximum precipitation for river basins in the Czech Republic. *Atmospheric Research*, 77(1-4), 407–421. <https://doi.org/10.1016/j.atmosres.2004.10.011>
- Rohlf, F. J. (1974). Methods of Comparing Classifications. *Annual Review of Ecology and Systematics*, 5(1), 101–113. <https://doi.org/10.1146/annurev.es.05.110174.000533>
- Romesburg, C. (2004). *Cluster Analysis For Researchers*. Lulu Press. https://books.google.com.pe/books?id=ZuIPv7OKm10C&printsec=frontcover&dq=Cluster+Analysis+for+Researchers&cd=1&redir_esc=y&hl=es#v=onepage&q&f=false
- Salas, J. D., Anderson, M. L., & Papalexiou, S. M. (2020). PMP and Climate Variability and Change: A Review. *Journal of Hydrologic Engineering*, 25(12). [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0002003](https://doi.org/10.1061/(ASCE)HE.1943-5584.0002003)
- Sarle, W. S. (1983). *SAS technical report A-108, Cubic clustering criterion*.

- Scott, A. J., & Symons, M. J. (1971). Clustering Methods Based on Likelihood Ratio Criteria. *Biometrics*, 27(2), 387. <https://doi.org/10.2307/2529003>
- Sen, P. K. (1968). Estimates of the Regression Coefficient Based on Kendall's Tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.1080/01621459.1968.10480934>
- Servicio Nacional de Meteorología e Hidrología del Perú. (s.f.). *Descarga de datos Meteorológicos a nivel nacional*. <https://www.senamhi.gob.pe/?p=descarga-datos-hidrometeorologicos>
- Servicio Nacional de Meteorología e Hidrología del Perú. (2017). *Regionalización de Precipitaciones Máximas en el Perú*. <https://www.senamhi.gob.pe/load/file/01402SENA-76.pdf>
- Servicio Nacional de Meteorología e Hidrología del Perú. (2020a). *Mapa de clasificación Climática Nacional*. <https://idsep.senamhi.gob.pe/geonetwork/srv/spa/catalog.search#/metadatas/9f18b911-64af-4e6b-bbef-272bb20195e4>
- Servicio Nacional de Meteorología e Hidrología del Perú. (2020b). *Regiones Homogéneas de Precipitación Máxima*. <http://catalogo.geoidep.gob.pe/metadata/srv/spa/catalog.search#/metadata/9aa01480-9f76-4663-b0a7-000930a27552>
- Servicio Nacional de Meteorología e Hidrología del Perú. (2020c). *Reporte de Condiciones Secas y húmedas en el Perú durante el año hidrológico 2019-2020*. 0–34. https://scholar.googleusercontent.com/scholar?q=cache:_M5UVN90yYcJ:scholar.google.com/+condiciones+hidrológicas+del+Perú&hl=es&as_sdt=0,5
- Servicio Nacional de Meteorología e Hidrología del Perú. (2021). *Módulo para la estimación de curvas de Intensidad-Duración-Frecuencia (IDF)*. <https://idsep.senamhi.gob.pe/dhi-idf/>
- Simondi, T. (2000). *FILExt*. <https://filext.com/es/>
- Stowhas, L. (2017). *Fundamentos de Hidrología aplicada*. Editorial USM. <https://bibliotecadigital.usm.cl/info/00566094>.
- Stowhas, L. (1983). Precipitaciones máximas diarias en Chile. *VI Congreso*

Nacional de Ingeniería Hidráulica, Departamento de Ingeniería Hidráulica de La Escuela de Ingeniería, PUC.

Sugai, M. R. B., & Fill, H. D. O. A. (1990). Tempo de Recorrência Associado à Precipitação Máxima Provável na Região Sul do Brasil. *Revista Brasileira de Engenharia*, 8(1), 110.

Tableau. (s.f.). *Por qué elegir el análisis visual*.
https://help.tableau.com/current/blueprint/es-es/bp_why_visual_analytics.htm#:~:text=El análisis visual es un,dentro de una interfaz visual.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the Number of Clusters in a Data Set via the Gap Statistic. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 63(2), 411–423.
<http://www.jstor.org/stable/2680607>

World Meteorological Organization. (1971). Climatic Change. Annexed III: Standard Tests of Significance to be Recommended in Routine Analysis of Climatic Fluctuations. Technical Note No. 79. Geneva: *World Meteorological Organization, Secretariat of the WMO, WMO-No. 19*, 58–71.

World Meteorological Organization. (1986). *Manual for estimation of probable maximum precipitation*.

World Meteorological Organization. (1994). *Guía de prácticas hidrológicas*.

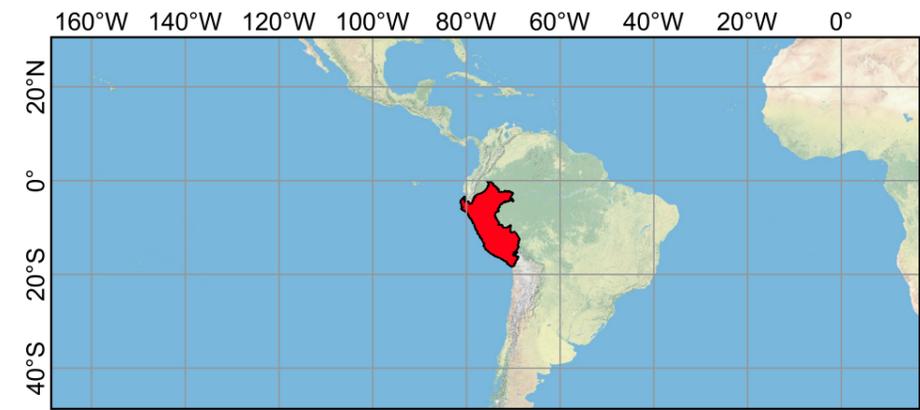
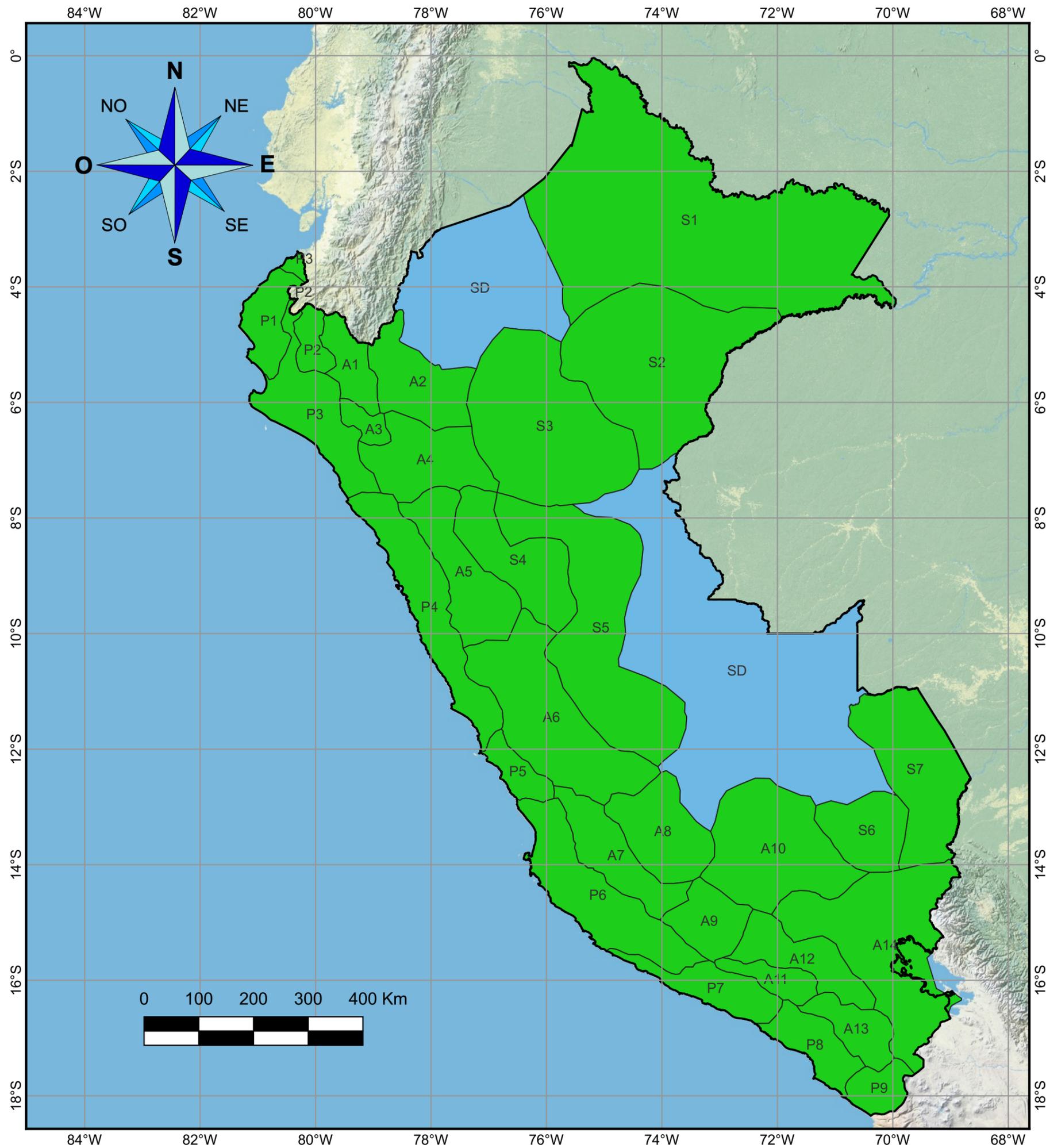
World Meteorological Organization. (2009). *Manual on estimation of probable maximum precipitation (PMP)* (WMO-No. 10). WMO.

Yue, S., Pilon, P., Phinney, B., & Cavadias, G. (2002). The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes*, 16(9), 1807–1829. <https://doi.org/10.1002/hyp.1095>

Anexos

Anexo A: Mapas utilizados para el desarrollo del proyecto	158
Anexo B: Estaciones empleadas en el análisis de la PMP para 24; 48 y 72 horas.....	163
Anexo C: Envoltentes de K para cada grupo de <i>clusters</i> para 24; 48 y 72 horas.....	181
Anexo D: Mapas relacionado a la duración de 24 horas	188
Anexo E: Mapas relacionado a la duración de 48 horas	196
Anexo F: Mapas relacionado a la duración de 72 horas.....	204

ANEXO A: MAPAS UTILIZADOS PARA EL DESARROLLO DEL PROYECTO



MAPA DE LOCALIZACIÓN
ESCALA : 1 / 150 000 000

LEYENDA	
REGIONES HOMOGÉNEAS DE PRECIPITACIÓN MÁXIMA	
■	A1 / A2 / A3 / A4 / A5 / A6 / A7 / A8 / A9 / A10 / A11 / A12 / A13 / P1 / P2 / P3 / P4 / P5 / P6 / P7 / P8 / P9 / S1 / S2 / S3 / S4 / S5 / S6 / S7
■	SIN DATOS

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.
3. LAS REGIONES HOMOGÉNEAS DE PRECIPITACIÓN MÁXIMA FUERON OBTENIDAS DEL CATÁLOGO NACIONAL DE METADATOS (SENAMHI 2020).

VISTA
ESCALA: 1 / 7 000 000



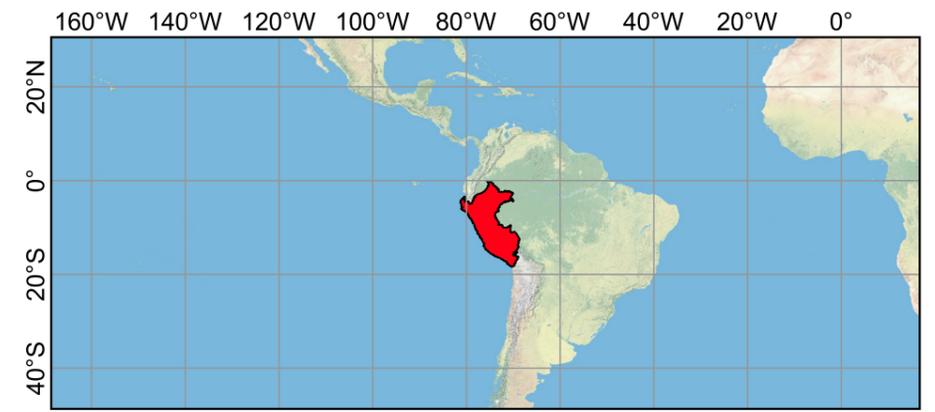
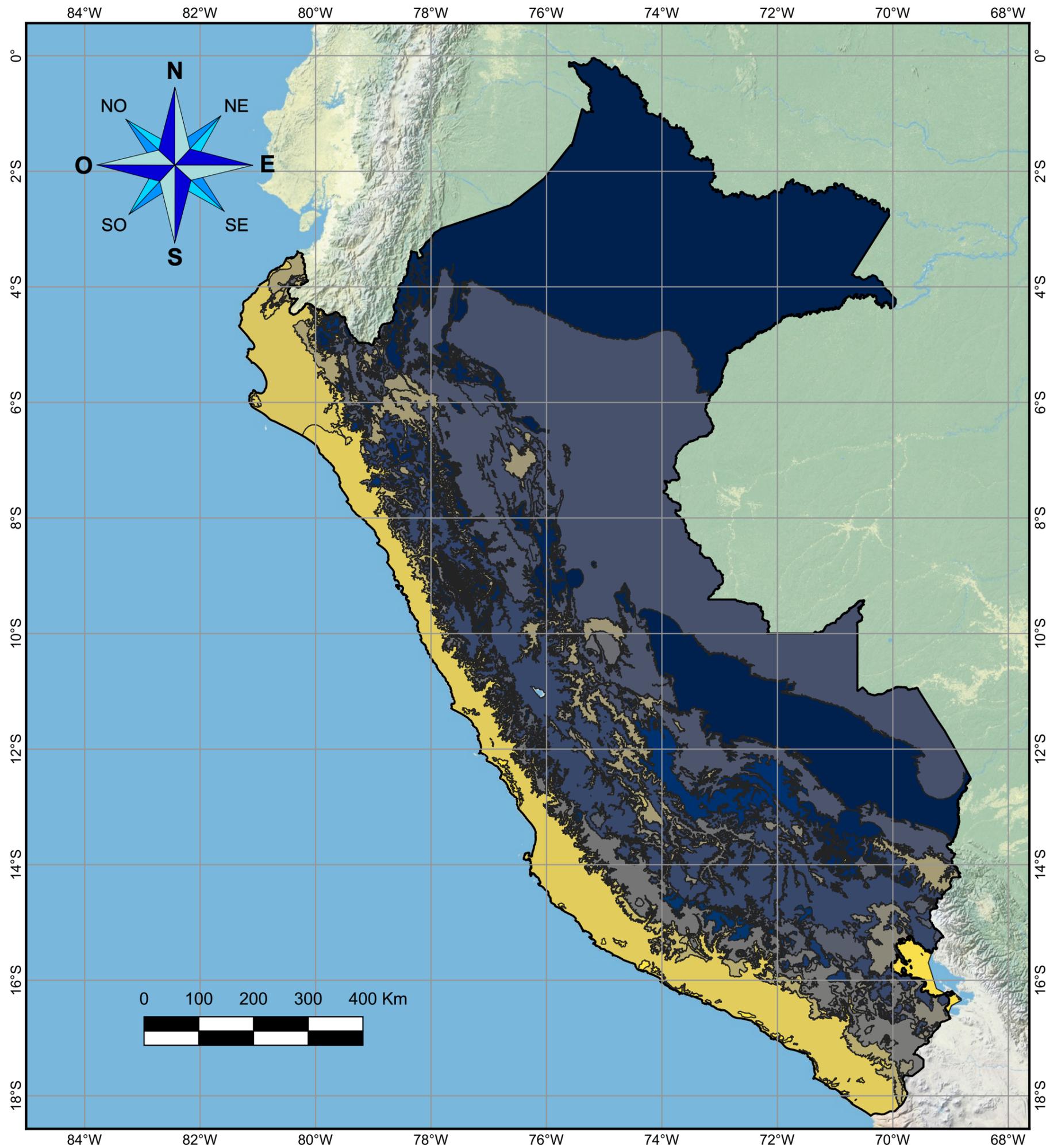
UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
REGIONES HOMOGÉNEAS DE PRECIPITACIÓN MÁXIMA

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA: DANTE RAMIRO LUCANO ALARCO	
SISTEMA DE COORDENADAS: WGS 84	Nº MAPA: A.1
LIMA - PERÚ 2024	



MAPA DE LOCALIZACIÓN
ESCALA : 1 / 150 000 000

LEYENDA		
Regiones climáticas Senamhi	B (r) C'	C (r) B'
A (r) A'	B (r) D'	D (i, p) A'
A (r) B'	C (i, p) A'	D (i, p) B'
A (r) C'	C (i, p) B'	D (i, p) C'
A (r) D'	C (i) A'	D (i) A'
B (i) B'	C (i) B'	D (i) B'
B (i) C'	C (i) C'	D (i) C'
B (i) D'	C (i) D'	D (i) D'
B (i) E'	C (i) E'	E (d) A'
B (o, i) B'	C (o, i) B'	E (d) B'
B (o, i) C'	C (o, i) C'	Glaciar
B (o, i) D'	C (o, i) D'	Lago Junín
B (r) A'	C (o, i) E'	Lago Titicaca
B (r) B'	C (r) A'	Loma

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.
3. LAS REGIONES CLIMÁTICAS PROVIENEN DEL MAPA DE CLASIFICACIÓN CLIMÁTICA DEL PERÚ EXTRAIDO DEL CATÁLOGO DE METADATOS CARTOGRÁFICOS (SENAMHI, 2020).

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESISTA:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
REGIONES CLIMÁTICAS DEL PERÚ

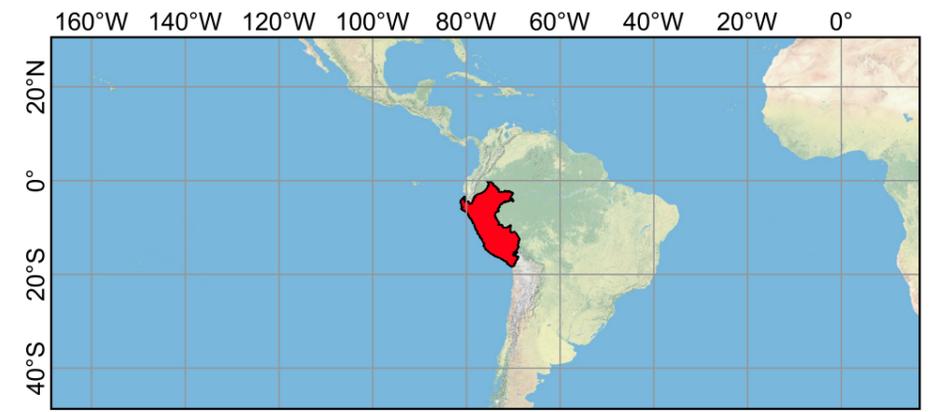
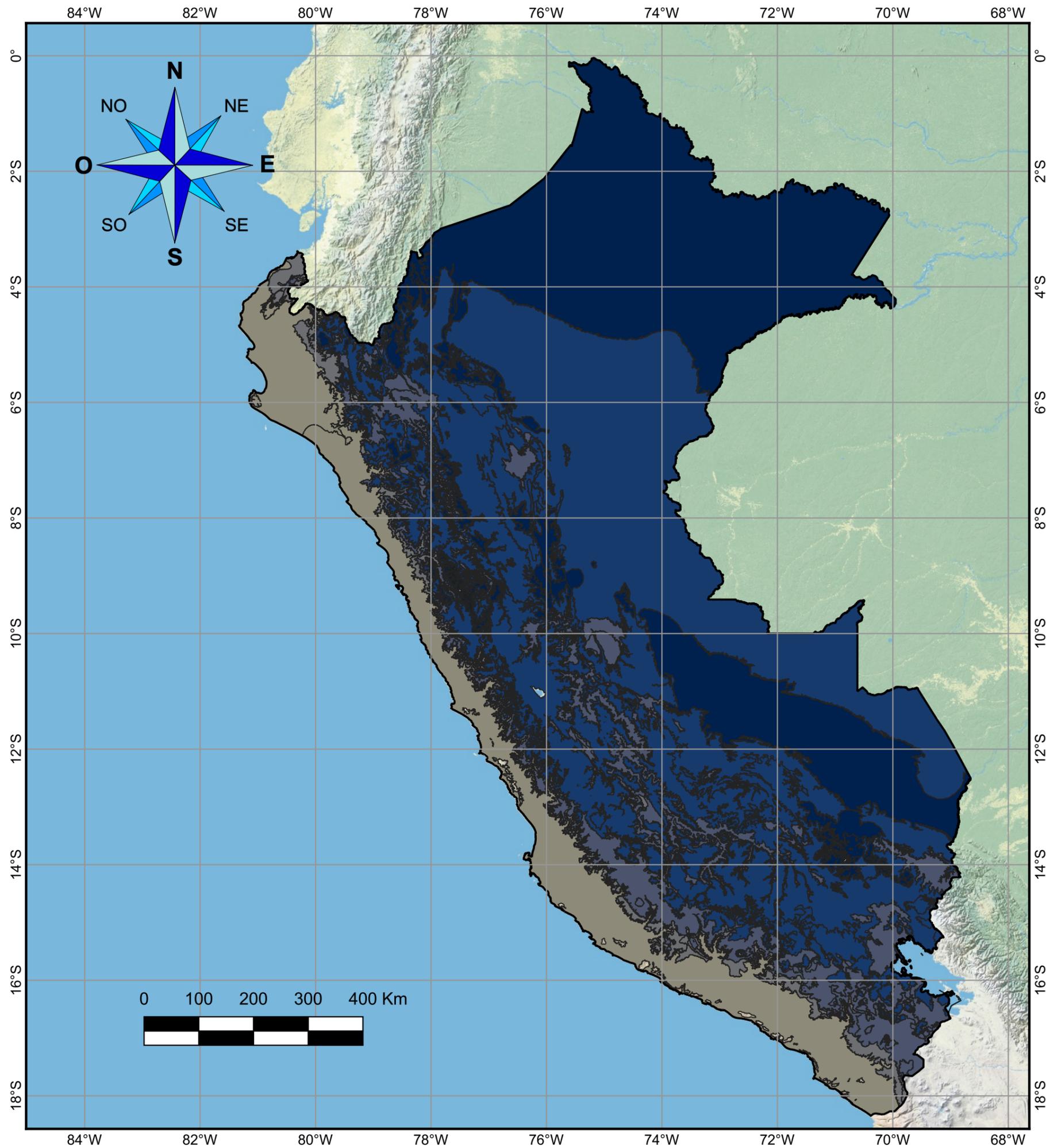
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

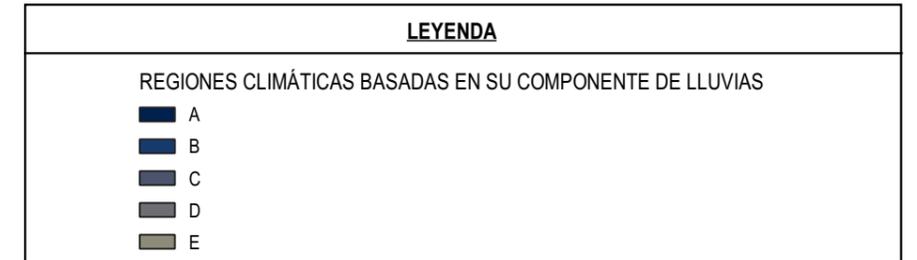
LIMA - PERÚ 2024

N° MAPA:
A.2



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.
3. LAS REGIONES CLIMÁTICAS PROVIENEN DEL MAPA DE CLASIFICACIÓN CLIMÁTICA DEL PERÚ EXTRAIDO DEL CATÁLOGO DE METADATOS CARTOGRÁFICOS (SENAMHI, 2020).
4. LAS PEQUEÑAS ZONAS VACÍAS CORRESPONDEN A LAGOS, GLACIARES Y LOMAS.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

REGIONES CLIMÁTICAS DEL PERÚ BASADAS EN SU COMPONENTE DE LLUVIAS

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.

“El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”

Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

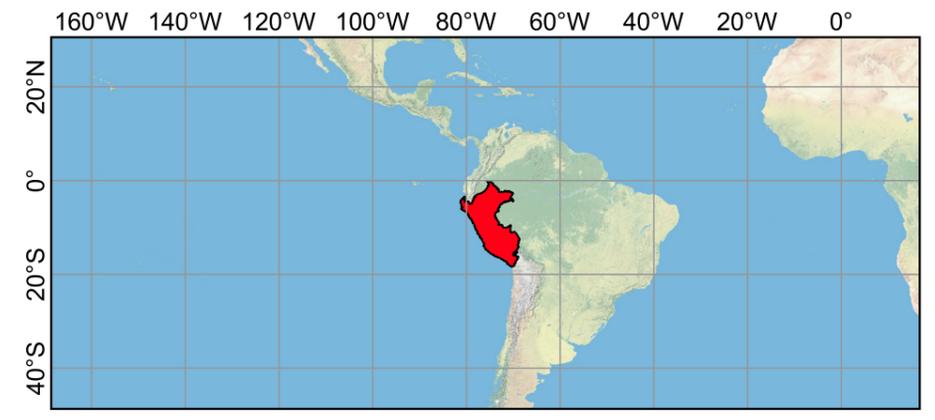
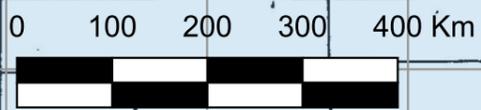
Nº MAPA:

A.3



Nº	DEPARTAMENTO
I	TUMBES
II	PIURA
III	LAMBAYEQUE
IV	LA LIBERTAD
V	CAJAMARCA
VI	AMAZONAS
VII	SAN MARTÍN
VIII	LORETO
VIII-a	UCAYALI
IX	PASCO
X	HUANUCO
XI	ANCASH
XII	LIMA
XIII	CALLAO
XIV	JUNÍN
XV	HUANCAVELICA
XVI	ICA
XVII	AREQUIPA
XVIII	AYACUCHO
XIX	APURÍMAC
XX	CUSCO
XXI	MADRE DE DIOS
XXII	PUNO
XXIII	MOQUEGUA
XXIV	TACNA

SIMBOLOGÍA	
*	ZONA NO DEFINIDA, por falta de información
X	NÚMERO DE ORDEN DEL DEPARTAMENTO
- - -	LÍMITE DEL DEPARTAMENTO
—	LÍMITE DE ZONA PLUVIOMÉTRICA



MAPA DE LOCALIZACIÓN
ESCALA: 1 / 150 000 000

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.
3. LAS ZONAS PLUVIOMÉTRICAS PROVIENEN DEL ESTUDIO DE LA HIDROLOGÍA DEL PERÚ (SERVICIO NACIONAL DE METEOROLOGÍA E HIDROLOGÍA DEL PERÚ, 1982-12).

VISTA
ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
ZONAS PLUVIOMÉTRICAS DEL PERÚ

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

Nº MAPA:
A.4

LIMA - PERÚ 2024

**ANEXO B: ESTACIONES EMPLEADAS EN EL ANÁLISIS DE LA PMP PARA
24; 48 Y 72 HORAS**

Tabla 1 Estaciones empleadas para el análisis de la PMP de 24 horas

Código ANA	Nombre	Código ANA	Nombre
486	CAPITANIA HUACHO	765	IMATA
800	MOLLENDO	658	PAUCARBAMBA
840	ILO	657	LIRCAY
853	LOCUMBA	156211	VILCASHUAMAN
536	SANTA ROSA02	156126	HUANCALPI
413	PUNTA CULEBRAS	155514	MILLOC
6617	CAMPO DE MARTE	821	RINCON DE LA CRUZ
543	ÑAÑA	152109	TOMA DE ZAMBA
602	CAMPO DE MARTE CGFAP	648	PILCHACA
624	PANTANOS DE VILLA	208	MALLARES
530	HUARMEY	152126	SONDORILLO
6642	SANTIAGO	747	CHALHUANCA
501	ALCANTARILLA	664	WUAYLLAPAMPA
418	HUACATAMBO	157409	TOROYA
619	SAN MARCOS02	625	ACOSTAMBO
831	OCOÑA	759	SICUANI
804	LA JOYA	157733	RAMIS
834	SANTA ELENA	157217	PAICO
830	PUNTA ATICO	157404	NUÑA
833	APLAO	728	PAMPACHIRI
858	ITE	156307	CHITAPAMPA
807	CALANA	157322	PUSA PUSA
535	ANDAHUASI	154103	CACHICADAN
856	RIACHUELO	152170	MINAS CAYLLOMA
158312	HACIENDA PUQUIO	593	CERRO DE PASCO
901	JORGE BASADRE	881	PIZCOMA
329	PUERTO CHICAMA	362	BAMBAMARCA
546	DONOSO	4431	CABANA
158300	ITE02	441	RECUAY
875	SAMA GRANDE	776	AYAVIRI02
805	PAMPA DE MAJES	477	SANTA ANA
862	CALIENTES	635	HUAYAO
158311	ILABAYA01	879	ILAVE
539	HUAYAN02	761	LLALLY
731	PALPA	686	PARURO
638	PACARAN	607	GRANJA KCAYRA
700	SAN CAMILO	155111	TINGO
863	ILABAYA02	748	ANTABAMBA
806	MOQUEGUA	114108	CURPAHUASI
410	LAREDO	663	LA QUINUA
345	CARTAVARIO	153101	CALLANCAS
344	CASA GRANDE	239	HUANCABAMBA01
640	HUAMANI	812	POMACANCHI
639	HUANCANO	155407	HUARAZ

Código ANA	Nombre	Código ANA	Nombre
151209	CHOSICA	824	PUTINA
544	LA CANTUTA	156210	VISCHONGO
332	REQUE	156128	LARAMARCA
325	TALLA GUADALUPE	783	ARAPA
599	ISLETA ANCON	755	SIBAYO
158315	CURIBAYA	155228	SAN PEDRO DE CHUCLU
331	FERREÑAFE	823	SANTA ROSA03
155107	PAMPA LIBRE	426	SANTIAGO ANTUNEZ DE MAYOLO
157102	PAMPA BLANCA	714	PEDREGAL
155213	SANTA EULALIA	153206	SINSICAP
746	CARAVELI	665	HUANCAPI
158321	PALCA	157205	PAMPAHUASI
301	LAMBAYEQUE	788	CAPACHICA
326	SAN PEDRO01	882	TAHUACO YUNGUYO
156113	SAN JUAN DE YANAC	687	ACOMAYO
155169	TOMAS	827	COJATA
299	PUCALA	739	PAMPA GALERAS
158302	OTORA	781	AZANGARO
157100	MALLUCHIMPANA	152130	SAN FELIPE
157313	AYO	357	COINA
157201	OTOCA	140305	CAJAMARCA
306	SIPAN	321	NAMORA
150900	CHAMANA	157332	HACIENDA COLONI
320	CAYALTI	792	PAUCARAY
152100	BERNAL	307	INCAHUASI
6640	CALLACA CO	153327	HUANGACOCOA
155106	ACO	157412	UMAYO
839	LA PAMPILLA	642	LAIVE
155122	AUTISHA	816	TARACO
150108	HACIENDA FERNANDEZ	764	CHUQUIBAMBILLA
396	MONTE GRANDE	782	LOS UROS
231	CHUSIS	552	SAN RAFAEL02
158313	CAIRANI	153108	QUEBRADA SHUGAR
151211	AMBAR	820	MAÑAZO
154107	CHACCHAN	684	ANTA ANCACHURO
155291	LAGUNA COCHAQUILLO	889	LARAQUERI
158303	COSCORI	815	PUCARA02
111175	SAN MATEO DE HUANCHOR	779	LAMPA02
150904	PARIACOTO	762	PAMPAHUTA
154113	COTAPARACO	157401	QUILLISANI
157314	CHOCO	811	TAMBOBAMBA
814	ARMA	157325	TISCO
157300	CHICHAS	754	LA ANGOSTURA
158202	LLUTA	13404001	CONDOROMA
548	MATUCANA	156131	PARIONA

Código ANA	Nombre	Código ANA	Nombre
155204	PACHANGARA	370	SAN MARCOS01
636	YAUYOS	388	LLAPA
440	AIJA	152166	LLACTO SAYAÑA
158205	SOCABAYA	152403	TAMANCO
230	LA ESPERANZA	241	EL LIMON
158301	QUINISTAQUILLAS	153111	LIVES
554	TARMA	152101	PANANGA
8791	ANTIQUILLA	351	SANTA CRUZ03
155102	LLACLLIN	682	ZURITE
155103	CAJACAY	152132	PORCULLA
156114	SAN JUAN DE CASTROVIRREYNA	822	ISLA SUANA
155203	CARAC	708	PUNO
151215	SANTANDER	786	HUANCANE
156121	CUSICANCHA	152129	SALLIQUE
155105	OCROS	6670	CALLABAMBA
997	TAZA OROYA	157420	CONIMA
151214	LANGA	369	SAN JUAN
158317	SUSAPAYA	676	MARANURA
150004	LAGUNA RAMON	153201	SAN BENITO
151206	CHAUTE	883	DESAGUADERO
151204	ARAHUAY	153215	YAMAGUAL
157200	LLAUTA	153235	CHOTANO LAJAS
151207	HUANCATA	240	TABACONAS
156109	CARANIA	880	JULI
151212	SAN PEDRO DE PILAS	140	LAS LOMAS
150903	MAYORARCA	352	CUTERVO
425	HUARI	232	CHILACO
541	OYON	236	OLMOS
154108	CAJAMARQUILLA	375	CHACHAPOYAS
157202	OCAÑA	303	CHOTA
151208	GORGOR	152128	HACIENDA SHUMAYA
151210	RIO BLANCO	152107	VIRREY
153203	CHILETE	787	HUARAYA MOHO
156110	HUANGASCAR	152204	EL PINTOR
155218	HUAROS	152161	PIRGAS
847	CHIGUATA	153110	UDIMA
247	SAN MIGUEL02	371	CELENDIN
643	COCAS	152213	PUENTE CHUNHUCA
797	AGUADA BLANCA	228	ARENALES
801	SALAMANCA	253	BAGUA CHICA
151213	SHEQUE	343	HUAMBOS
876	CANDARAVE	359	GRANJA PORCON
749	COTAHUASI	136	CAÑAVERAL
158332	LA FRONTERA	152134	COLASAY
155205	PALLAC	557	OXAPAMPA

Código ANA	Nombre	Código ANA	Nombre
864	YANAQUIHUA	3114	PUCHACA
155225	SAN JOSE DE PARAC	152123	HUAR HUAR SALALA
877	TARATA	135	EL SALTO
445	CHAVIN	252	JAEN
158328	PAUCARANI	134	PAPAYAL
150300	PUENTE SULLANA	152210	MAGUNCHAL
155202	SANTA CRUZ04	679	MACHU PICCHU
155101	TAPACOCHA	152144	MONTERO
156118	ACOBAMBILLA	250	CHONTALI
851	UBINAS	262	PASABAR
994	POMACOCHA	353	COCHABAMBA
157210	PAUCACORRAL	242	SAN IGNACIO
156119	TOTORA	152112	CHALACO
799	HUASACACHE	152111	HACIENDA BIGOTE
751	PMPACOLCA	152141	SICCHEZ
157315	HUAMBO	200164	Estacion 164
155229	RICRAN	152108	FRIAS
849	EL FRAYLE	272	JAZAN
157307	PUICA	255	CHULUCANAS
683	URUBAMBA02	341	LLAMA
745	PAUZA	152135	SAN JUAN DE LOS ALISOS
649	HUANCAVELICA	248	HUARMACA
115129	MACHAHUAY	152103	LANCONES
154110	PIRA	153300	JEPELACIO
157310	ANDAHUA	152156	ANIA CABUYAL
489	COSMOS	237	AYABACA
502	SURASACA	572	PUERTO OCOPA
155200	PACCHO	216	PARTIDOR
333	JAYANCA LA VIÑA	152147	BARRIOS
302	OYOTUN	381	SISA
155212	PARQUIN	150303	ALIVIADEROS
736	PUQUIO	454	CARPISH
157317	MADRIGAL	152118	CANCHAQUE
750	CHUQUIBAMBA	790	TAMBOPATA
157312	CHACHAS	561	SAN RAMON02
156111	VILCA	590	IÑAPARI
315	HACIENDA TUNAD	152202	ARDILLA SOLANA BAJA
158208	PILLONES	384	LA UNION01
155112	PARIACANCHA	153326	SHAMBOYACU
553	HUARIACA	152110	SANTO DOMINGO
660	HUANTA	153316	PILLUANA
158326	CAPAZO	310	EL PORVENIR
316	AYLAMBO	153311	CUÑUMBUQUE
669	ANDAHUAYLAS	152142	AUL C MEMBRILLO
156214	CARHUANCA	219	NARANJILLO

Código ANA	Nombre	Código ANA	Nombre
155217	LACHAQUI	385	SAUCE
737	ANDAMARCA02	150112	MATAPALO
475	YANAHUANCA	153307	SAN PABLO01
151503	HUACHOS	153345	CUZCO
860	CHUAPALCA	383	LAMAS
158308	CALACOA	150001	SAN PEDRO02
157326	PULLHUAY AYAHUASI	297	TINGO DE PONAZA
404	HUANUCO	382	BELLAVISTA02
154106	MOLLEPATA01	153313	PICOTA
151506	SANTIAGO DE HUIRPACANCHA	476	PICHANAKY
155115	CARHUACAYAN	377	RIOJA
398	SALPO	153312	NUEVO LIMA
156102	SAN LAZARO DE ESCOMARCA	155421	PERENE
157311	ORCOPAMPA	571	SATIPO
207	MIRAFLORES	152140	PICO DE LORO
209	EL ALTO	401	TARAPOTO
685	CALCA	153306	SACANCHE
549	MARCAPOMACOCHA	556	POZUZO
157207	PAMPAMARCA	153226	SORITOR
150901	MALVAS	5572	VILLA RICA
157209	CERRO CONDORILLO	139	CABO INGA
153212	HACIENDA LLAGADEN	211	SAPOSOA
743	CORACORA	3308	ALAO
155232	RUNATULLO	152409	SANTA MARIA DE NANAY
155224	SANTIAGO DE TUNA	152153	ALAMOR
795	CABANACONDE	153314	SAN ANTONIO01
156204	URUBAMBA01	152404	SILVA MERINO
659	ACOMBAMBA	150003	EL CIRUELO
156122	TAMBO	154208	TULUMAYO
604	LA OROYA	153322	PELEJO
574	USIBAMBA	152402	BETRAÑA
878	MAZO CRUZ	387	CONTAMANA
112190	HUANTA GORE	453	LA DIVISORIA
153214	HACIENDA LLUCAN	154305	MASISEA
861	VOLACOTA	447	TOCACHE
803	CRUCERO ALTO	465	CACHICOTO
605	ABANCAY	279	NAUTA
777	MACUSANI	173	MUYUY
785	MUÑANI	153315	SHANUSI
157328	SUMBAY	151	MORONA
996	PACHACHACA02	281	GENARO HERRERA
155234	PARIAMARCA	480	TOURNAVISTA
677	CARAHUASI	446	TANANTA
156104	AYAVIRI01	140406	TINGO MARIA
760	PAÑE	291	CABALLOCOCHA

Código ANA	Nombre	Código ANA	Nombre
873	CARUMAS	180	PEBAS
155450	YURICOCHA	150206	SANTA CRUZ01
560	COMAS	3317	PONGO DE CAYNARACHI
667	QUEROBAMBA	278	SAN RAMON01
756	COMBAPATA	172	TAMSHIYACU
661	HUAMANGA	478	PUERTO INCA
157204	PUTACCASA	150201	MANITI
706	LOS LIBERTADORES	2412	SAN ALEJANDRO
155231	INGENIO	153400	TIRUNTAN
594	SAN JUAN DE JARPA	459	EL MARONAL
153218	LA LLICA	150208	PUNCHANA
6620	COLCABMBA	691	PILCOPATA
778	PROGRESO	150207	LA LIBERTAD
391	JESUS	6401	SAN GABAN
692	OCONGATE	153320	PUCALLPA01
503	JAUJA	462	AGUAYTIA
157206	LUCANAS	150204	TIMICURILLO
647	TUNEL CERO	693	QUINCÉMIL
690	CCATCCA		

Fuente: Elaboración propia.

Tabla 2 Estaciones empleadas para el análisis de la PMP de 48 horas

Código ANA	Nombre	Código ANA	Nombre
486	CAPITANIA HUACHO	759	SICUANI
800	MOLLENDO	157204	PUTACCASA
840	ILO	155228	SAN PEDRO DE CHUCLU
413	PUNTA CULEBRAS	873	CARUMAS
536	SANTA ROSA02	765	IMATA
853	LOCUMBA	607	GRANJA KCAYRA
6617	CAMPO DE MARTE	824	PUTINA
602	CAMPO DE MARTE CGFAP	157101	CORDOVA
530	HUARMEY	362	BAMBAMARCA
543	ÑAÑA	728	PAMPACHIRI
6642	SANTIAGO	239	HUANCABAMBA01
624	PANTANOS DE VILLA	156307	CHITAPAMPA
501	ALCANTARILLA	706	LOS LIBERTADORES
899	LA YADARA	776	AYAVIRI02
619	SAN MARCOS02	155100	PARARIN
418	HUACATAMBO	594	SAN JUAN DE JARPA
804	LA JOYA	157409	TOROYA
631	LA CAPILLA 2	154200	ONGON
831	OCOÑA	157221	PECOPE
834	SANTA ELENA	552	SAN RAFAEL02
830	PUNTA ATICO	441	RECUAY
833	APLAO	781	AZANGARO

Código ANA	Nombre	Código ANA	Nombre
535	ANDAHUASI	156126	HUANCALPI
858	ITE	155514	MILLOC
856	RIACHUELO	760	PAÑE
546	DONOSO	663	LA QUINUA
329	PUERTO CHICAMA	156210	VISCHONGO
158312	HACIENDA PUQUIO	827	COJATA
807	CALANA	761	LLALLY
805	PAMPA DE MAJES	821	RINCON DE LA CRUZ
539	HUAYAN02	114108	CURPAHUASI
901	JORGE BASADRE	156211	VILCASHUAMAN
862	CALIENTES	460	HUALLANCA
158300	ITE02	593	CERRO DE PASCO
731	PALPA	783	ARAPA
158311	ILABAYA01	755	SIBAYO
700	SAN CAMILO	4431	CABANA
638	PACARAN	158310	ICHUÑA
863	ILABAYA02	157322	PUSA PUSA
806	MOQUEGUA	748	ANTABAMBA
344	CASA GRANDE	686	PARURO
410	LAREDO	780	CABANILLAS
345	CARTAVARIO	155111	TINGO
639	HUANCANO	153101	CALLANCAS
640	HUAMANI	321	NAMORA
151209	CHOSICA	764	CHUQUIBAMBILLA
544	LA CANTUTA	157206	LUCANAS
332	REQUE	816	TARACO
599	ISLETA ANCON	687	ACOMAYO
331	FERREÑAFE	879	ILAVE
301	LAMBAYEQUE	426	SANTIAGO ANTUNEZ DE MAYOLO
157102	PAMPA BLANCA	881	PIZACOMA
326	SAN PEDRO01	823	SANTA ROSA03
155107	PAMPA LIBRE	812	POMACANCHI
746	CARAVELI	152101	PANANGA
299	PUCALA	156128	LARAMARCA
158321	PALCA	0	YURACMAYO
320	CAYALTI	241	EL LIMON
158302	OTORA	140305	CAJAMARCA
152100	BERNAL	157412	UMAYO
155169	TOMAS	152170	MINAS CAYLLOMA
306	SIPAN	882	TAHUACO YUNGUYO
150108	HACIENDA FERNANDEZ	739	PAMPA GALERAS
156113	SAN JUAN DE YANAC	820	MAÑAZO
231	CHUSIS	815	PUCARA02
6640	CALLACA CO	157332	HACIENDA COLONI
150900	CHAMANA	154112	HUACAMARCANGA

Código ANA	Nombre	Código ANA	Nombre
157100	MALLUCHIMPANA	370	SAN MARCOS01
839	LA PAMPILLA	642	LAIVE
157201	OTOCA	153206	SINSICAP
396	MONTE GRANDE	665	HUANCABI
230	LA ESPERANZA	152403	TAMANCO
158303	COSCORI	157217	PAICO
155122	AUTISHA	307	INCAHUASI
158325	TOQUELA	788	CAPACHICA
158313	CAIRANI	762	PAMPAHUTA
111175	SAN MATEO DE HUANCHOR	782	LOS UROS
150904	PARIACOTO	779	LAMPA02
158205	SOCABAYA	157325	TISCO
154107	CHACCHAN	889	LARAQUERI
155291	LAGUNA COCHAQUILLO	152130	SAN FELIPE
554	TARMA	357	COINA
151211	AMBAR	157205	PAMPAHUASI
157300	CHICHAS	811	TAMBOBAMBA
158202	LLUTA	153108	QUEBRADA SHUGAR
548	MATUCANA	754	LA ANGOSTURA
997	TAZA OROYA	752	SANTO TOMAS
155203	CARAC	140	LAS LOMAS
151215	SANTANDER	156131	PARIONA
151206	CHAUTE	152113	NAGAY MATLACAS
150004	LAGUNA RAMON	157401	QUILLISANI
156114	SAN JUAN DE CASTROVIRREYNA	682	ZURITE
247	SAN MIGUEL02	708	PUNO
154113	COTAPARACO	351	SANTA CRUZ03
157314	CHOCO	153111	LIVES
636	YAUYOS	6670	CALLABAMBA
814	ARMA	786	HUANCANE
440	AIJA	676	MARANURA
153203	CHILETE	152129	SALLIQUE
155103	CAJACAY	153107	HACIENDA PUCARA
151212	SAN PEDRO DE PILAS	152166	LLACTO SAYAÑA
151214	LANGA	714	PEDREGAL
150300	PUENTE SULLANA	822	ISLA SUANA
151204	ARAHUAY	253	BAGUA CHICA
151207	HUANCATA	153235	CHOTANO LAJAS
157200	LLAUTA	695	OLLACHEA
155207	PACHAMACHAY	236	OLMOS
155105	OCROS	388	LLAPA
542	PICOY	152204	EL PINTOR
156121	CUSICANCHA	152213	PUENTE CHUNHUCA
425	HUARI	153327	HUANGACocha
158317	SUSAPAYA	240	TABACONAS

Código ANA	Nombre	Código ANA	Nombre
541	OYON	880	JULI
333	JAYANCA LA VIÑA	883	DESAGUADERO
555	HUASAHUASI	374	HUAMACHUCO
157202	OCAÑA	153201	SAN BENITO
155223	CARAMPOMA	152107	VIRREY
156110	HUANGASCAR	232	CHILACO
150903	MAYORARCA	375	CHACHAPOYAS
151210	RIO BLANCO	153215	YAMAGUAL
151205	CANCHACALLA	152161	PIRGAS
799	HUASACACHE	136	CAÑAVERAL
155218	HUAROS	352	CUTERVO
156109	CARANIA	152128	HACIENDA SHUMAYA
864	YANAQUIHUA	369	SAN JUAN
155102	LLACLLIN	252	JAEN
209	EL ALTO	305	NIEPO
847	CHIGUATA	3114	PUCHACA
797	AGUADA BLANCA	152132	PORCULLA
302	OYOTUN	371	CELENDIN
876	CANDARAVE	343	HUAMBOS
155205	PALLAC	152210	MAGUNCHAL
151213	SHEQUE	787	HUARAYA MOHO
877	TARATA	135	EL SALTO
643	COCAS	228	ARENALES
445	CHAVIN	817	ISLA SOTO
404	HUANUCO	134	PAPAYAL
315	HACIENDA TUNAD	242	SAN IGNACIO
207	MIRAFLORES	132	PUERTO PIZARRO
683	URUBAMBA02	250	CHONTALI
151208	GORGOR	359	GRANJA PORCON
155225	SAN JOSE DE PARAC	557	OXAPAMPA
749	COTAHUASI	262	PASABAR
156118	ACOBAMBILLA	152134	COLASAY
316	AYLAMBO	353	COCHABAMBA
801	SALAMANCA	272	JAZAN
155201	ANDAJES	679	MACHU PICCHU
158328	PAUCARANI	152119	HUARA DE VERAS
809	CAY CAY	152103	LANCONES
540	CAJATAMBO	572	PUERTO OCOPA
157315	HUAMBO	153300	JEPELACIO
996	PACHACHACA02	200164	Estacion 164
848	PUQUINA	152111	HACIENDA BIGOTE
851	UBINAS	153316	PILLUANA
745	PAUZA	384	LA UNION01
849	EL FRAYLE	255	CHULUCANAS
155229	RICRAN	153103	TOCMOCHE

Código ANA	Nombre	Código ANA	Nombre
994	POMACOCHA	152144	MONTERO
574	USIBAMBA	153326	SHAMBOYACU
155202	SANTA CRUZ04	153311	CUÑUMBUQUE
685	CALCA	790	TAMBOPATA
604	LA OROYA	381	SISA
489	COSMOS	150303	ALIVIADEROS
553	HUARIACA	152112	CHALACO
157312	CHACHAS	310	EL PORVENIR
751	PMPACOLCA	153307	SAN PABLO01
646	ACNOCOCHA	153345	CUZCO
238	SAUSAL DE CULUCAN	152202	ARDILLA SOLANA BAJA
158332	LA FRONTERA	152141	SICCHEZ
669	ANDAHUAYLAS	219	NARANJILLO
156119	TOTORA	382	BELLAVISTA02
649	HUANCAVELICA	152108	FRIAS
750	CHUQUIBAMBA	216	PARTIDOR
208	MALLARES	561	SAN RAMON02
156214	CARHUANCA	153313	PICOTA
736	PUQUIO	590	IÑAPARI
661	HUAMANGA	152135	SAN JUAN DE LOS ALISOS
155112	PARIACANCHA	297	TINGO DE PONAZA
154106	MOLLEPATA01	152156	ANIA CABUYAL
860	CHUAPALCA	153312	NUEVO LIMA
153212	HACIENDA LLAGADEN	152147	BARRIOS
155212	PARQUIN	383	LAMAS
608	VIQUES	248	HUARMACA
475	YANAHUANCA	152118	CANCHAQUE
155200	PACCHO	150112	MATAPALO
157310	ANDAHUA	377	RIOJA
157307	PUICA	401	TARAPOTO
737	ANDAMARCA02	153306	SACANCHE
156111	VILCA	385	SAUCE
115129	MACHAHUAY	234	LA TINA
158208	PILLONES	211	SAPOSOA
154110	PIRA	150001	SAN PEDRO02
153214	HACIENDA LLUCAN	152110	SANTO DOMINGO
777	MACUSANI	153226	SORITOR
677	CARAHUASI	152142	AUL C MEMBRILLO
157209	CERRO CONDORILLO	571	SATIPO
156204	URUBAMBA01	3308	ALAO
151503	HUACHOS	5572	VILLA RICA
155234	PARIAMARCA	556	POZUZO
659	ACOMBAMBA	153314	SAN ANTONIO01
785	MUÑANI	152140	PICO DE LORO
112190	HUANTA GORE	139	CABO INGA

Código ANA	Nombre	Código ANA	Nombre
156102	SAN LAZARO DE ESCOMARCA	152409	SANTA MARIA DE NANAY
157311	ORCOPAMPA	152153	ALAMOR
605	ABANCAY	387	CONTAMANA
690	CCATCCA	154305	MASISEA
155101	TAPACOCHA	152404	SILVA MERINO
398	SALPO	153322	PELEJO
158326	CAPAZO	177	SANTA CLOTILDE
150901	MALVAS	152402	BETRAÑA
560	COMAS	447	TOCACHE
152126	SONDORILLO	154208	TULUMAYO
151506	SANTIAGO DE HUIRPACANCHA	173	MUYUY
878	MAZO CRUZ	150003	EL CIRUELO
158308	CALACOA	151	MORONA
391	JESUS	453	LA DIVISORIA
803	CRUCERO ALTO	281	GENARO HERRERA
155231	INGENIO	278	SAN RAMON01
778	PROGRESO	180	PEBAS
155115	CARHUACAYAN	150206	SANTA CRUZ01
503	JAUJA	153315	SHANUSI
477	SANTA ANA	446	TANANTA
795	CABANACONDE	172	TAMSHIYACU
756	COMBAPATA	480	TOURNAVISTA
692	OCONGATE	140406	TINGO MARIA
664	WUAYLLAPAMPA	465	CACHICOTO
156122	TAMBO	150201	MANITI
155217	LACHAQUI	478	PUERTO INCA
657	LIRCAY	150208	PUNCHANA
152109	TOMA DE ZAMBA	153400	TIRUNTAN
635	HUAYAO	2412	SAN ALEJANDRO
743	CORACORA	150207	LA LIBERTAD
648	PILCHACA	459	EL MARONAL
157207	PAMPAMARCA	466	LA MORADA
844	PISAC	3317	PONGO DE CAYNARACHI
157733	RAMIS	691	PILCOPATA
641	VILLA DE ARMAS	150204	TIMICURILLO
155407	HUARAZ	153320	PUCALLPA01
156104	AYAVIRIO1	6671	CHONTACHACA
647	TUNEL CERO	6401	SAN GABAN
153218	LA LLICA	462	AGUAYTIA
861	VOLACOTA	693	QUINCEMIL

Fuente: Elaboración propia.

Tabla 3 Estaciones empleadas para el análisis de la PMP de 72 horas

Código ANA	Nombre	Código ANA	Nombre
486	CAPITANIA HUACHO	743	CORACORA

Código ANA	Nombre	Código ANA	Nombre
800	MOLLEDO	765	IMATA
840	ILO	776	AYAVIRI02
413	PUNTA CULEBRAS	156104	AYAVIRI01
536	SANTA ROSA02	667	QUEROBAMBA
6617	CAMPO DE MARTE	460	HUALLANCA
853	LOCUMBA	152101	PANANGA
602	CAMPO DE MARTE CGFAP	157409	TOROYA
543	ÑAÑA	241	EL LIMON
530	HUARMEY	641	VILLA DE ARMAS
624	PANTANOS DE VILLA	873	CARUMAS
6642	SANTIAGO	552	SAN RAFAEL02
501	ALCANTARILLA	781	AZANGARO
899	LA YADARA	861	VOLACOTA
619	SAN MARCOS02	156126	HUANCALPI
418	HUACATAMBO	706	LOS LIBERTADORES
804	LA JOYA	157204	PUTACCASA
831	OCOÑA	728	PAMPACHIRI
834	SANTA ELENA	783	ARAPA
830	PUNTA ATICO	157101	CORDOVA
833	APLAO	761	LLALLY
794	TACAMA	155514	MILLOC
535	ANDAHUASI	594	SAN JUAN DE JARPA
858	ITE	441	RECUAY
329	PUERTO CHICAMA	663	LA QUINUA
856	RIACHUELO	827	COJATA
546	DONOSO	158310	ICHUÑA
158312	HACIENDA PUQUIO	780	CABANILLAS
805	PAMPA DE MAJES	157221	PECOPE
539	HUAYAN02	816	TARACO
807	CALANA	156211	VILCASHUAMAN
901	JORGE BASADRE	443	PMABAMBA
862	CALIENTES	748	ANTABAMBA
731	PALPA	812	POMACANCHI
158300	ITE02	156210	VISCHONGO
158311	ILABAYA01	747	CHALHUANCA
700	SAN CAMILO	821	RINCON DE LA CRUZ
638	PACARAN	687	ACOMAYO
806	MOQUEGUA	155111	TINGO
863	ILABAYA02	686	PARURO
410	LAREDO	426	SANTIAGO ANTUNEZ DE MAYOLO
344	CASA GRANDE	820	MAÑAZO
345	CARTAVARIO	140	LAS LOMAS
639	HUANCANO	153101	CALLANCAS
640	HUAMANI	370	SAN MARCOS01

Código ANA	Nombre	Código ANA	Nombre
332	REQUE	764	CHUQUIBAMBILLA
544	LA CANTUTA	4431	CABANA
599	ISLETA ANCON	157322	PUSA PUSA
151209	CHOSICA	456	JACAS CHICO
325	TALLA GUADALUPE	157412	UMAYO
331	FERREÑAFE	114108	CURPAHUASI
301	LAMBAYEQUE	307	INCAHUASI
157102	PAMPA BLANCA	879	ILAVE
326	SAN PEDRO01	156212	CHICAYOC
746	CARAVELI	642	LAIVE
299	PUCALA	760	PAÑE
155107	PAMPA LIBRE	665	HUANCAPI
152100	BERNAL	815	PUCARA02
158321	PALCA	881	PIZACOMA
320	CAYALTI	155100	PARARIN
158302	OTORA	140305	CAJAMARCA
150108	HACIENDA FERNANDEZ	882	TAHUACO YUNGUYO
306	SIPAN	152170	MINAS CAYLLOMA
231	CHUSIS	153206	SINSICAP
156100	ANTIOQUIA	762	PAMPAHUTA
155169	TOMAS	156128	LARAMARCA
156113	SAN JUAN DE YANAC	739	PAMPA GALERAS
230	LA ESPERANZA	157332	HACIENDA COLONI
6640	CALLACA CO	6670	CALLABAMBA
839	LA PAMPILLA	153107	HACIENDA PUCARA
150900	CHAMANA	157206	LUCANAS
396	MONTE GRANDE	823	SANTA ROSA03
157100	MALLUCHIMPANA	811	TAMBOBAMBA
157201	OTOCA	152204	EL PINTOR
158303	COSCORI	788	CAPACHICA
155122	AUTISHA	779	LAMPA02
150904	PARIACOTO	157325	TISCO
158313	CAIRANI	0	YURACMAYO
158205	SOCABAYA	157401	QUILLISANI
158325	TOQUELA	782	LOS UROS
247	SAN MIGUEL02	152129	SALLIQUE
554	TARMA	253	BAGUA CHICA
997	TAZA OROYA	152130	SAN FELIPE
154107	CHACCHAN	152213	PUENTE CHUNHUCA
111175	SAN MATEO DE HUANCHOR	152403	TAMANCO
150004	LAGUNA RAMON	754	LA ANGOSTURA
157300	CHICHAS	153108	QUEBRADA SHUGAR
151206	CHAUTE	157217	PAICO
156114	SAN JUAN DE CASTROVIRREYNA	357	COINA

Código ANA	Nombre	Código ANA	Nombre
155291	LAGUNA COCHAQUILLO	157205	PAMPAHUASI
151211	AMBAR	152113	NAGAY MATLACAS
548	MATUCANA	153111	LIVES
151215	SANTANDER	889	LARAQUERI
155203	CARAC	708	PUNO
158202	LLUTA	786	HUANCANE
153203	CHILETE	351	SANTA CRUZ03
156123	SANTIAGO DE CHOCORVOS	156131	PARIONA
150300	PUENTE SULLANA	154112	HUACAMARCANGA
157314	CHOCO	13404001	CONDOROMA
154113	COTAPARACO	152107	VIRREY
333	JAYANCA LA VIÑA	240	TABACONAS
155103	CAJACAY	236	OLMOS
440	AIJA	682	ZURITE
151214	LANGA	676	MARANURA
404	HUANUCO	154101	JULCAN
207	MIRAFLORES	152166	LLACTO SAYAÑA
814	ARMA	153235	CHOTANO LAJAS
151204	ARAHUAY	232	CHILACO
555	HUASAHUASI	252	JAEN
209	EL ALTO	373	CAJBAMBA
425	HUARI	375	CHACHAPOYAS
156121	CUSICANCHA	822	ISLA SUANA
542	PICOY	319	SAN PABLO02
155105	OCROS	388	LLAPA
157202	OCAÑA	714	PEDREGAL
157200	LLAUTA	695	OLLACHEA
541	OYON	883	DESAGUADERO
683	URUBAMBA02	303	CHOTA
155223	CARAMPOMA	152161	PIRGAS
864	YANAQUIHUA	153201	SAN BENITO
151207	HUANCATA	880	JULI
799	HUASACACHE	136	CAÑAVERAL
508	PAMPAS01	152128	HACIENDA SHUMAYA
302	OYOTUN	3114	PUCHACA
151210	RIO BLANCO	352	CUTERVO
156110	HUANGASCAR	153110	UDIMA
151205	CANCHACALLA	135	EL SALTO
847	CHIGUATA	371	CELENDIN
316	AYLAMBO	369	SAN JUAN
155218	HUAROS	134	PAPAYAL
315	HACIENDA TUNAD	132	PUERTO PIZARRO
158317	SUSAPAYA	153327	HUANGACocha
877	TARATA	152210	MAGUNCHAL
155207	PACHAMACHAY	242	SAN IGNACIO

Código ANA	Nombre	Código ANA	Nombre
604	LA OROYA	262	PASABAR
155225	SAN JOSE DE PARAC	654	QUEBRADA YANATILE
155205	PALLAC	272	JAZAN
574	USIBAMBA	787	HUARAYA MOHO
797	AGUADA BLANCA	250	CHONTALI
151213	SHEQUE	152132	PORCULLA
150903	MAYORARCA	228	ARENALES
208	MALLARES	153300	JEPELACIO
156118	ACOBAMBILLA	152134	COLASAY
643	COCAS	152103	LANCONES
158328	PAUCARANI	153316	PILLUANA
553	HUARIACA	359	GRANJA PORCON
685	CALCA	384	LA UNION01
155201	ANAJES	153311	CUÑUMBUQUE
745	PAUZA	557	OXAPAMPA
749	COTAHUASI	572	PUERTO OCOPA
157315	HUAMBO	255	CHULUCANAS
156109	CARANIA	200164	Estacion 164
155229	RICRAN	790	TAMBOPATA
489	COSMOS	382	BELLAVISTA02
801	SALAMANCA	153326	SHAMBOYACU
157312	CHACHAS	152111	HACIENDA BIGOTE
151208	GORGOR	310	EL PORVENIR
849	EL FRAYLE	153307	SAN PABLO01
851	UBINAS	153345	CUZCO
994	POMACOCHA	381	SISA
238	SAUSAL DE CULUCAN	153313	PICOTA
848	PUQUINA	219	NARANJILLO
155202	SANTA CRUZ04	152202	ARDILLA SOLANA BAJA
155102	LLACLIN	150303	ALIVIADEROS
751	PMPACOLCA	297	TINGO DE PONAZA
646	ACNOCOCHA	153312	NUEVO LIMA
669	ANDAHUAYLAS	401	TARAPOTO
661	HUAMANGA	152144	MONTERO
156214	CARHUANCA	153103	TOCMOCHE
152126	SONDORILLO	152141	SICCHEZ
750	CHUQUIBAMBA	152159	ALTAMIZA
158208	PILLONES	153306	SACANCHE
475	YANAHUANCA	590	IÑAPARI
608	VIQUES	152112	CHALACO
153212	HACIENDA LLAGADEN	216	PARTIDOR
154106	MOLLEPATA01	561	SAN RAMON02
158332	LA FRONTERA	152108	FRIAS
153214	HACIENDA LLUCAN	377	RIOJA
649	HUANCAVELICA	152147	BARRIOS

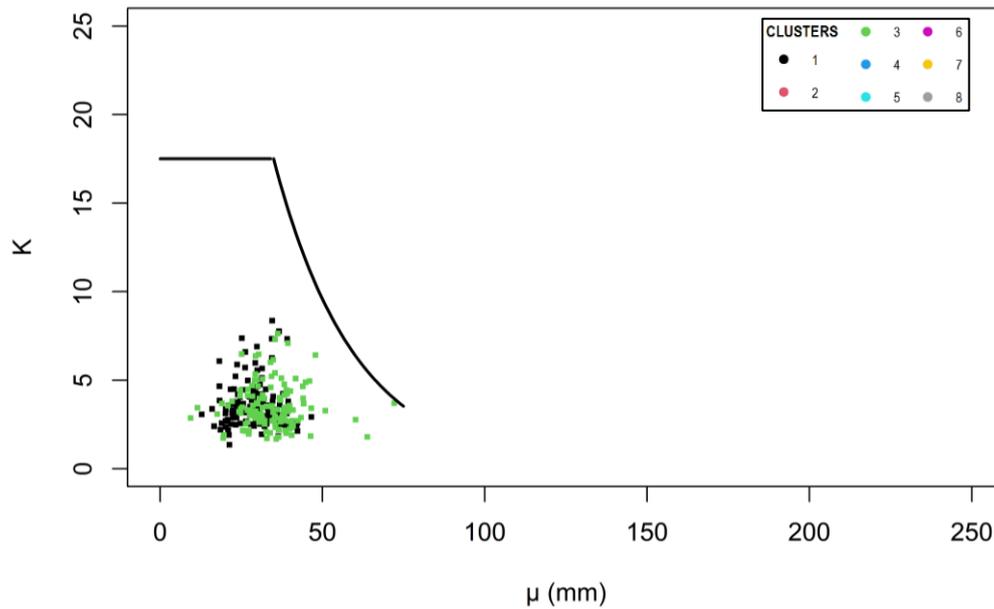
Código ANA	Nombre	Código ANA	Nombre
155212	PARQUIN	383	LAMAS
6200	SALCABMABA	152135	SAN JUAN DE LOS ALISOS
115129	MACHAHUAY	152156	ANIA CABUYAL
157310	ANDAHUA	150112	MATAPALO
736	PUQUIO	152118	CANCHAQUE
777	MACUSANI	385	SAUCE
737	ANDAMARCA02	248	HUARMACA
156119	TOTORA	155421	PERENE
677	CARAHUASI	476	PICHANAKY
503	JAUJA	237	AYABACA
860	CHUAPALCA	211	SAPOSOA
778	PROGRESO	234	LA TINA
785	MUÑANI	153226	SORITOR
690	CCATCCA	3308	ALAO
150901	MALVAS	150001	SAN PEDRO02
152109	TOMA DE ZAMBA	153314	SAN ANTONIO01
151503	HUACHOS	571	SATIPO
155200	PACCHO	5572	VILLA RICA
155228	SAN PEDRO DE CHUCLU	556	POZUZO
155231	INGENIO	152142	AUL C MEMBRILLO
391	JESUS	154305	MASISEA
659	ACOMBAMBA	152409	SANTA MARIA DE NANAY
635	HUAYAO	387	CONTAMANA
156102	SAN LAZARO DE ESCOMARCA	152153	ALAMOR
156111	VILCA	152140	PICO DE LORO
156204	URUBAMBA01	139	CABO INGA
157311	ORCOPAMPA	152404	SILVA MERINO
560	COMAS	153322	PELEJO
756	COMBAPATA	152402	BETRAÑA
477	SANTA ANA	151	MORONA
648	PILCHACA	177	SANTA CLOTILDE
151506	SANTIAGO DE HUIRPACANCHA	173	MUYUY
605	ABANCAY	278	SAN RAMON01
158326	CAPAZO	447	TOCACHE
844	PISAC	150003	EL CIRUELO
154110	PIRA	281	GENARO HERRERA
878	MAZO CRUZ	291	CABALLOCOCHA
155407	HUARAZ	180	PEBAS
398	SALPO	453	LA DIVISORIA
664	WUAYLLAPAMPA	150206	SANTA CRUZ01
803	CRUCERO ALTO	480	TOURNAVISTA
158308	CALACOA	153315	SHANUSI
157209	CERRO CONDORILLO	172	TAMSHIYACU
625	ACOSTAMBO	478	PUERTO INCA
795	CABANAONDE	465	CACHICOTO

Código ANA	Nombre	Código ANA	Nombre
112190	HUANTA GORE	150208	PUNCHANA
156122	TAMBO	153400	TIRUNTAN
157207	PAMPAMARCA	150201	MANITI
657	LIRCAY	2412	SAN ALEJANDRO
155115	CARHUACAYAN	140406	TINGO MARIA
157733	RAMIS	446	TANANTA
607	GRANJA KCAYRA	150207	LA LIBERTAD
155101	TAPACOCHA	459	EL MARONAL
239	HUANCABAMBA01	466	LA MORADA
153218	LA LLICA	150204	TIMICURILLO
824	PUTINA	6671	CHONTACHACA
155217	LACHAQUI	153320	PUCALLPA01
362	BAMBAMARCA	6401	SAN GABAN
759	SICUANI	462	AGUAYTIA
156307	CHITAPAMPA	693	QUINCEMIL

Fuente: Elaboración propia.

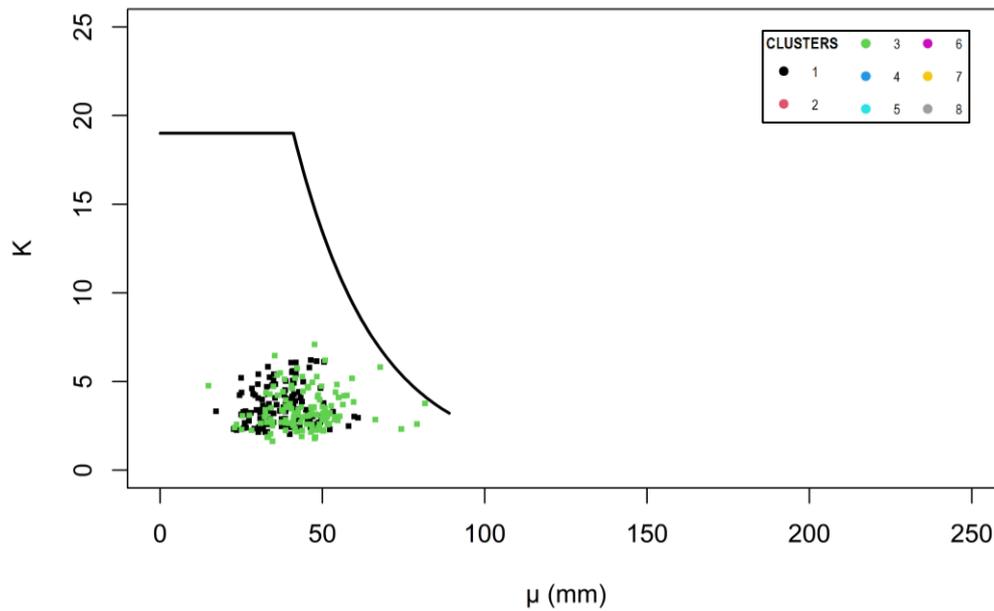
**ANEXO C: CURVAS ENVOLVENTES DE K PARA CADA GRUPO DE
CLUSTERS PARA 24; 48 Y 72 HORAS**

Gráfica 1 Envoltente de los cluster 1 y 3 para 24 horas



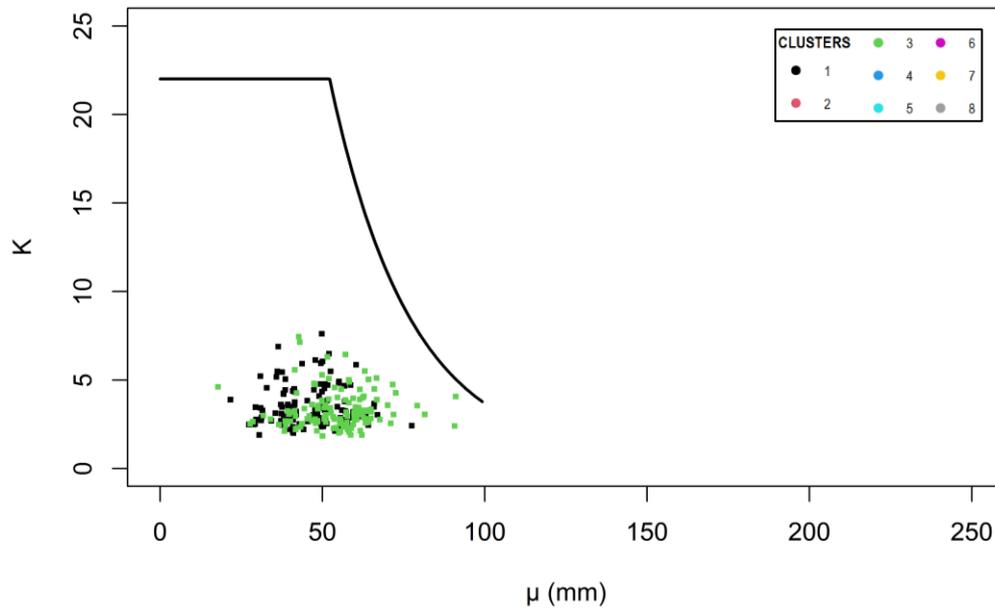
Fuente: Elaboración propia.

Gráfica 2 Envoltente de los cluster 1 y 3 para 48 horas



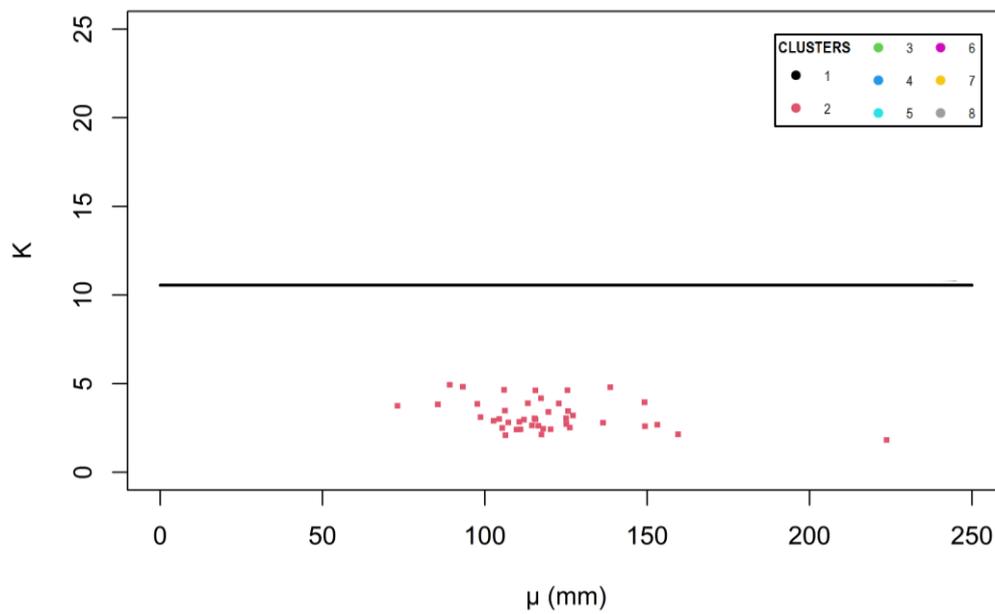
Fuente: Elaboración propia.

Gráfica 3 Envoltente de los cluster 1 y 3 para 72 horas



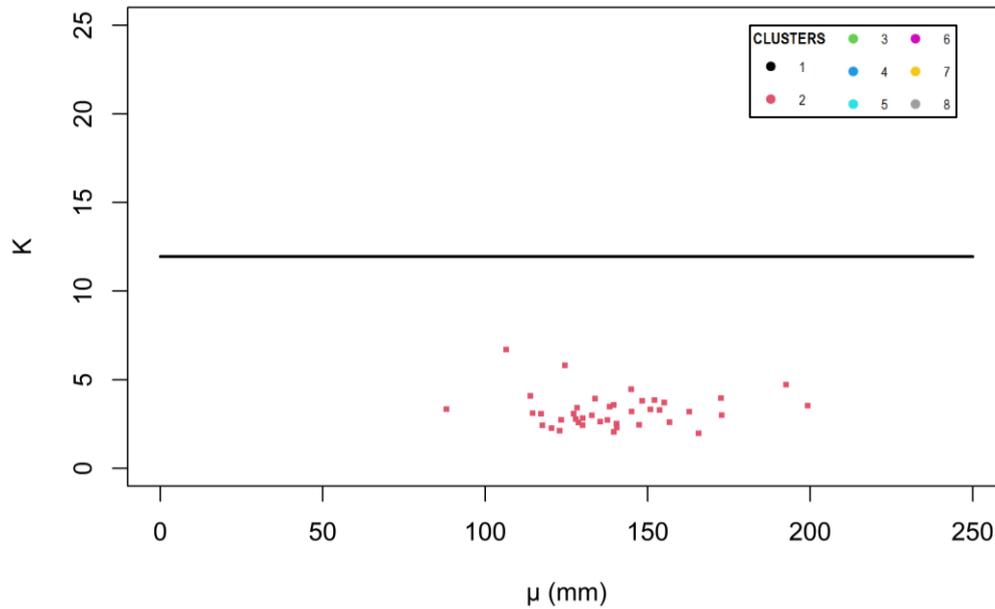
Fuente: Elaboración propia.

Gráfica 4 Envoltente del cluster 2 para 24 horas



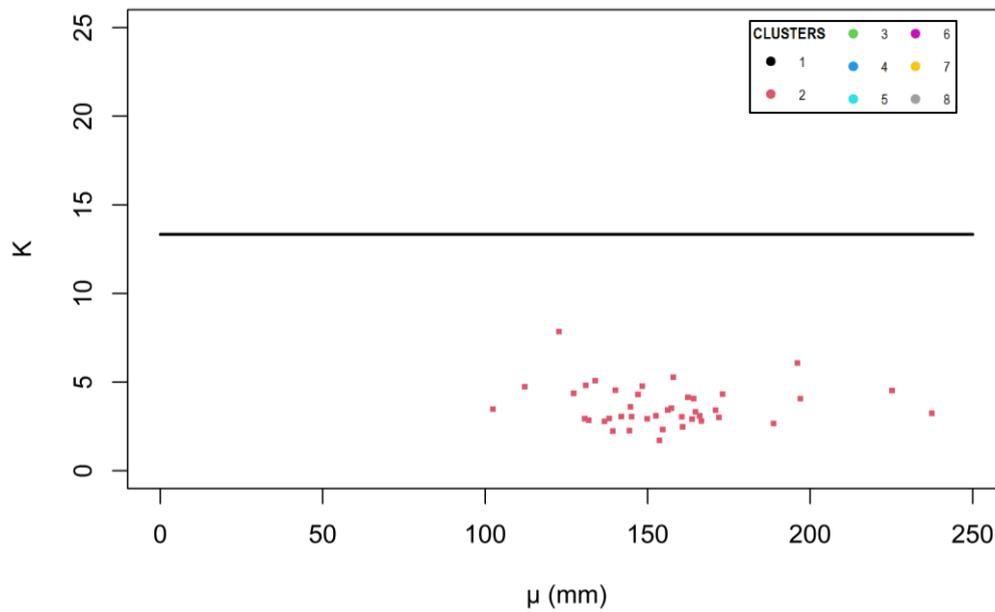
Fuente: Elaboración propia.

Gráfica 5 Envoltente del cluster 2 para 48 horas



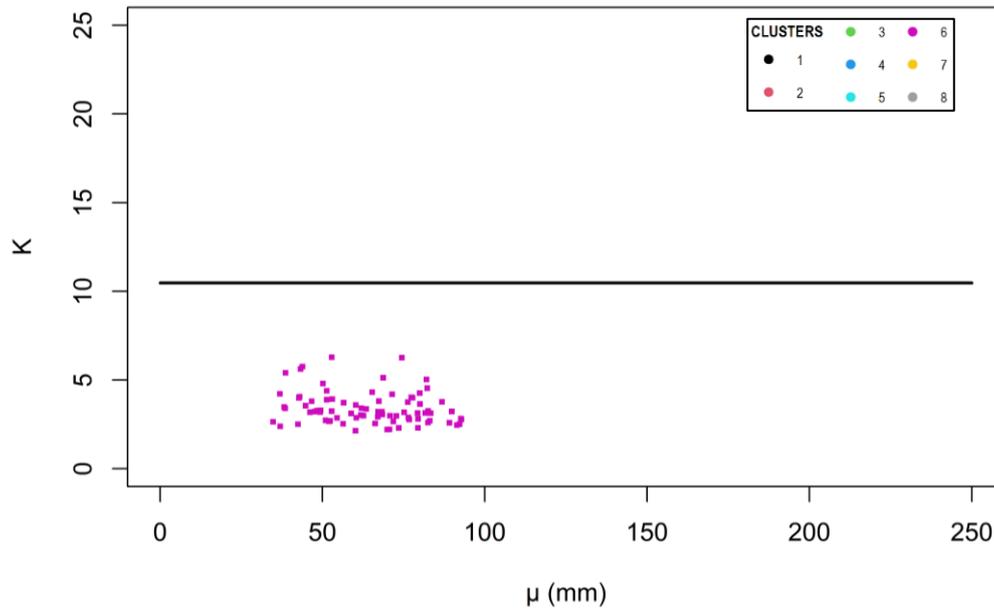
Fuente: Elaboración propia.

Gráfica 6 Envoltente del cluster 2 para 72 horas



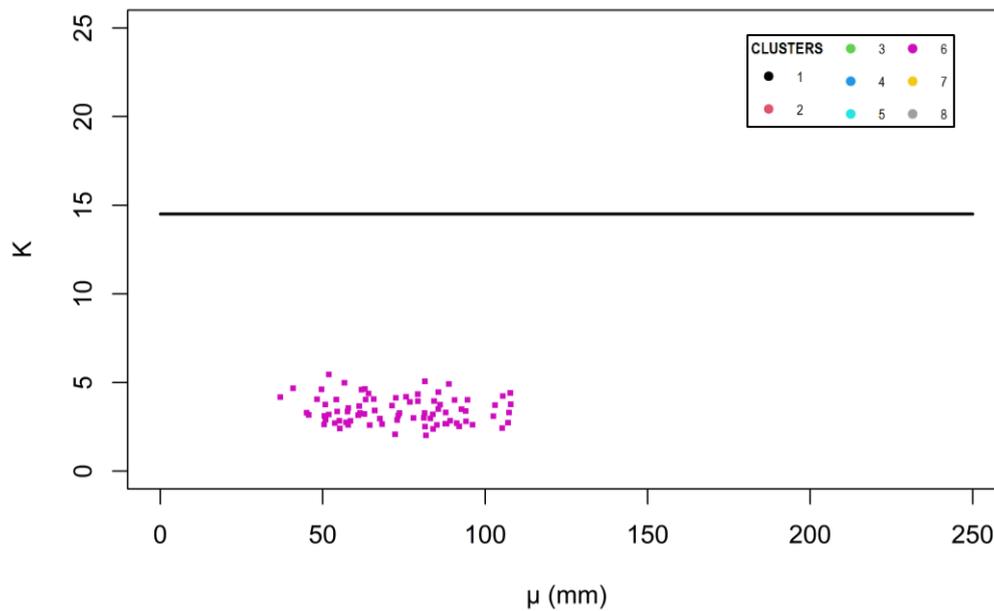
Fuente: Elaboración propia.

Gráfica 7 Envoltente del cluster 6 para 24 horas



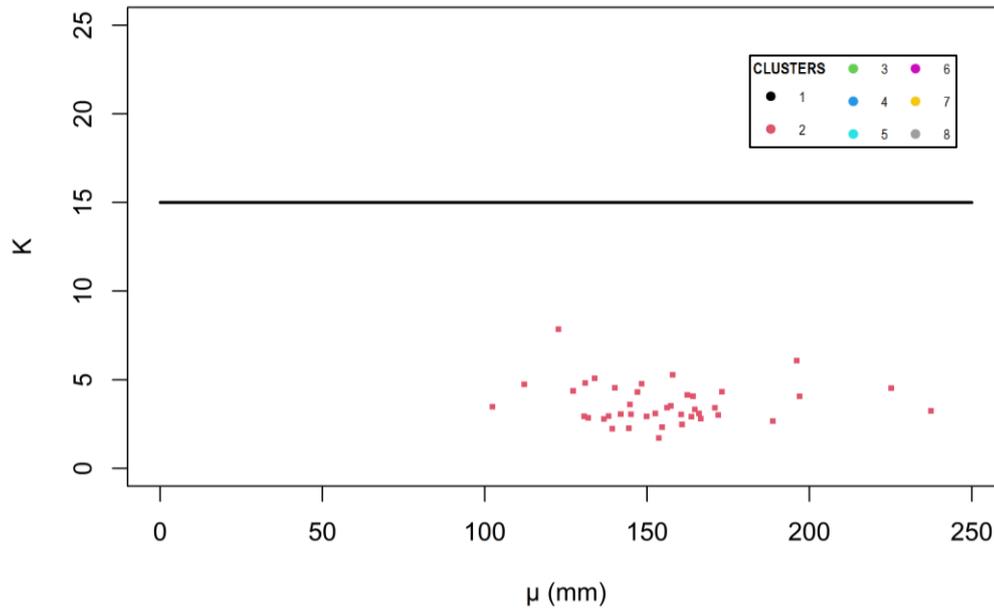
Fuente: Elaboración propia.

Gráfica 8 Envoltente del cluster 6 para 48 horas



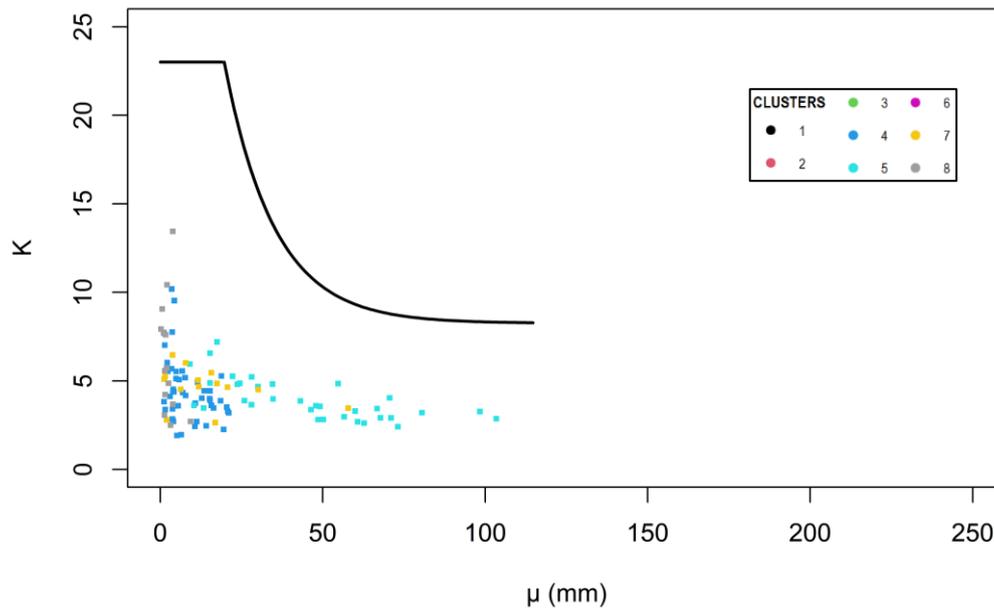
Fuente: Elaboración propia.

Gráfica 9 Envoltente del cluster 6 para 72 horas



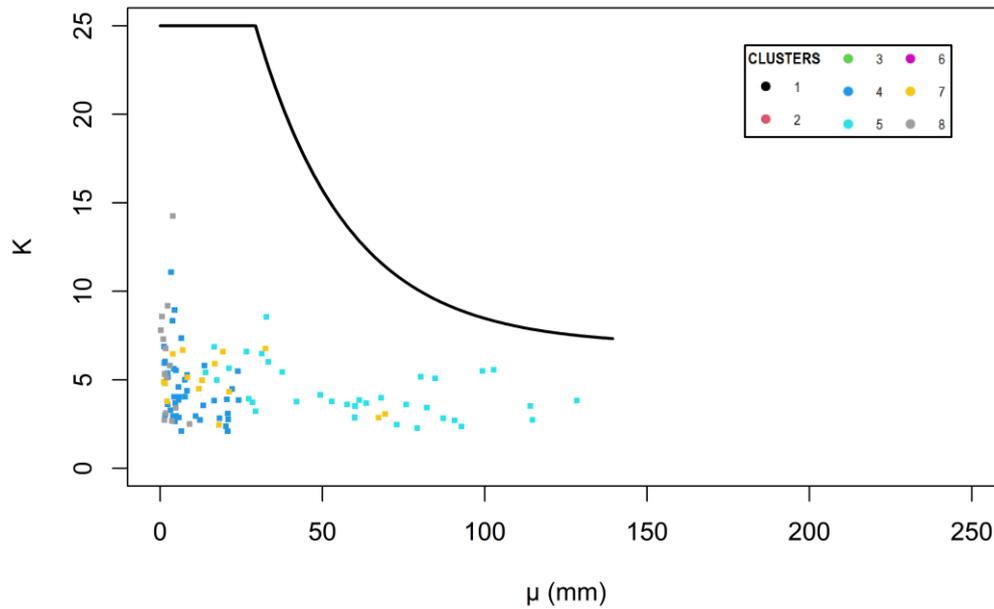
Fuente: Elaboración propia.

Gráfica 10 Envoltente del cluster 4; 5; 7 y 8 para 24 horas



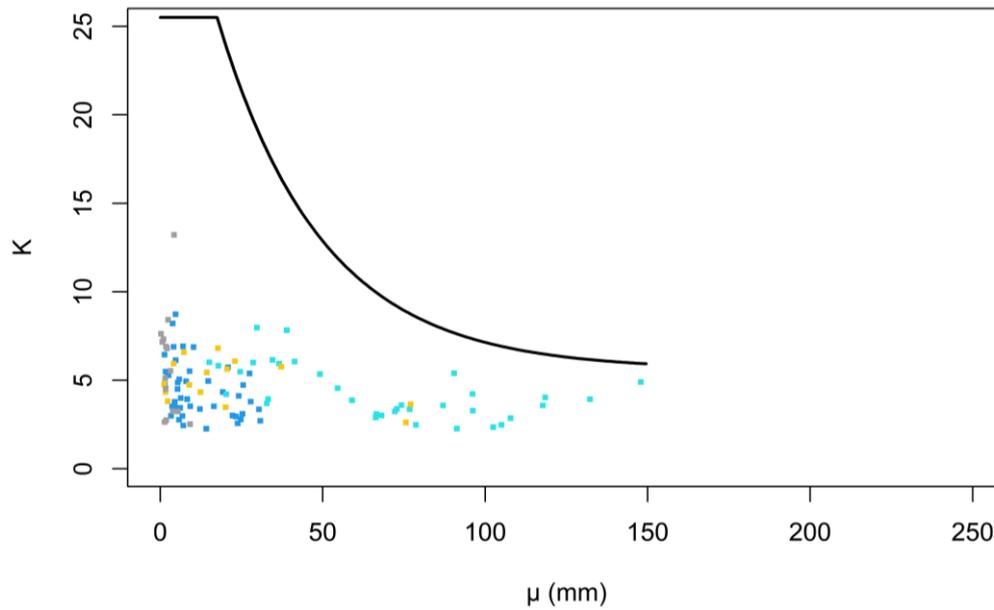
Fuente: Elaboración propia.

Gráfica 11 Envoltente del cluster 4; 5; 7 y 8 para 48 horas



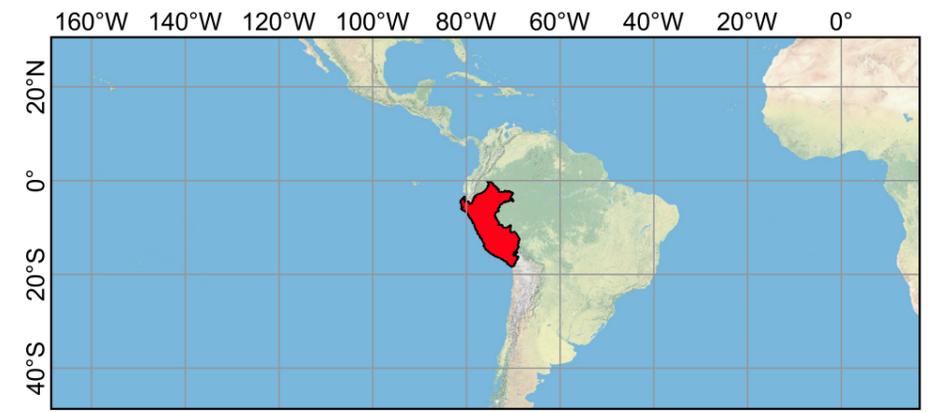
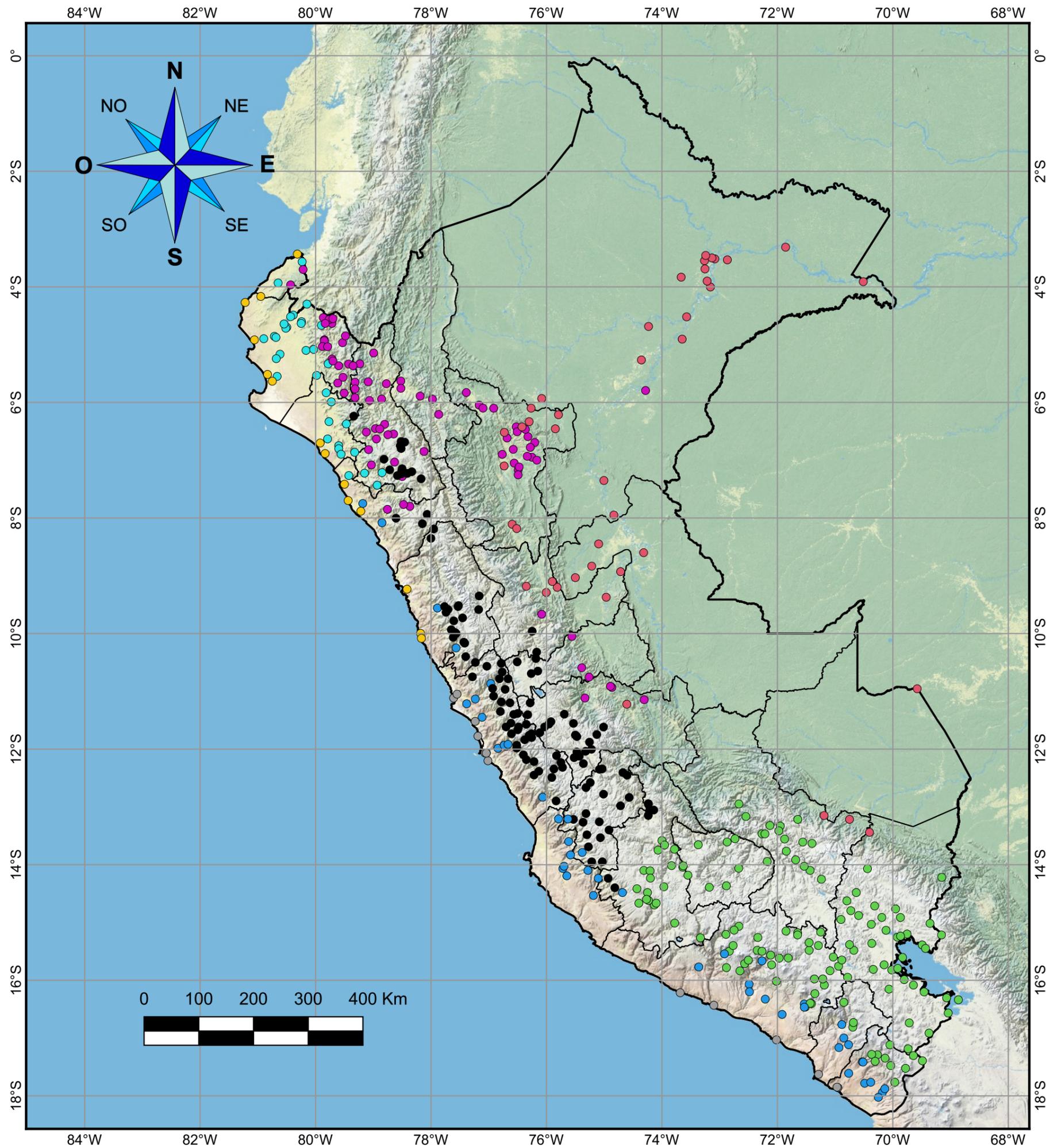
Fuente: Elaboración propia.

Gráfica 12 Envoltente del cluster 4; 5; 7 y 8 para 72 horas



Fuente: Elaboración propia.

ANEXO D: MAPAS RELACIONADOS A LA DURACIÓN DE 24 HORAS



MAPA DE LOCALIZACIÓN
ESCALA : 1 / 150 000 000

LEYENDA		
Clusters para el caso de 24 horas	● 3	● 6
● 1	● 4	● 7
● 2	● 5	● 8

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA
ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
CLUSTERS DE ESTACIONES PARA EL ANÁLISIS DE LA PMP DE 24 HORAS

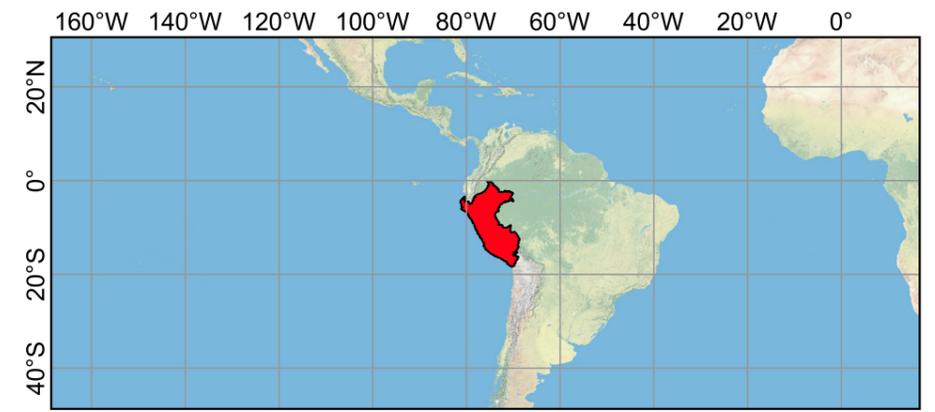
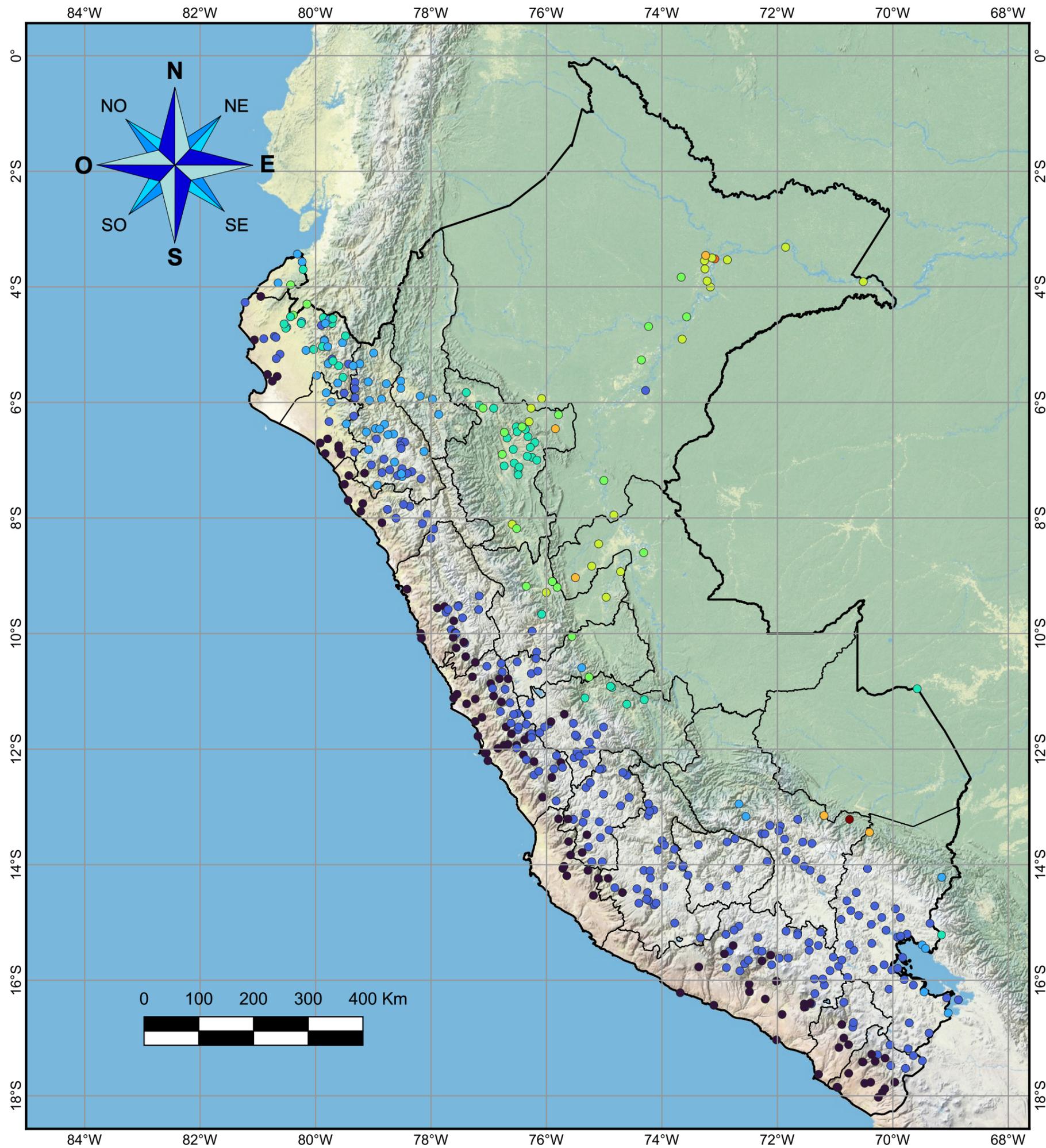
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

LIMA - PERÚ 2024

Nº MAPA:
D.1



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Precipitación media máxima anual de 24 horas (mm)	● 23 - 45	● 90 - 112	● 157 - 179
	● 45 - 67	● 112 - 134	● 179 - 201
	● 67 - 90	● 134 - 157	● 201 - 224
	● 0 - 23		

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MEDIA MÁXIMA ANUAL DE 24 HORAS PUNTUAL

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

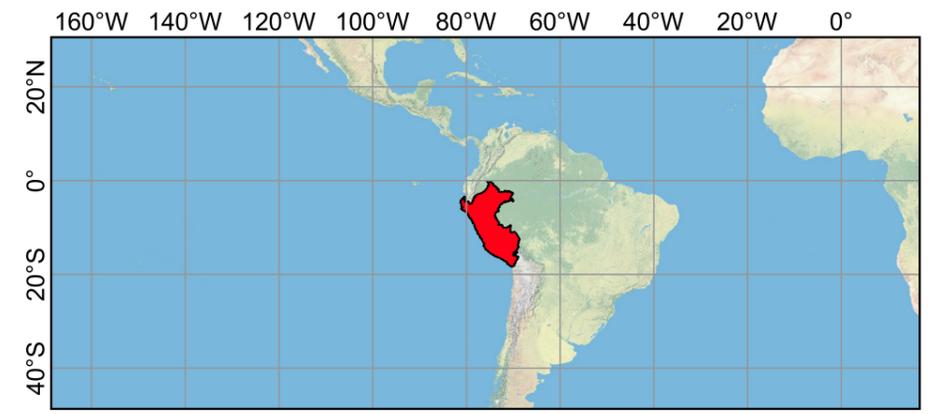
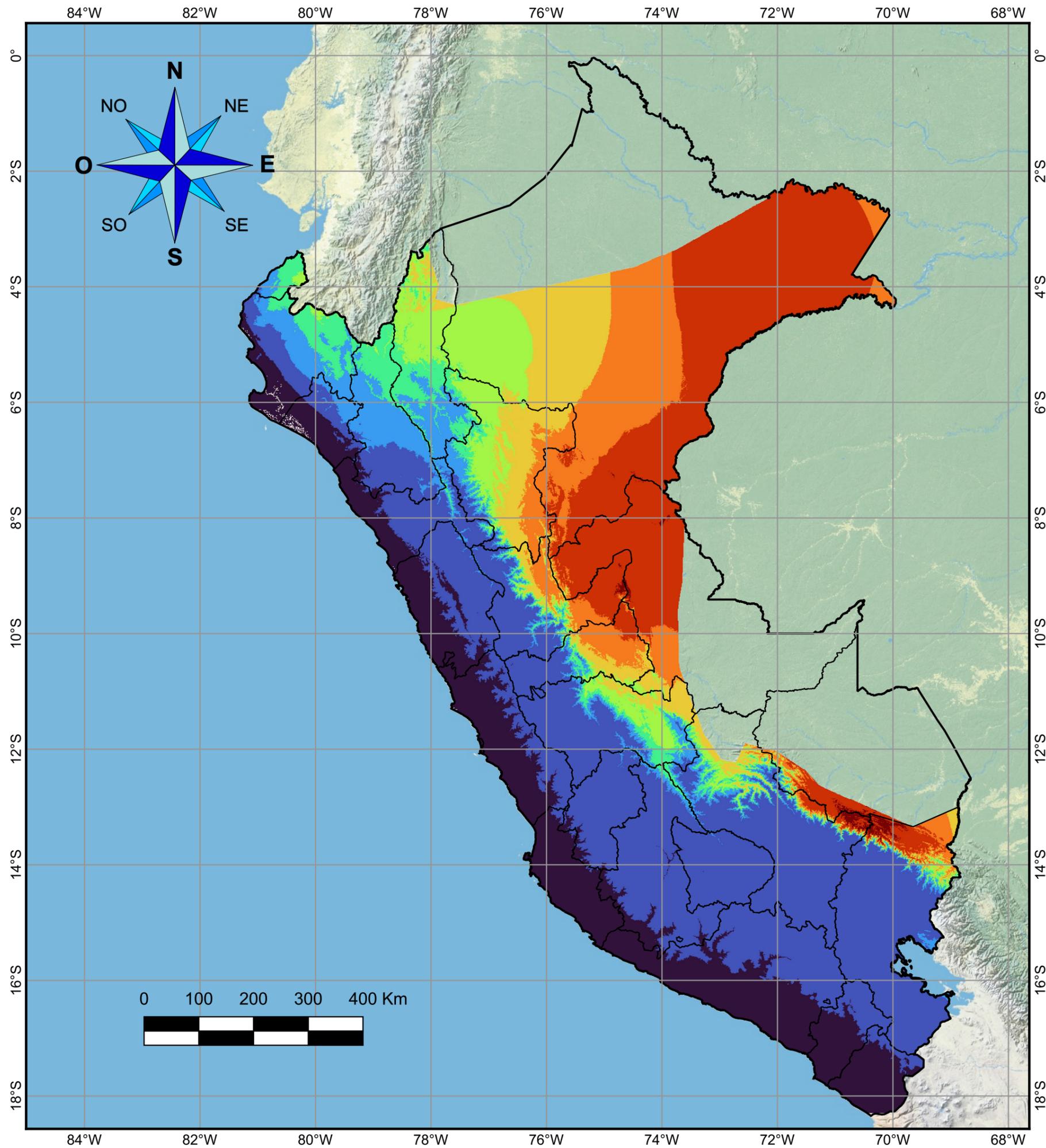
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

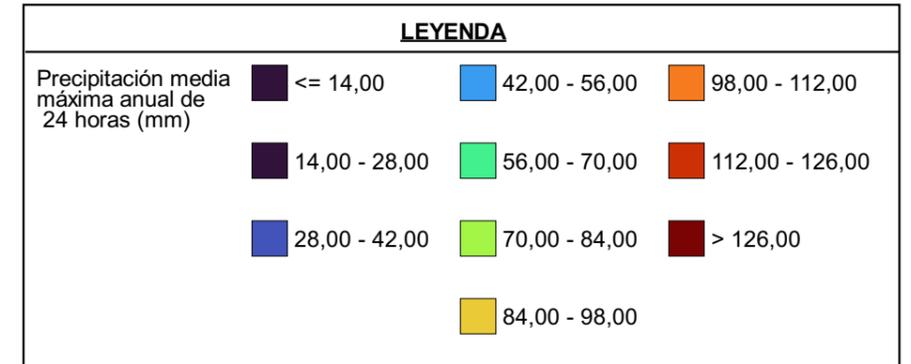
Nº MAPA:

D.2



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

PLANTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MEDIA MÁXIMA ANUAL DE 24 HORAS RASTERIZADA

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

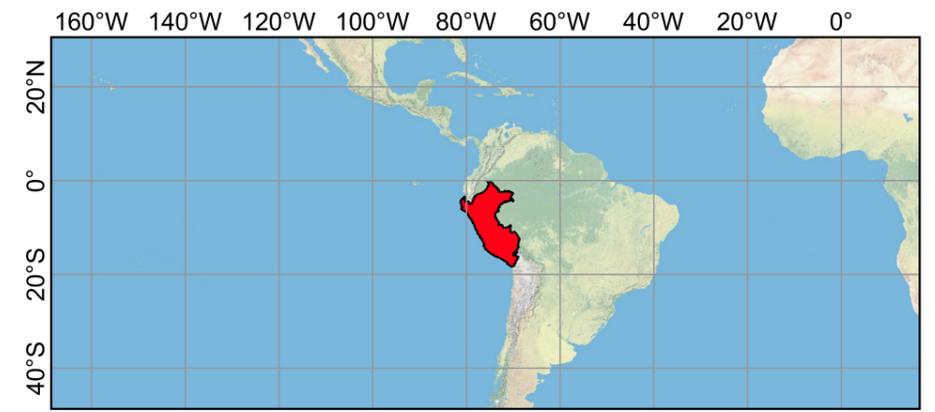
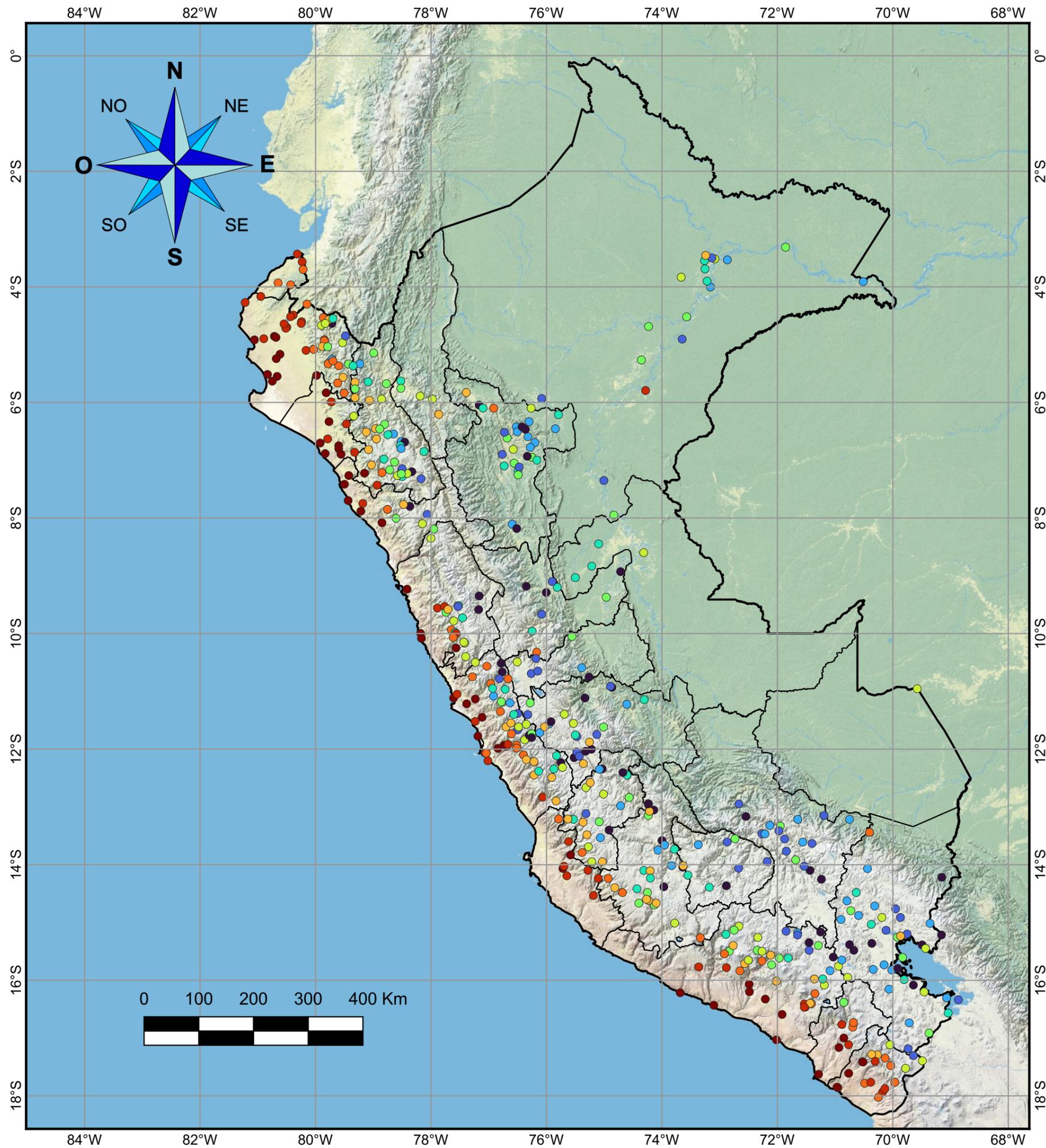
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

D.3



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Coeficiente de variación de la precipitación máxima anual de 24h	● 0.229 - 0.258	● 0.306 - 0.333	● 0.456 - 0.638
	● 0.258 - 0.284	● 0.333 - 0.376	● 0.638 - 0.969
	● 0.141 - 0.229	● 0.376 - 0.456	● 0.969 - 2.665

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

COEFICIENTE DE VARIACIÓN DE LA PRECIPITACIÓN MÁXIMA ANUAL DE 24 HORAS PUNTUAL

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.

“El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”

Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

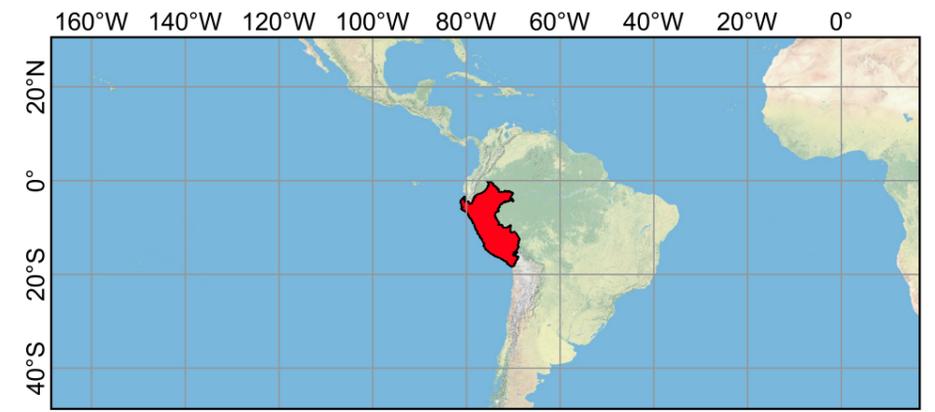
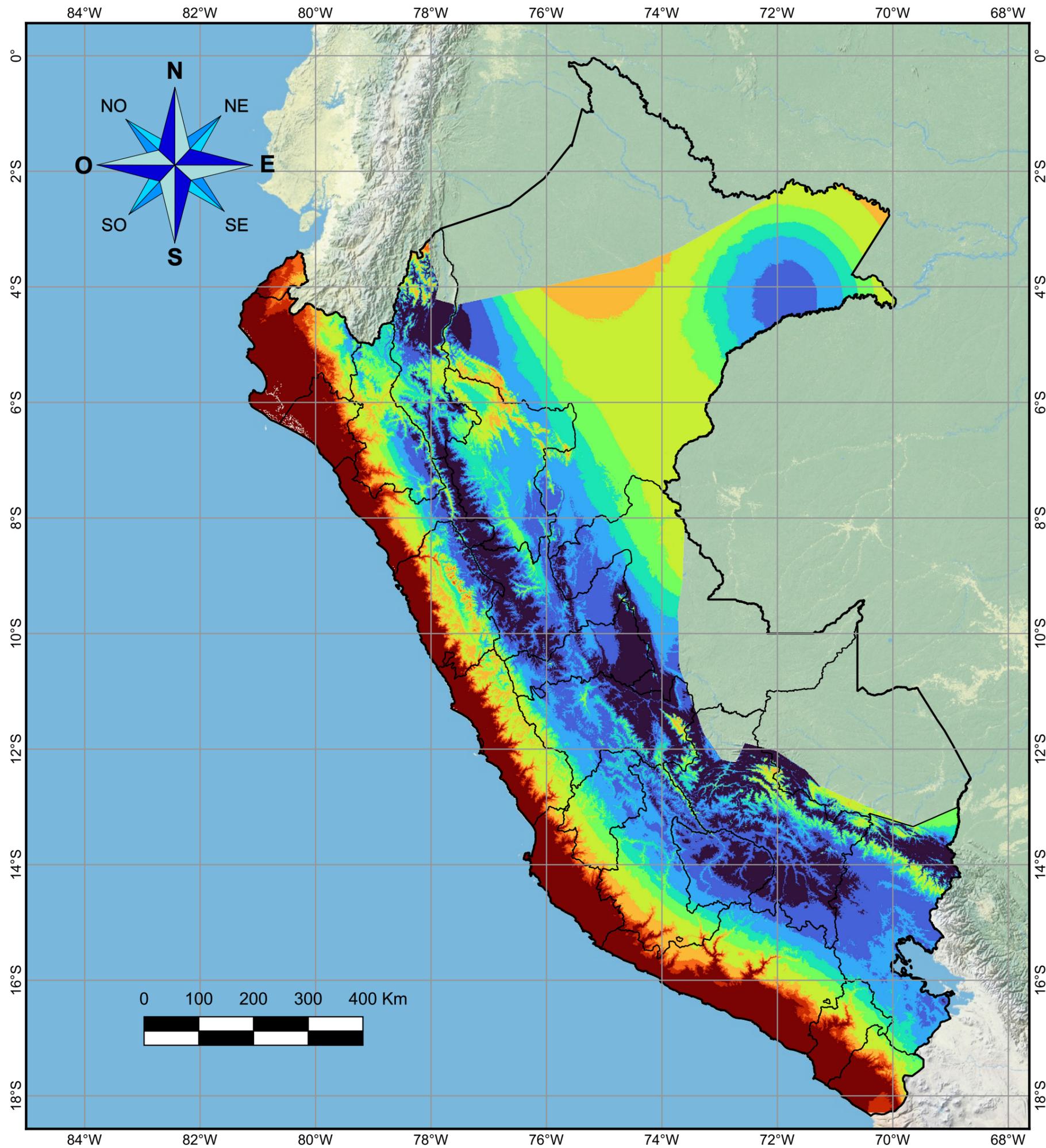
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

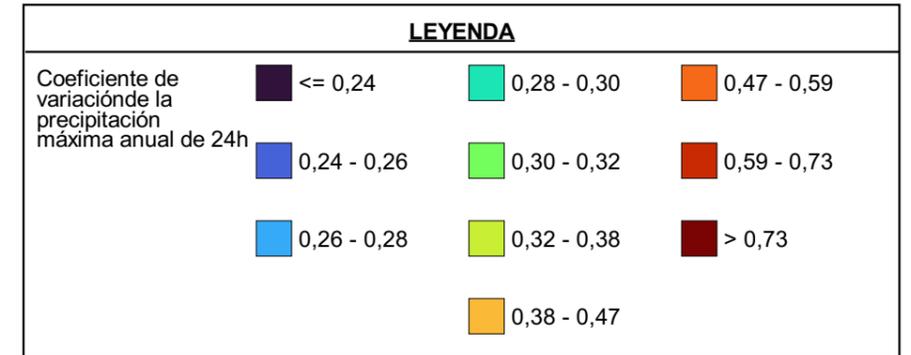
Nº MAPA:

D.4



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

COEFICIENTE DE VARIACIÓN DE LA PRECIPITACIÓN MÁXIMA ANUAL DE 24 HORAS RASTERIZADO

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

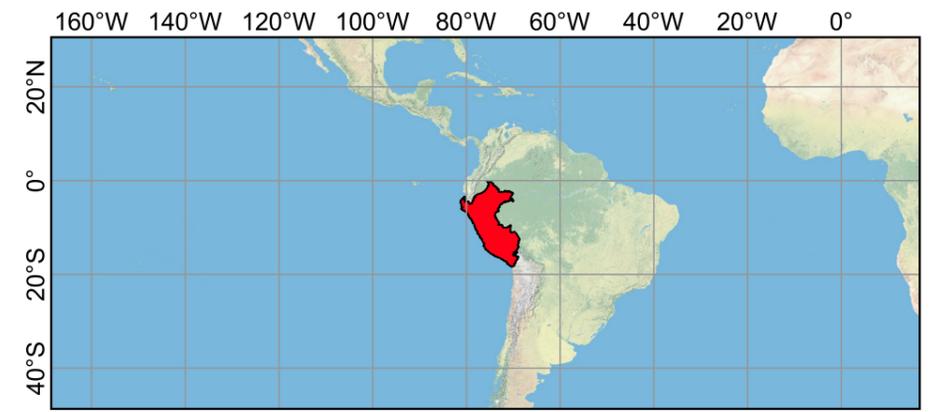
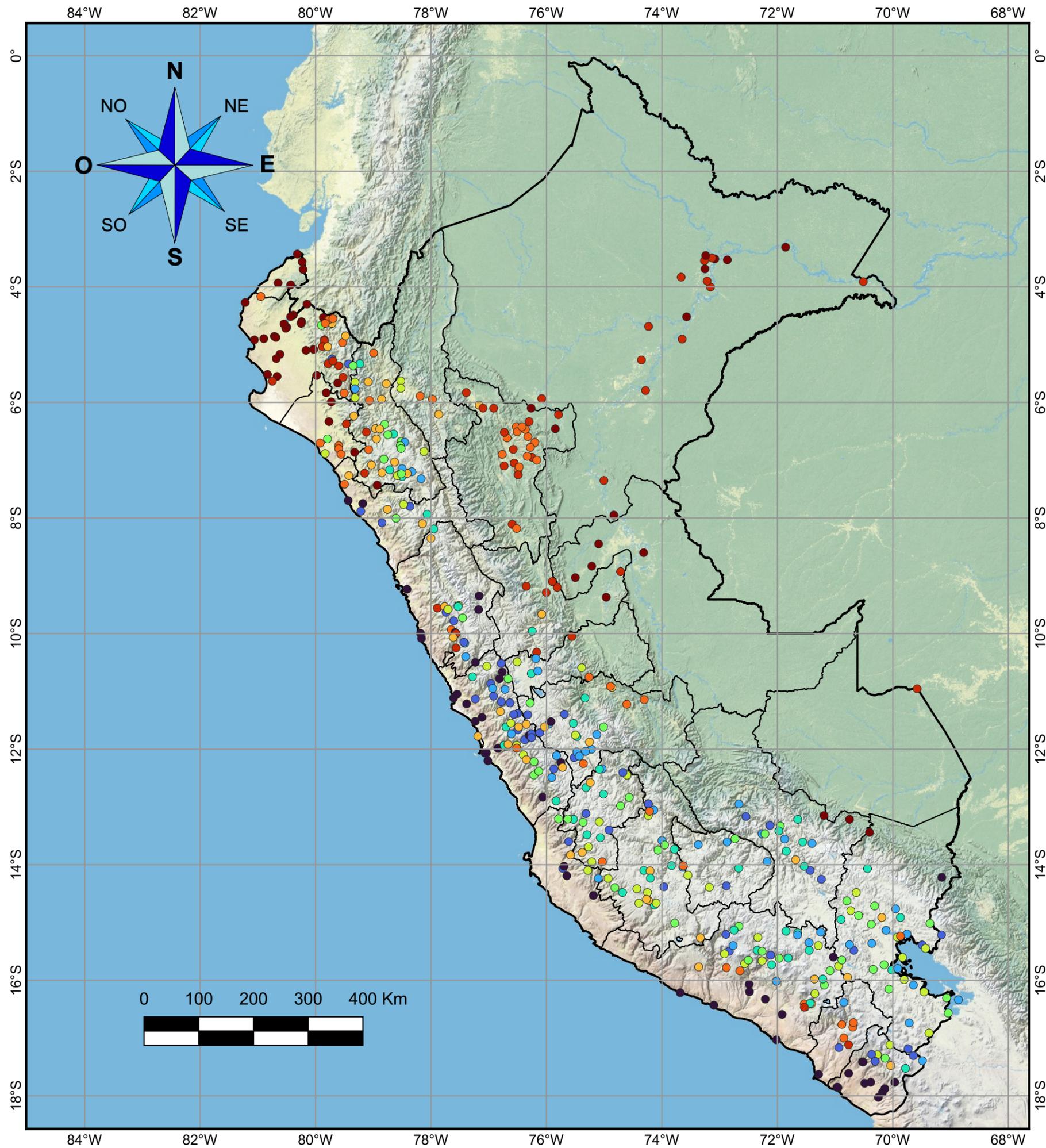
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

D.5



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Precipitación Máxima Probable (PMP) de 24 horas (mm)	● 120 - 159	● 195 - 210	● 274 - 341
	● 159 - 178	● 210 - 230	● 341 - 486
	● 178 - 195	● 230 - 274	● 486 - 1333
	● 11 - 120		

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
PRECIPITACIÓN MÁXIMA PROBABLE DE 24 HORAS PUNTUAL

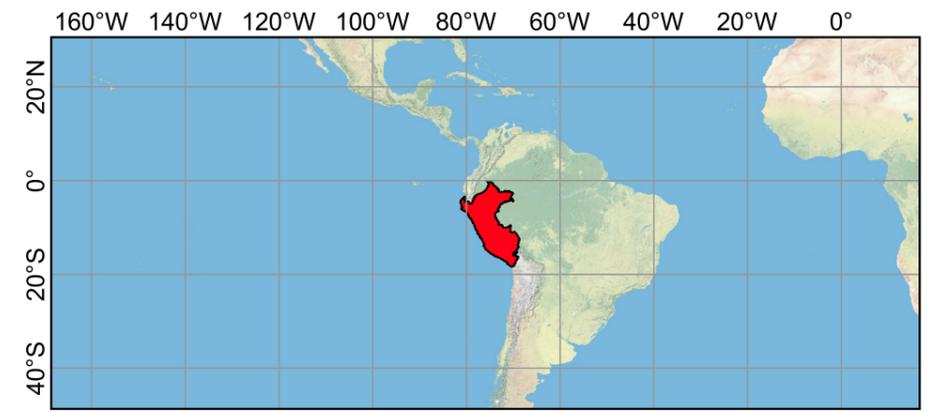
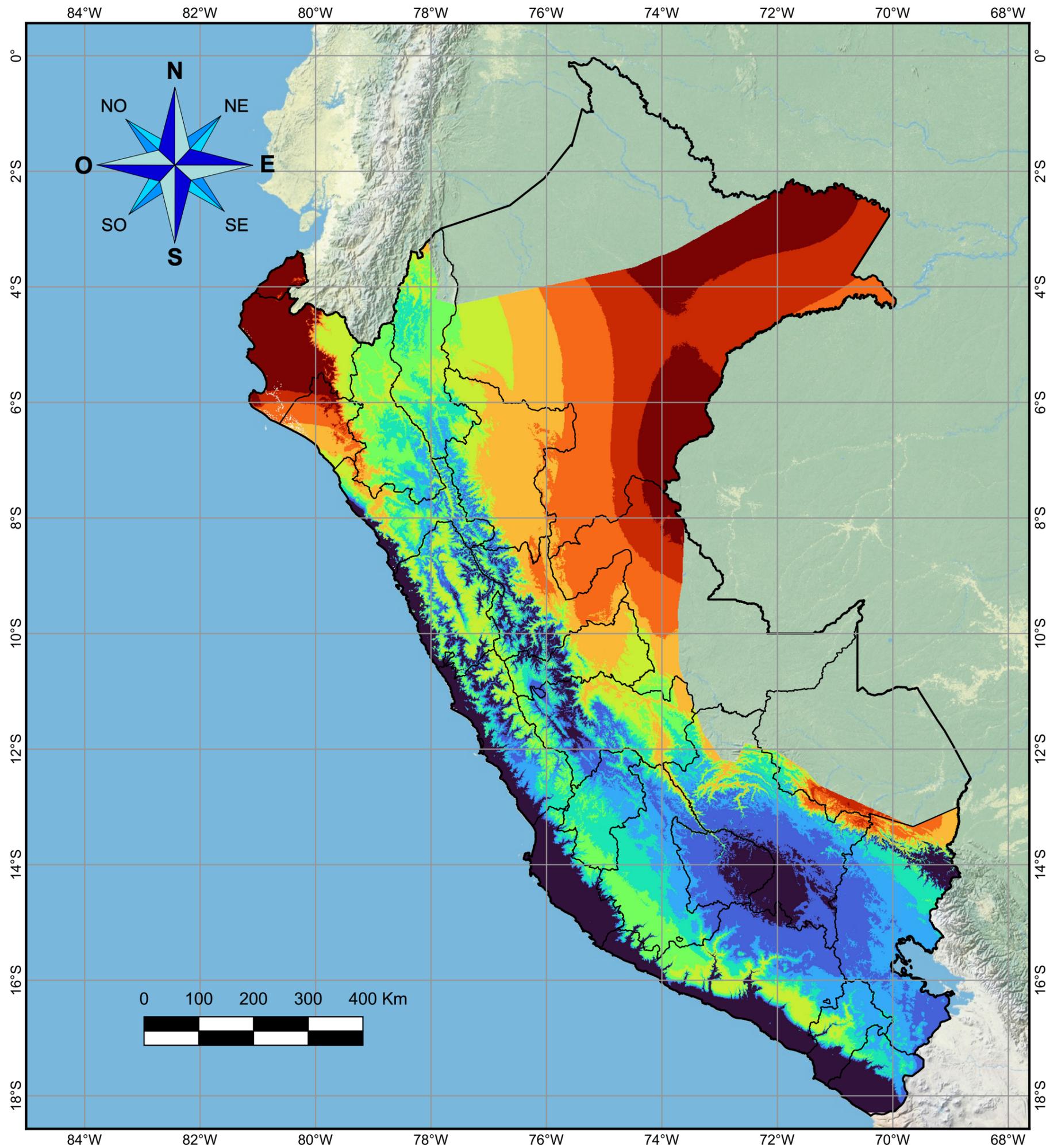
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

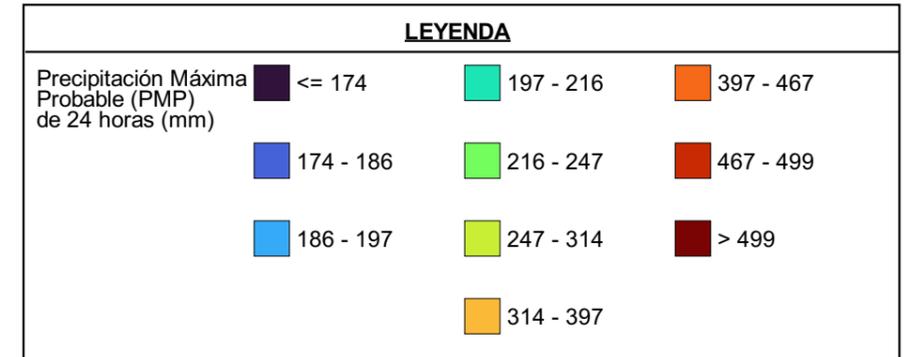
LIMA - PERÚ 2024

Nº MAPA:
D.6



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MÁXIMA PROBABLE DE 24 HORAS RASTERIZADA

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:

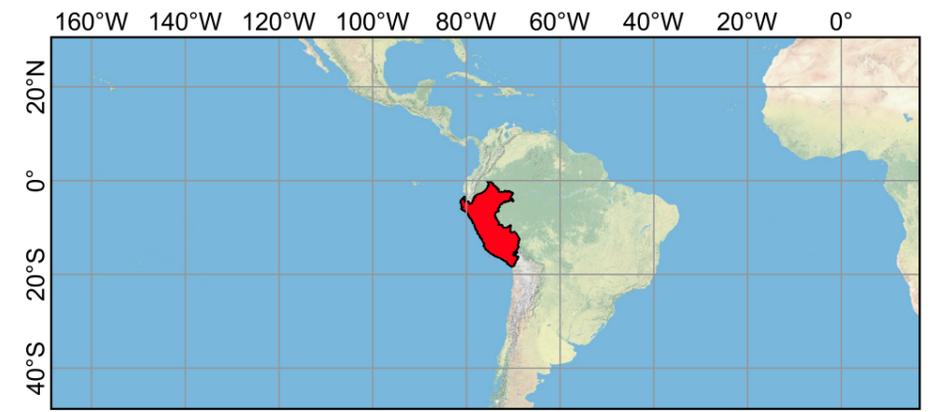
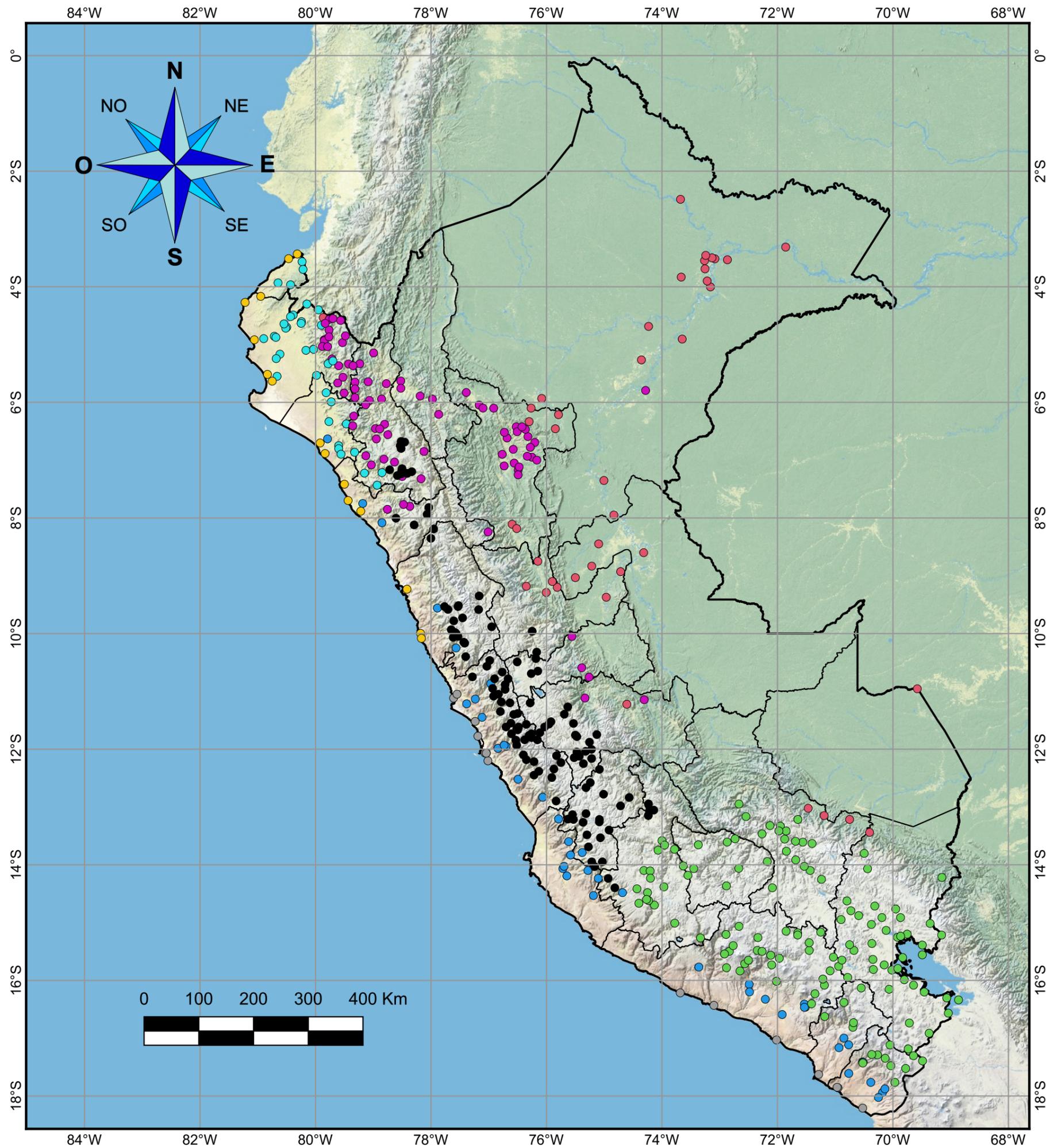
WGS 84

LIMA - PERÚ 2024

Nº MAPA:

D.7

ANEXO E: MAPAS RELACIONADOS A LA DURACIÓN DE 48 HORAS



MAPA DE LOCALIZACIÓN
ESCALA : 1 / 150 000 000

LEYENDA		
Clusters para el caso de 48 horas	● 3	● 6
● 1	● 4	● 7
● 2	● 5	● 8

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA
ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
CLUSTERS DE ESTACIONES PARA EL ANALISIS DE LA PMP DE 48 HORAS

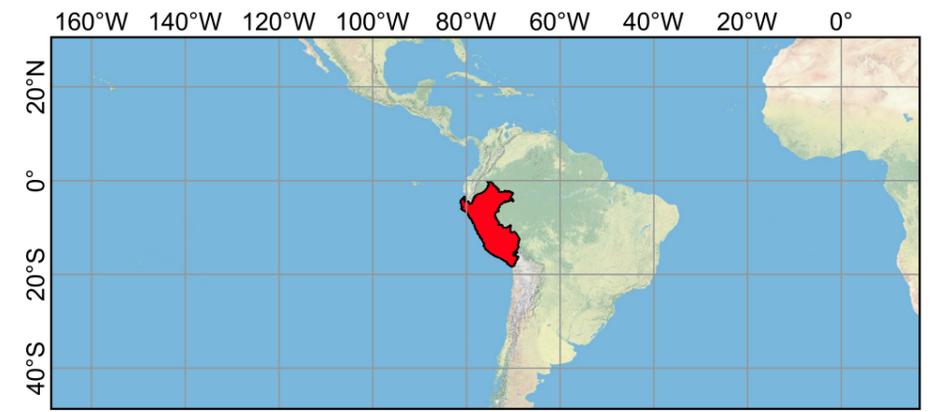
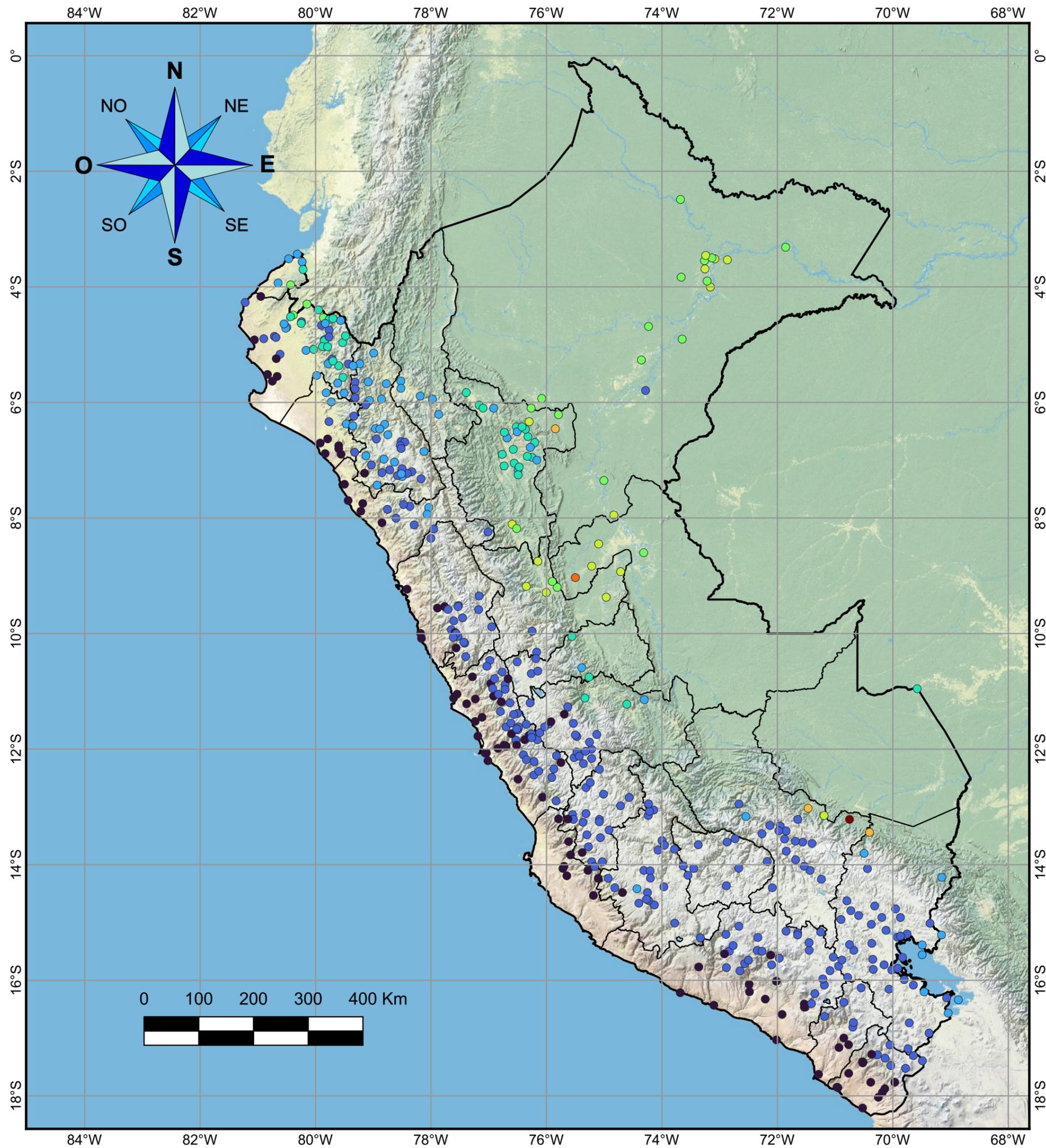
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

LIMA - PERÚ 2024

Nº MAPA:
E.1



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Precipitación media máxima anual de 48 horas (mm)	● 28 - 55	● 111 - 138	● 193 - 221
	● 55 - 83	● 138 - 166	● 221 - 249
	● 83 - 111	● 166 - 193	● 249 - 276
	● 0 - 28		

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MEDIA MÁXIMA ANUAL DE 48 HORAS PUNTUAL

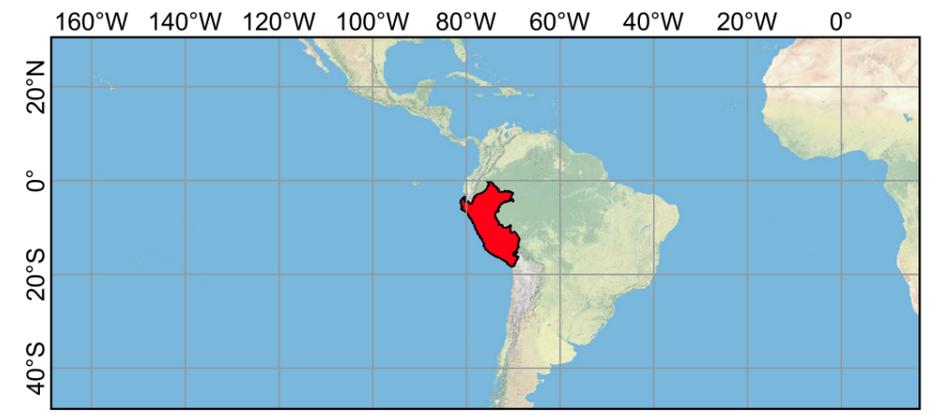
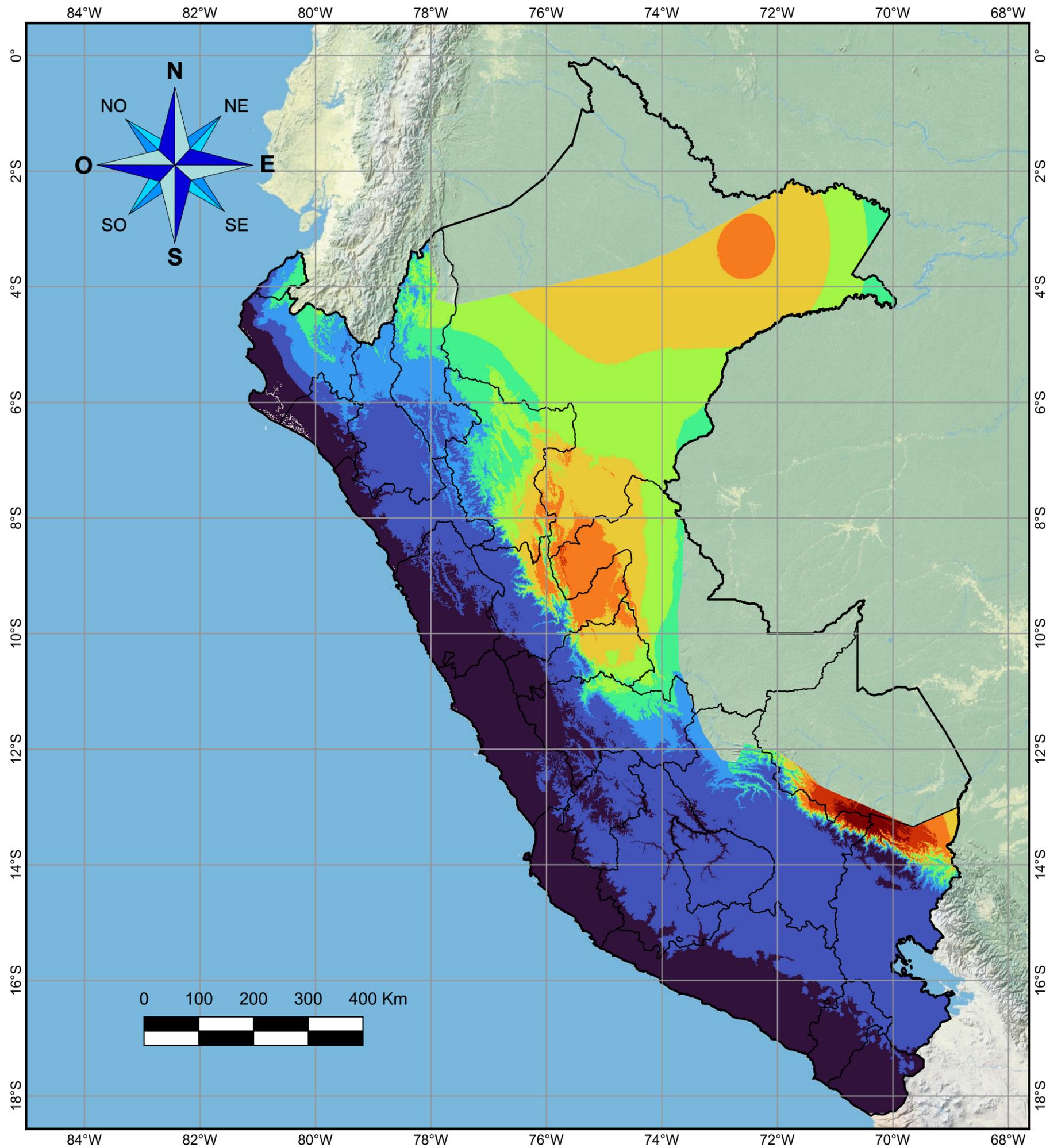
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
 DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
 WGS 84

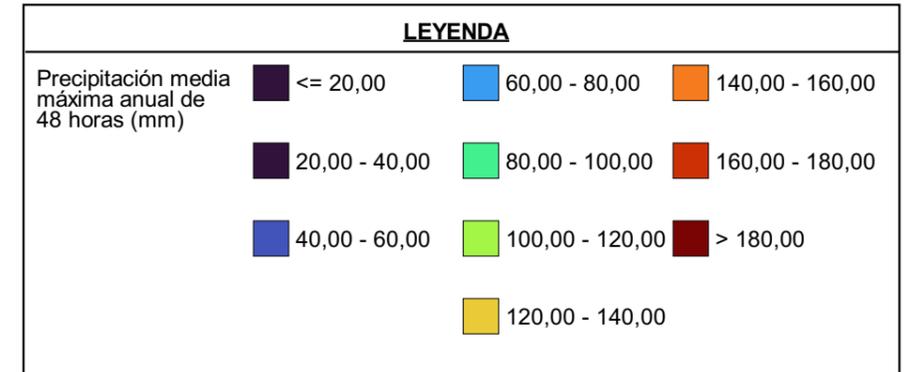
LIMA - PERÚ 2024

Nº MAPA:
 E.2



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
 “REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:
 PRECIPITACIÓN MEDIA MÁXIMA ANUAL DE 48 HORAS RASTERIZADA

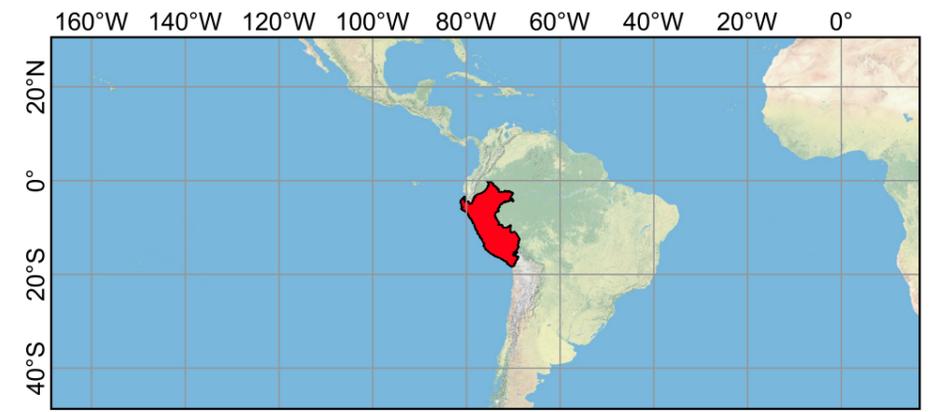
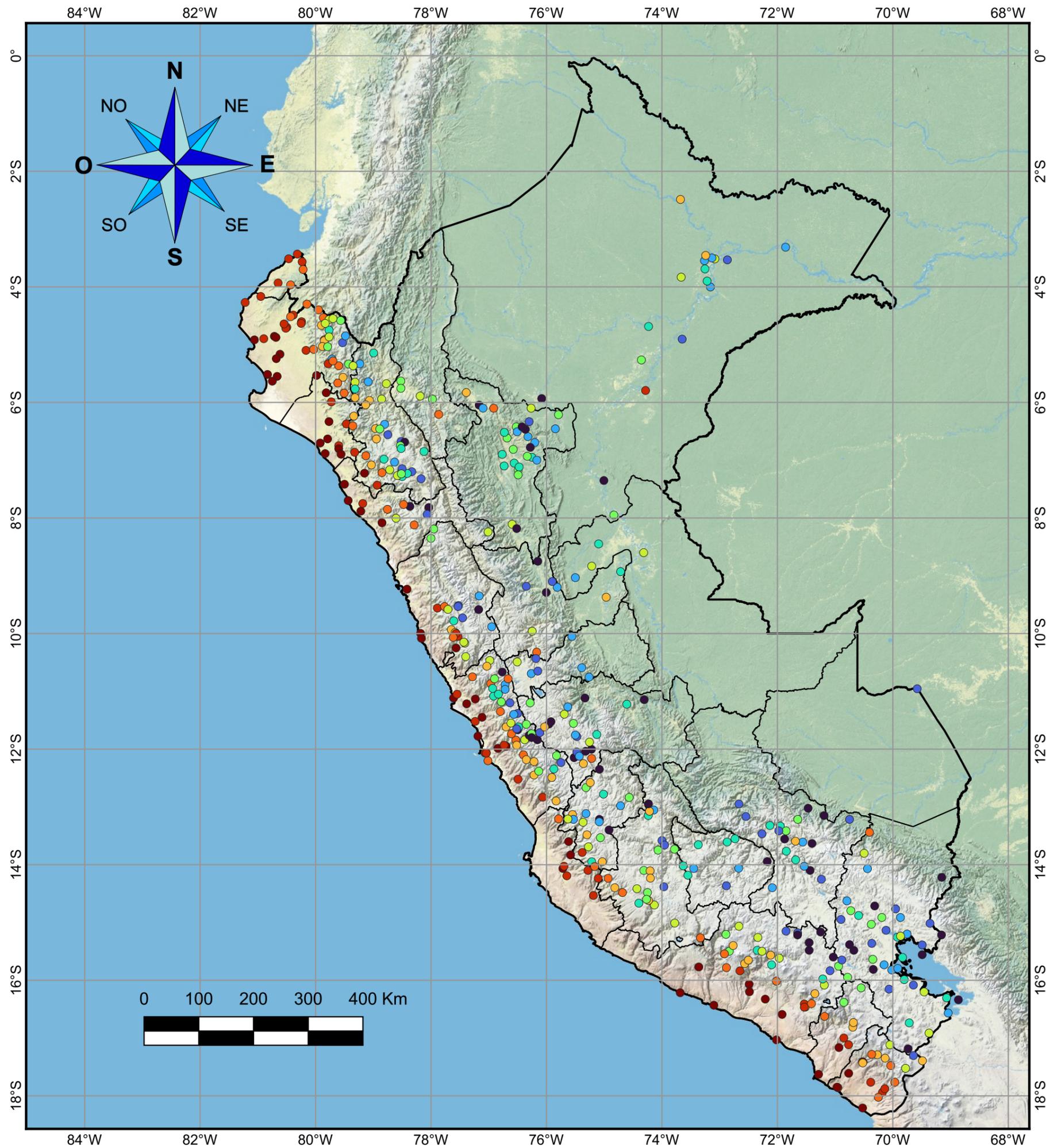
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
 DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
 WGS 84

LIMA - PERÚ 2024

Nº MAPA:
 E.3



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Coeficiente de variación de la precipitación máxima anual de 48h	● 0,228 - 0,253	● 0,292 - 0,321	● 0,468 - 0,632
	● 0,253 - 0,268	● 0,321 - 0,371	● 0,632 - 0,972
	● 0,172 - 0,228	● 0,371 - 0,468	● 0,972 - 2,683
	● 0,268 - 0,292		

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:
COEFICIENTE DE VARIACIÓN DE LA PRECIPITACIÓN MÁXIMA ANUAL DE 48 HORAS PUNTUAL

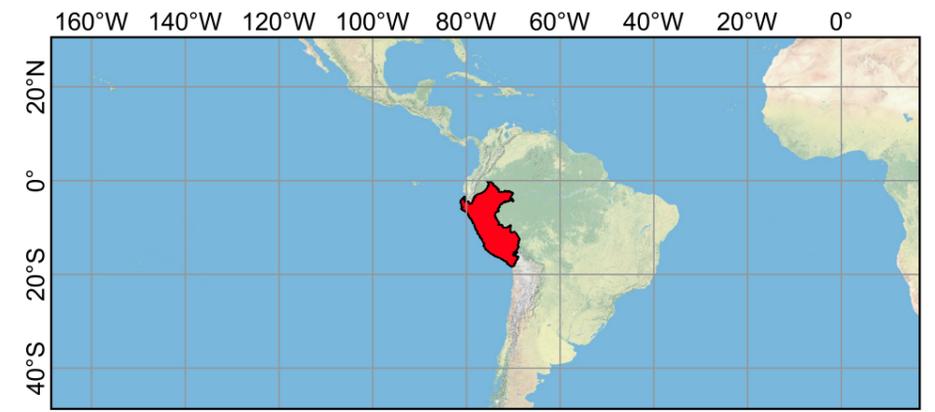
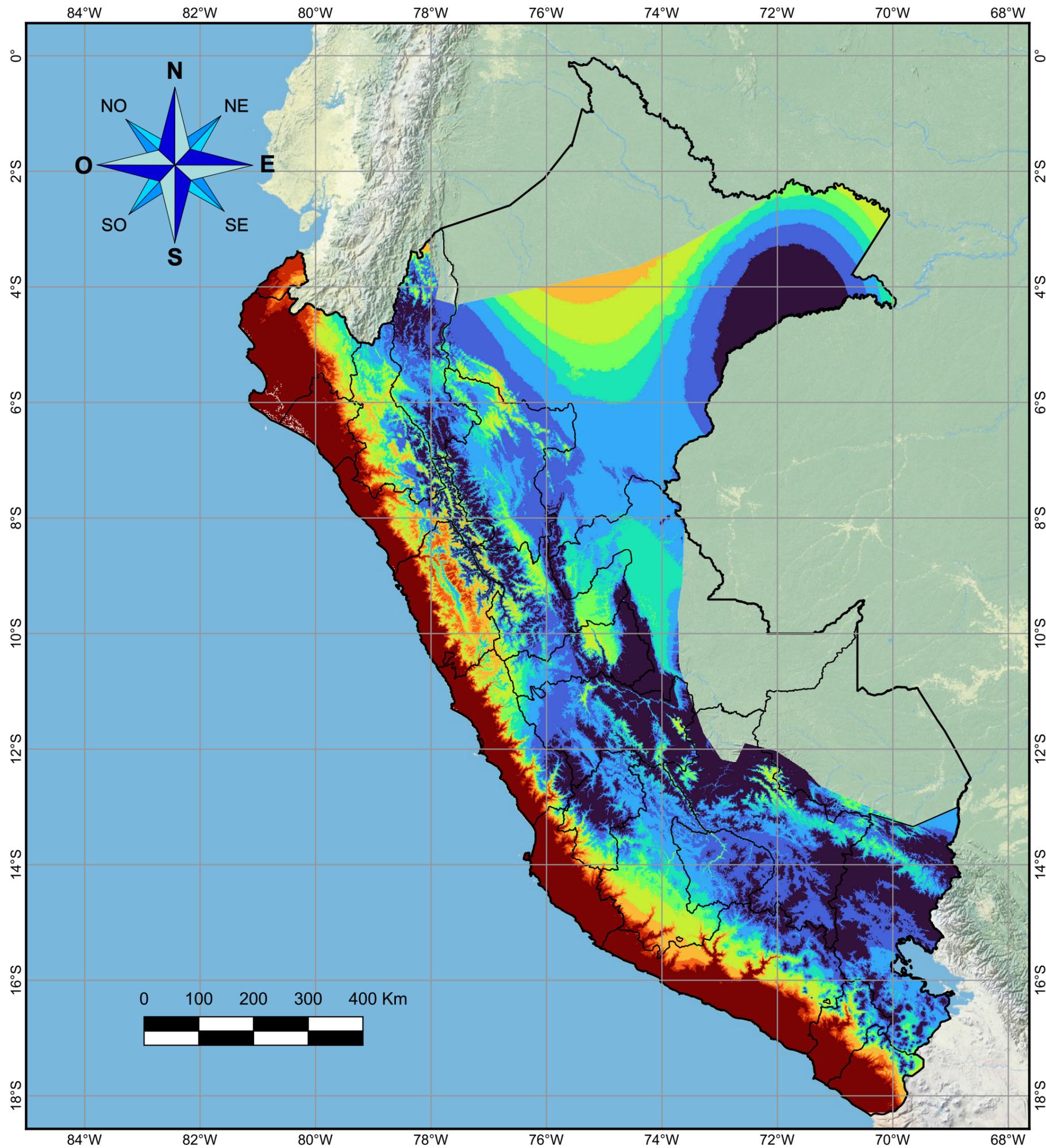
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
 DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
 WGS 84

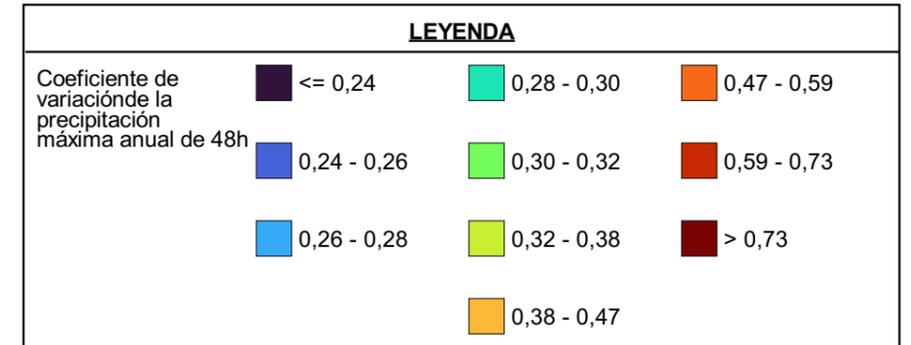
LIMA - PERÚ 2024

Nº MAPA:
 E.4



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

COEFICIENTE DE VARIACIÓN DE LA PRECIPITACIÓN MÁXIMA ANUAL DE 48 HORAS RASTERIZADO

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

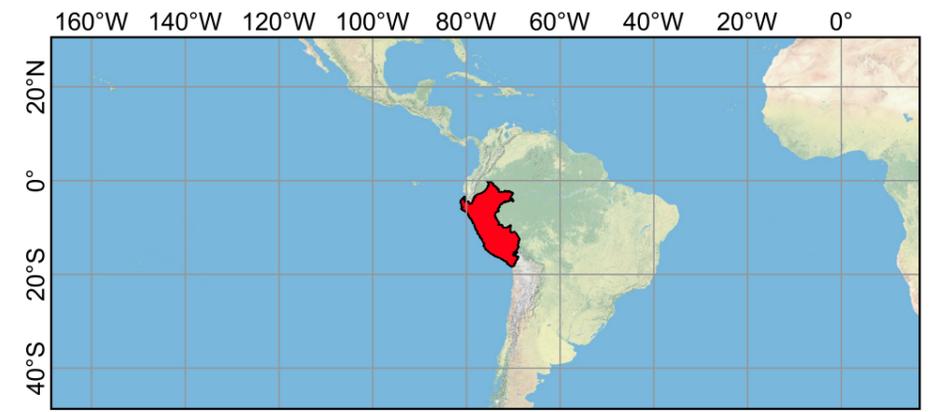
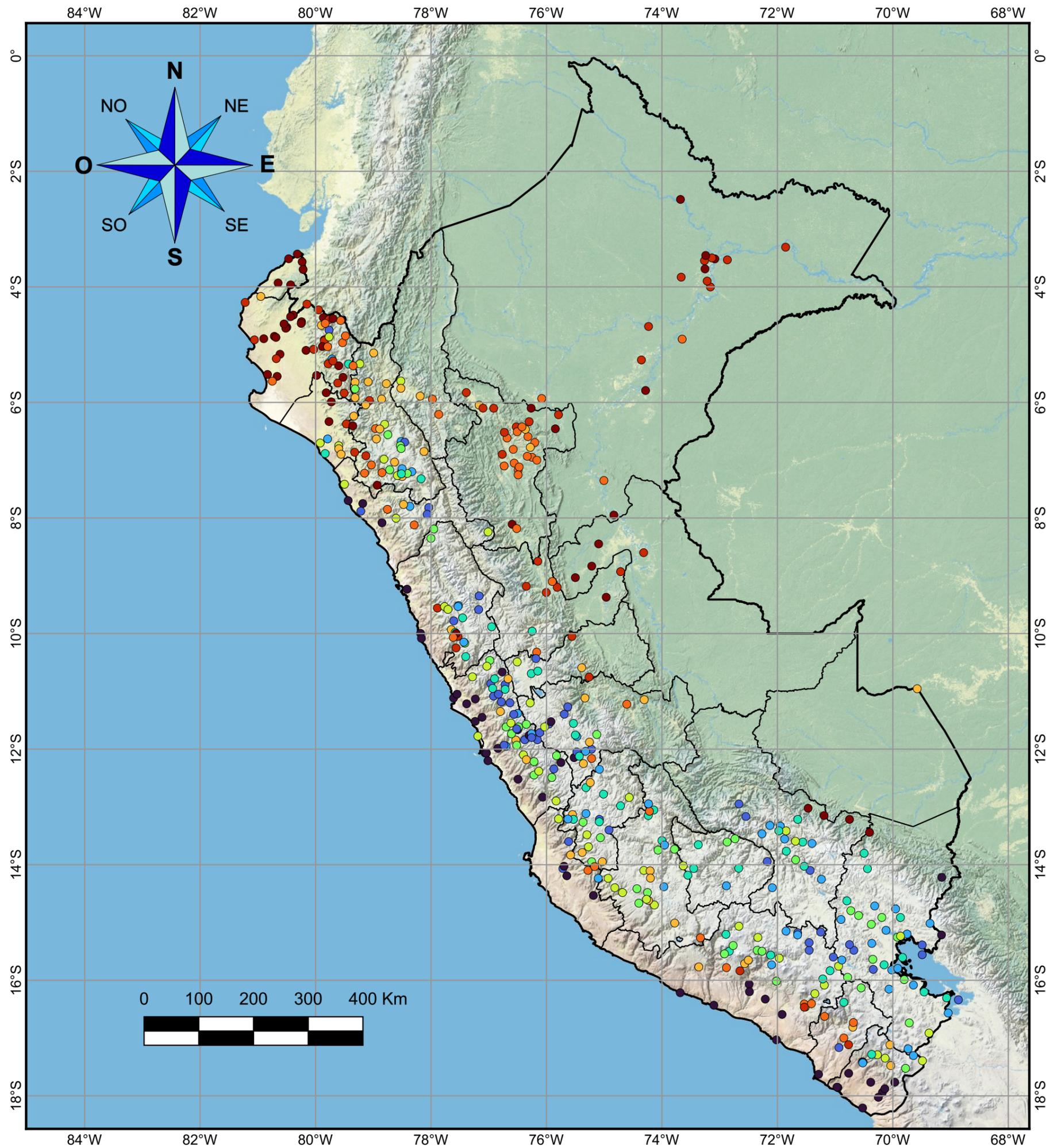
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

E.5



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA		
Precipitación Máxima Probable (PMP) de 48 horas (mm)	● 166 - 205	● 245 - 268
	● 205 - 223	● 268 - 314
	● 223 - 245	● 314 - 389
	● 11 - 166	● 389 - 491
		● 491 - 640
		● 640 - 1645

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MÁXIMA PROBABLE DE 48 HORAS PUNTUAL

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
“El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

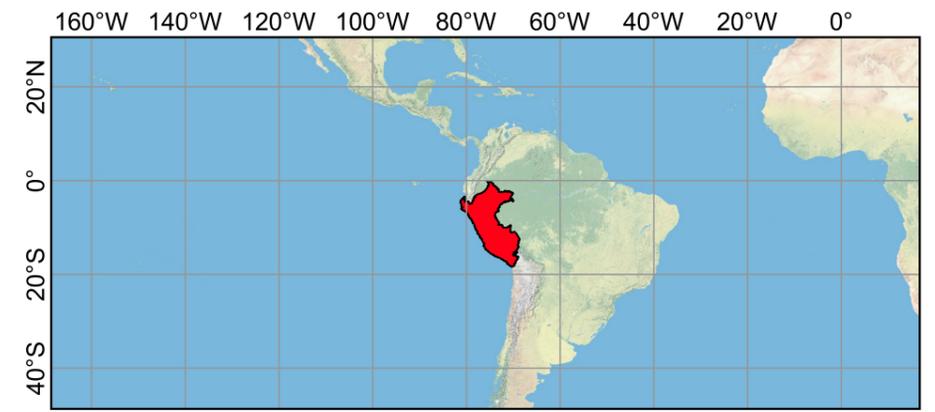
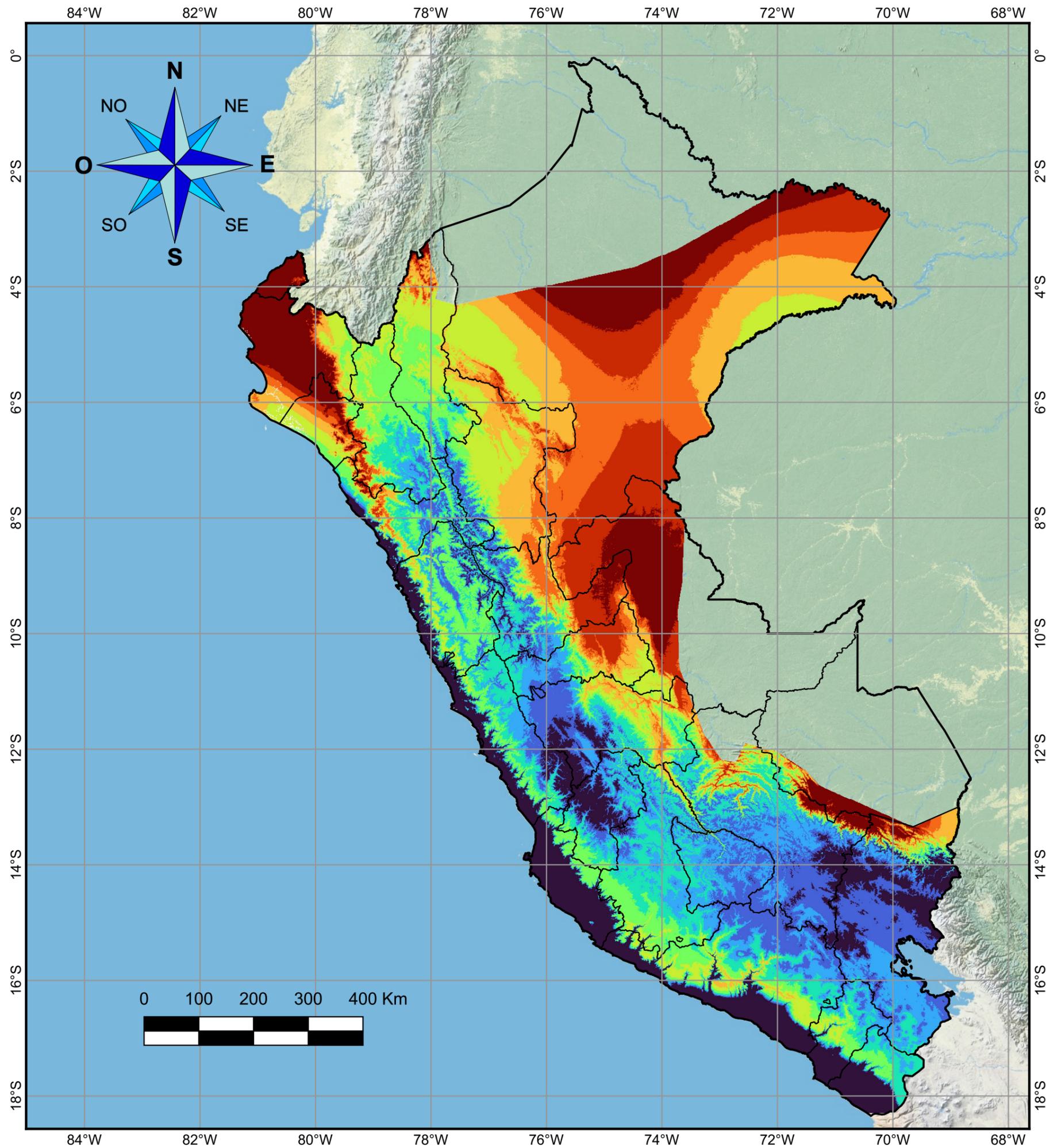
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

Nº MAPA:

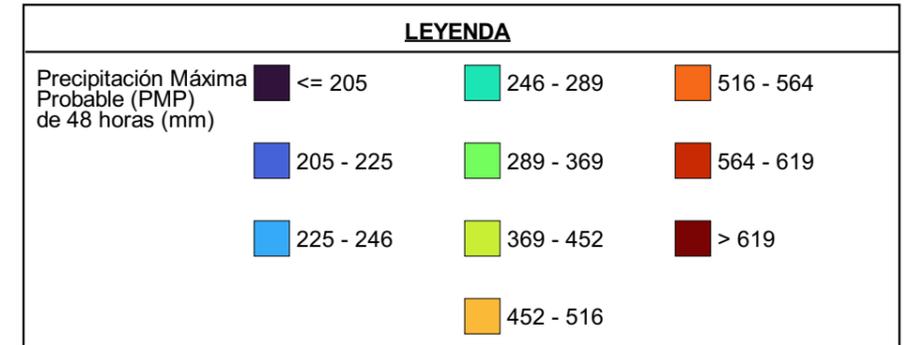
LIMA - PERÚ 2024

E.6



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MÁXIMA PROBABLE DE 48 HORAS RASTERIZADA

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:

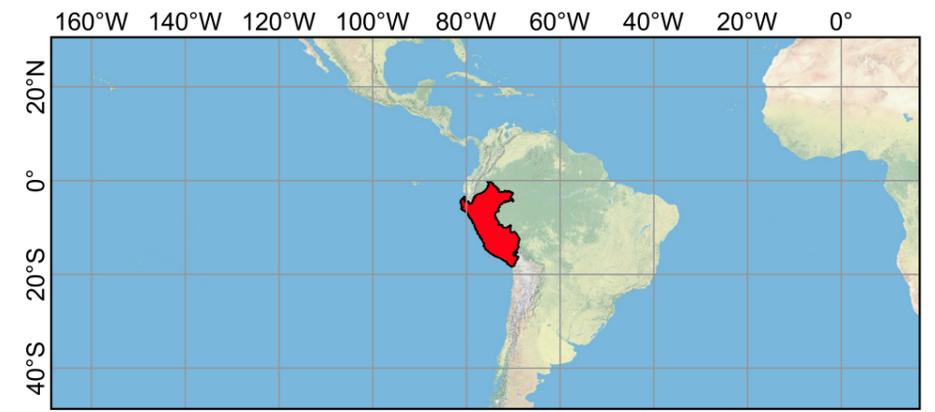
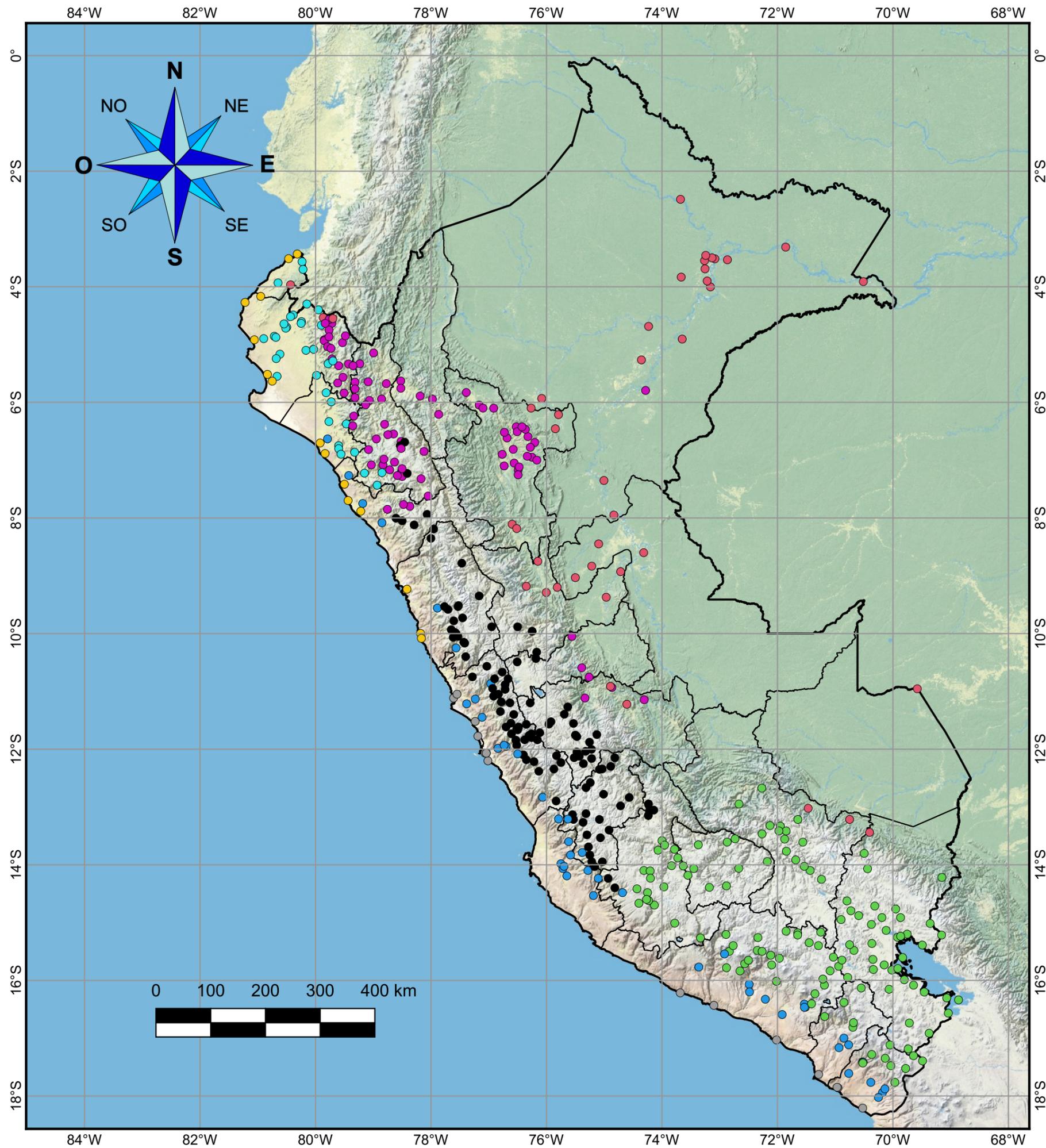
WGS 84

LIMA - PERÚ 2024

Nº MAPA:

E.7

ANEXO F: MAPAS RELACIONADOS A LA DURACIÓN DE 72 HORAS



MAPA DE LOCALIZACIÓN
ESCALA : 1 / 150 000 000

LEYENDA		
Clusters para el caso de 72 horas	● 3	● 6
● 1	● 4	● 7
● 2	● 5	● 8

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA
ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
CLUSTERS DE ESTACIONES PARA EL ANÁLISIS DE LA PMP DE 72 HORAS

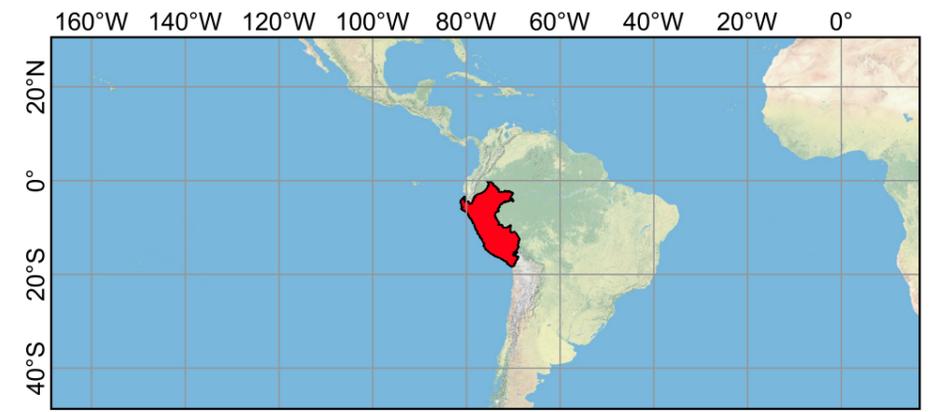
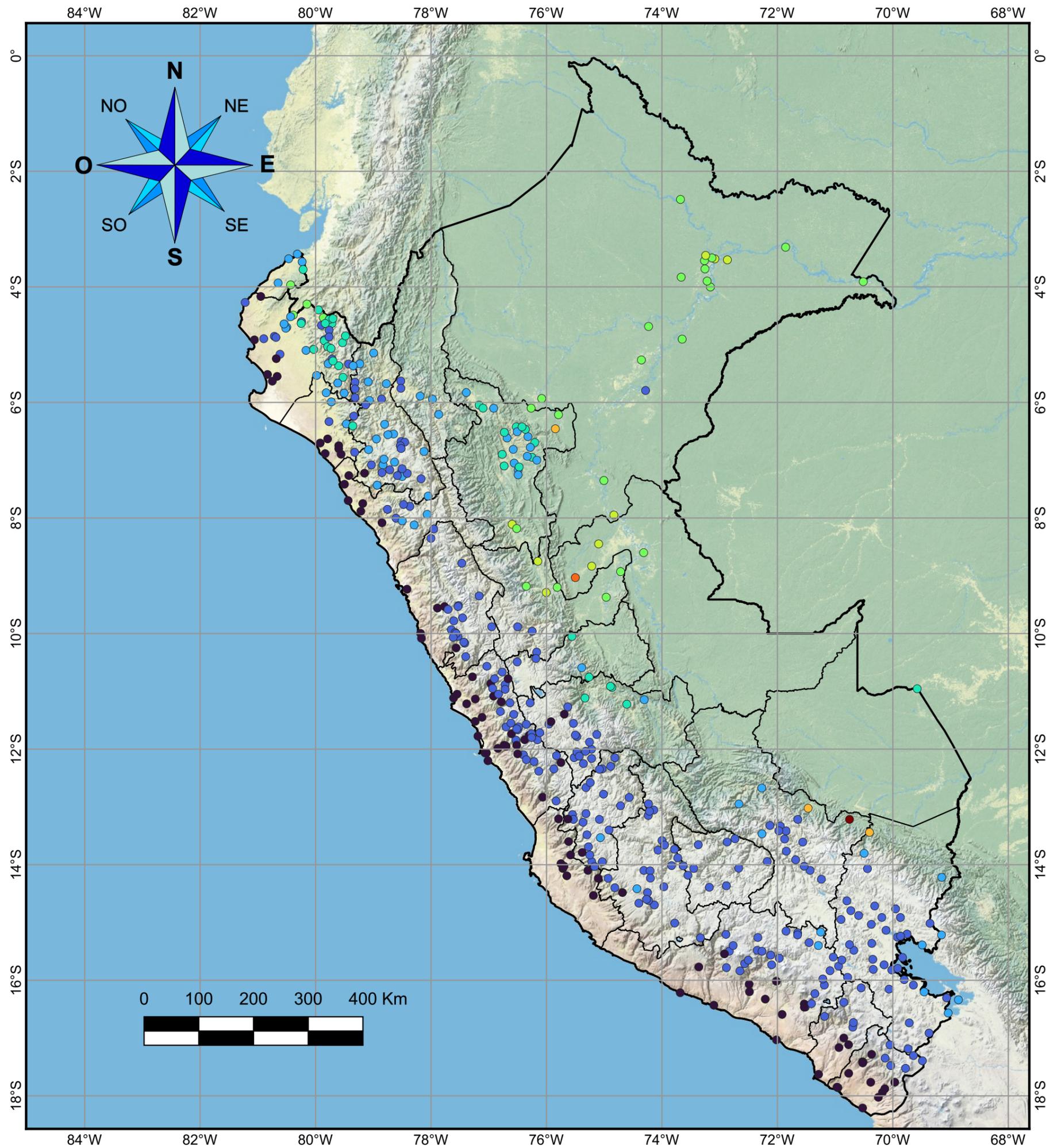
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

LIMA - PERÚ 2024

Nº MAPA:
F.1



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Precipitación media máxima anual de 72 horas (mm)	● 33 - 65	● 130 - 163	● 228 - 260
● 0 - 33	● 65 - 98	● 163 - 195	● 260 - 293
	● 98 - 130	● 195 - 228	● 293 - 326

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MEDIA MÁXIMA ANUAL DE 72 HORAS PUNTUAL

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

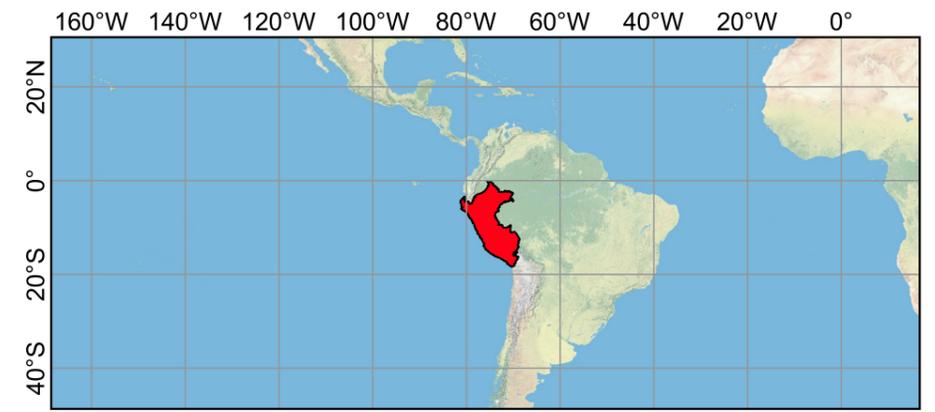
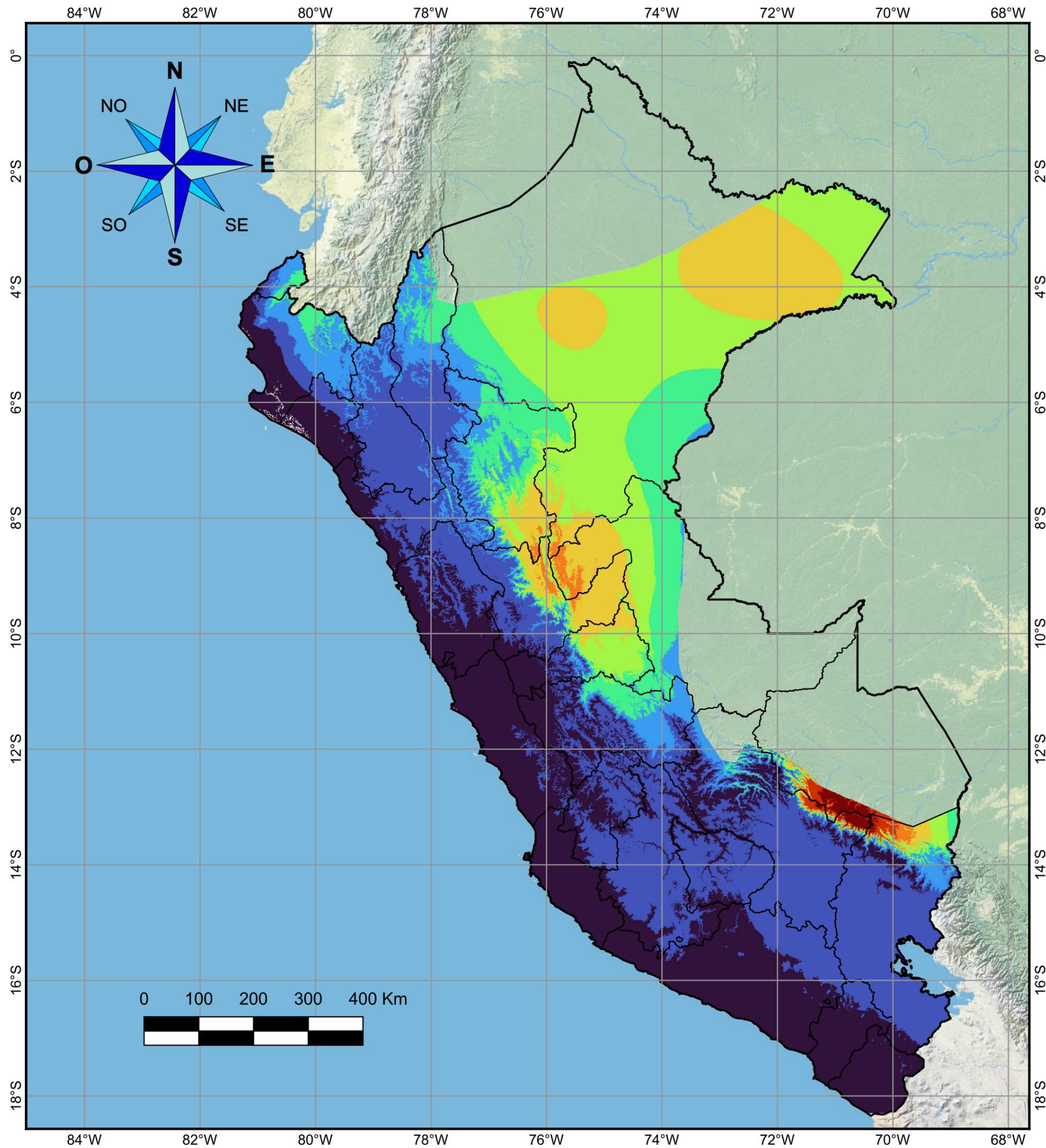
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

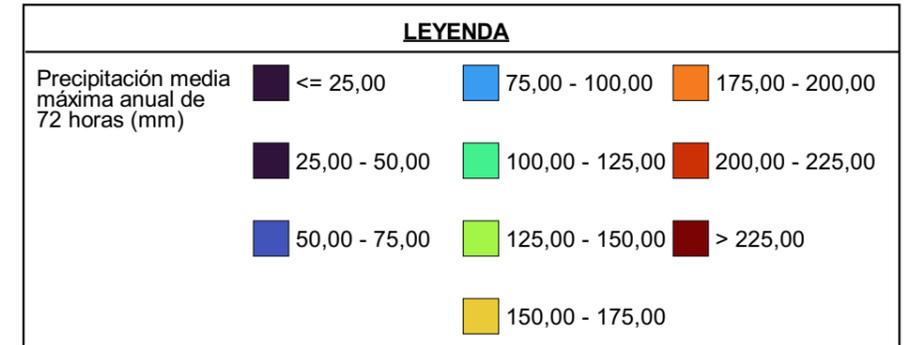
Nº MAPA:

F.2



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MEDIA MÁXIMA ANUAL DE 72 HORAS RASTERIZADA

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

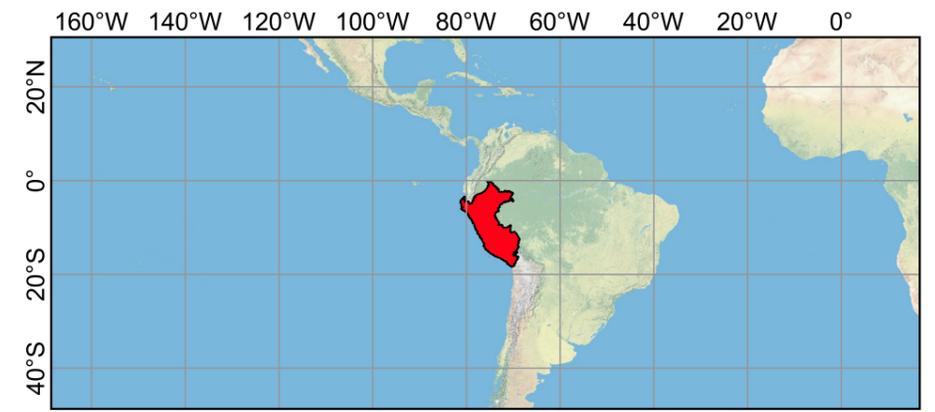
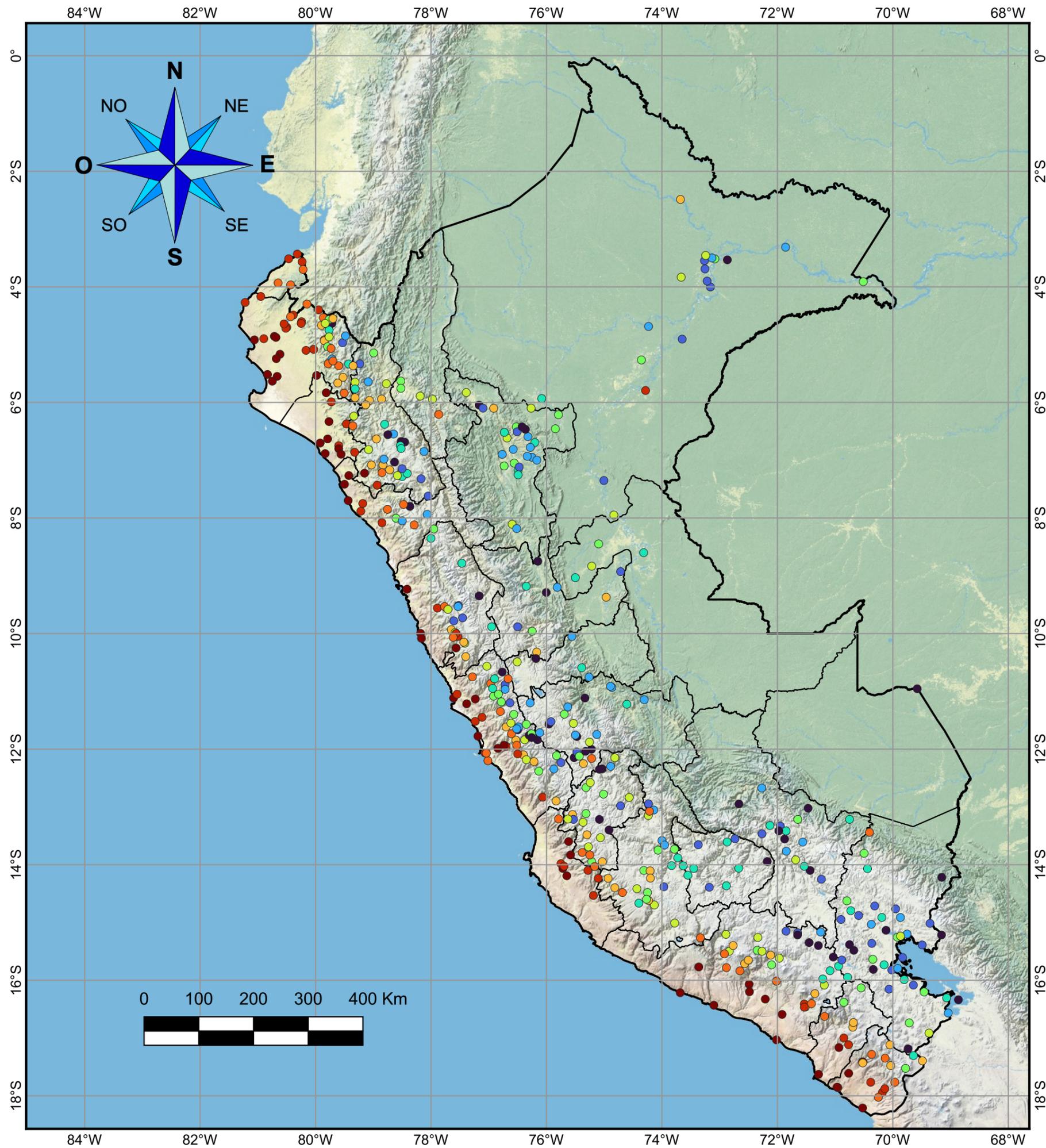
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

F.3



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Coeficiente de variación de la precipitación máxima anual de 72h	● 0.23 - 0.255	● 0.289 - 0.321	● 0.479 - 0.673
	● 0.255 - 0.273	● 0.321 - 0.38	● 0.673 - 0.987
	● 0.18 - 0.23	● 0.38 - 0.479	● 0.987 - 3.071
	● 0.273 - 0.289		

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:
"REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ"

NOMBRE DEL MAPA:
COEFICIENTE DE VARIACIÓN DE LA PRECIPITACIÓN MÁXIMA ANUAL DE 72 HORAS PUNTUAL

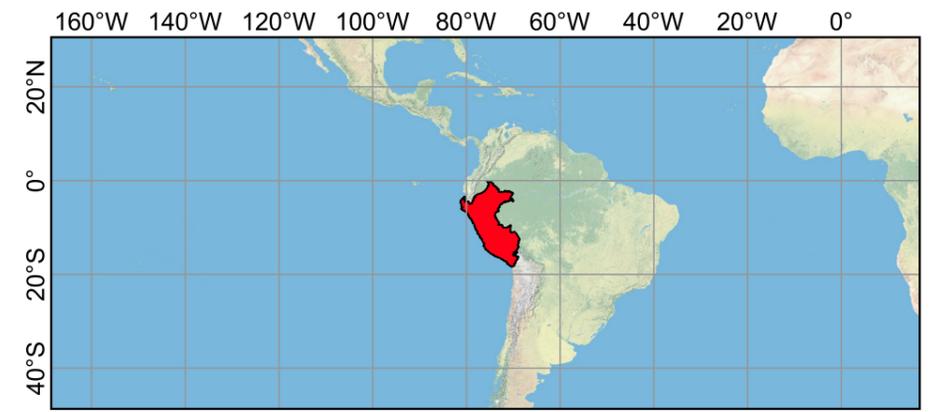
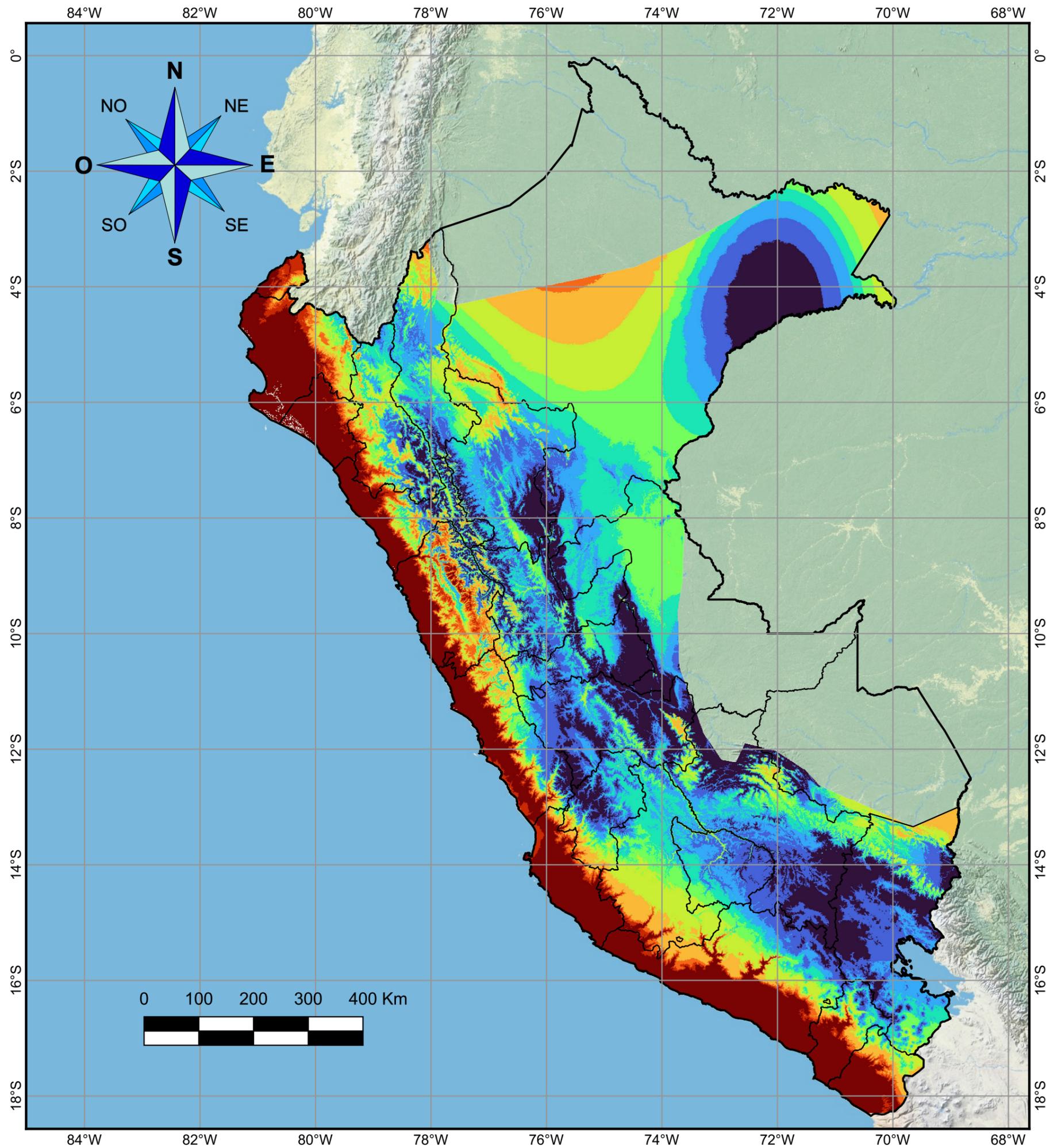
© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
"El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos."
Lucano Alarco, Dante Ramiro
C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:
DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:
WGS 84

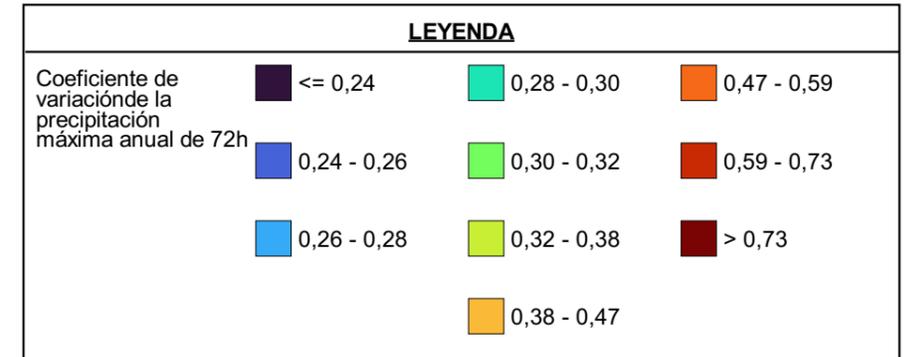
LIMA - PERÚ 2024

Nº MAPA:
F.4



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000



NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

COEFICIENTE DE VARIACIÓN DE LA PRECIPITACIÓN MÁXIMA ANUAL DE 72 HORAS RASTERIZADO

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

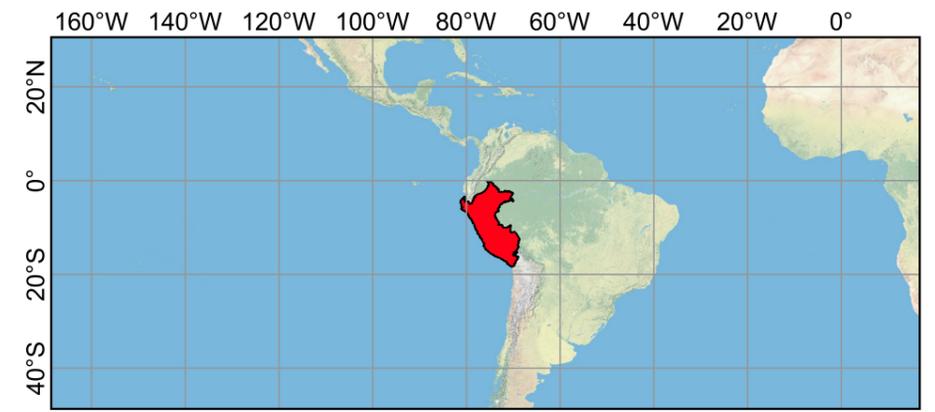
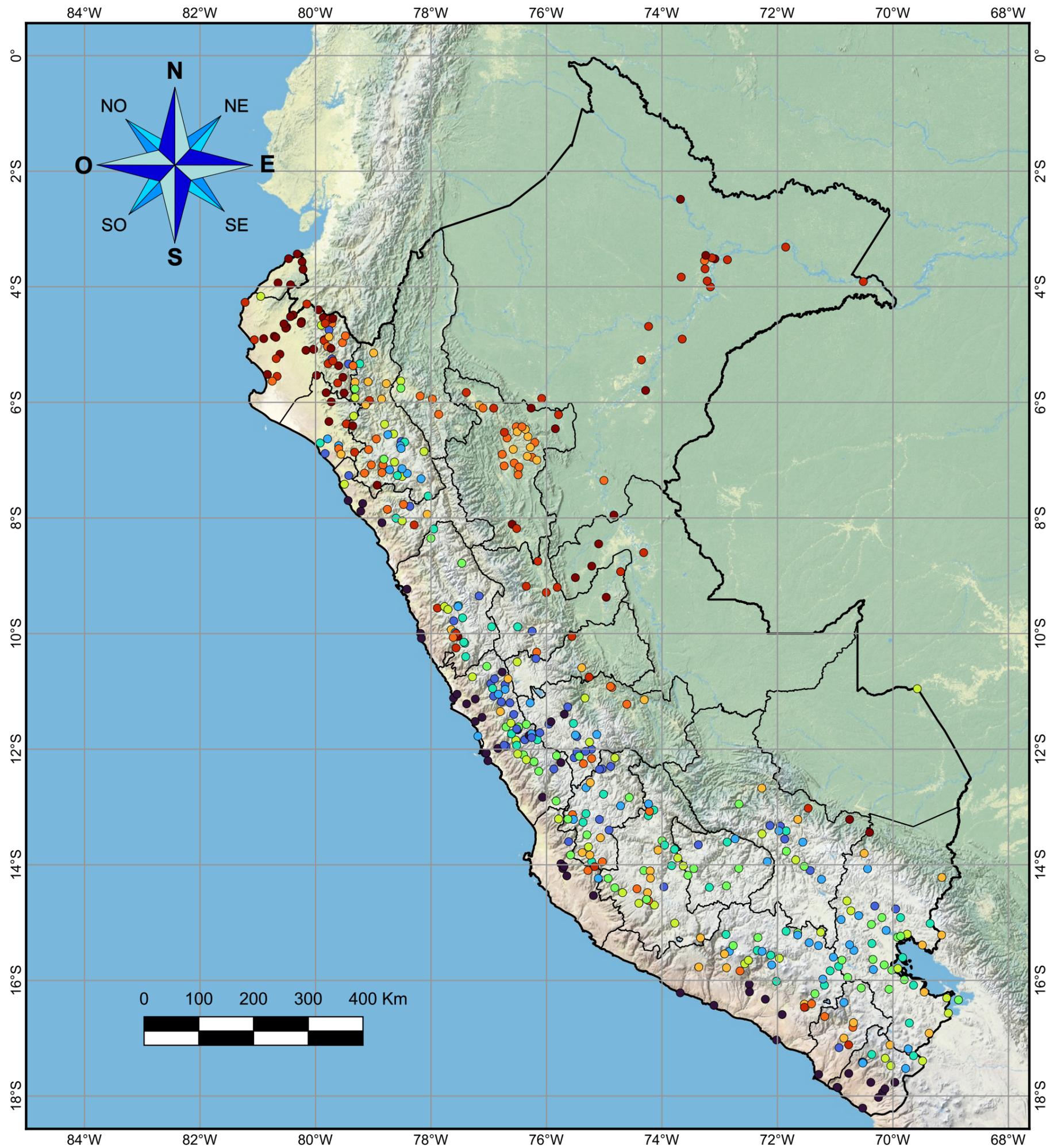
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

F.5



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA		
Precipitación Máxima Probable (PMP) de 72 horas (mm)	● 194 - 256	● 323 - 359
● 11 - 194	● 256 - 294	● 359 - 402
● 294 - 323	● 402 - 468	● 468 - 593
● 593 - 791	● 791 - 1934	

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MÁXIMA PROBABLE DE 72 HORAS PUNTUAL

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

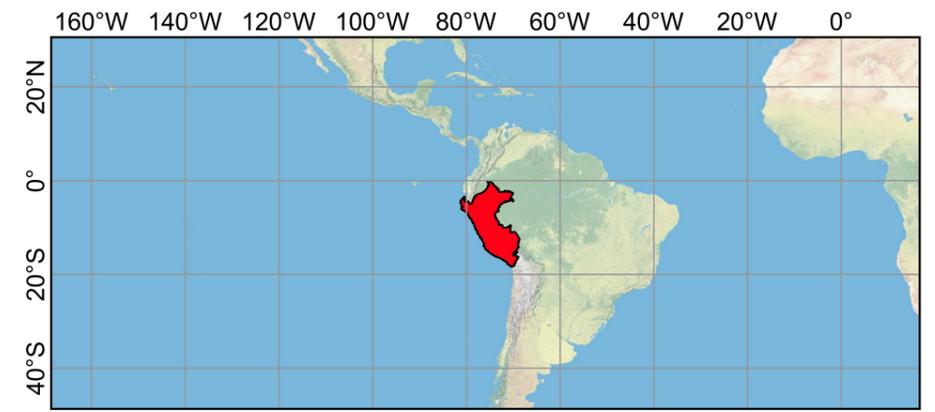
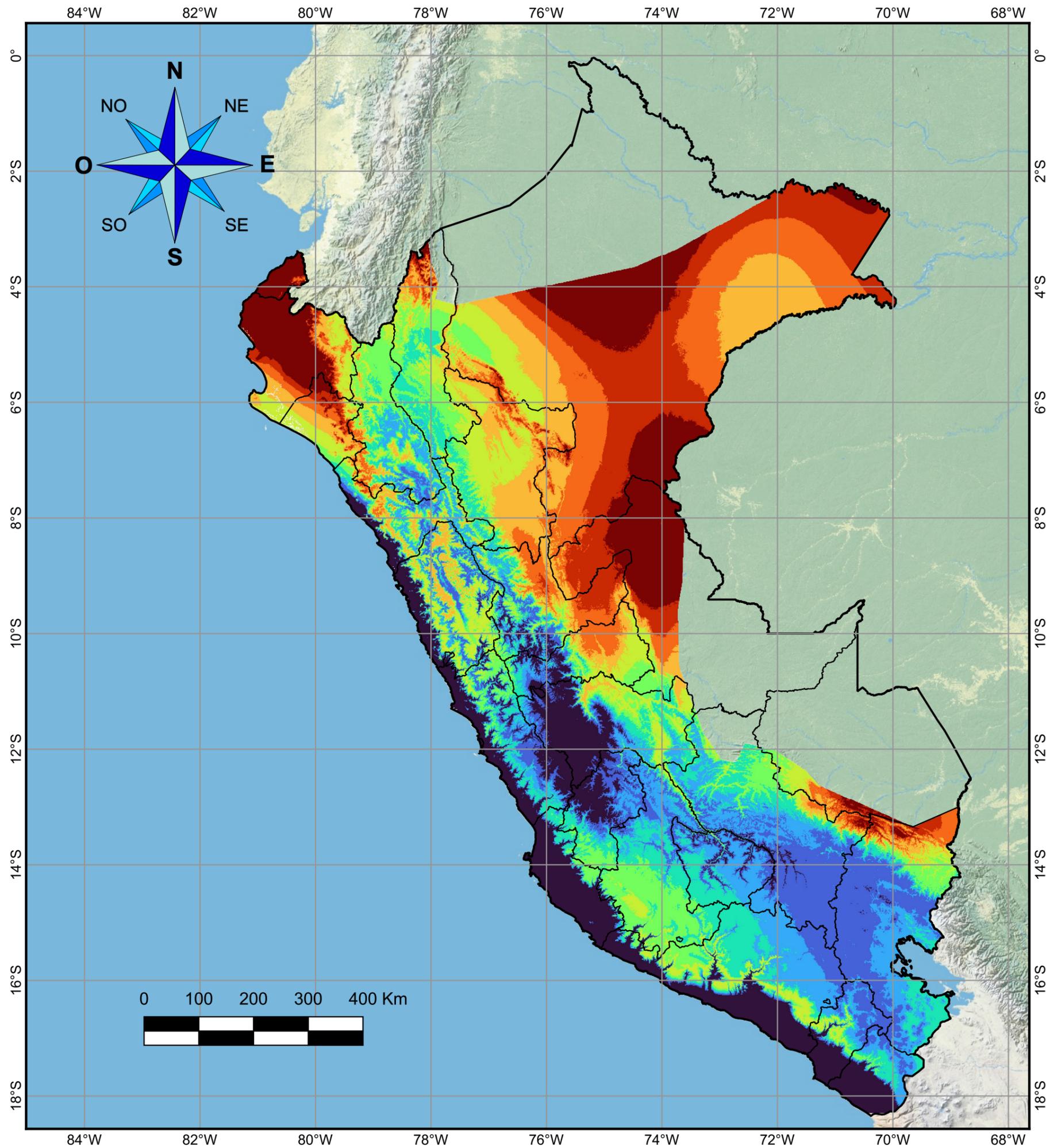
SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

F.6



MAPA DE LOCALIZACIÓN

ESCALA : 1 / 150 000 000

LEYENDA			
Precipitación Máxima Probable (PMP) de 72 horas (mm)	<= 283	346 - 390	645 - 718
	283 - 315	390 - 450	718 - 776
	315 - 346	450 - 547	> 776
		547 - 645	

NOTAS:

1. EL ÁREA DE ESTUDIO ESTÁ UBICADA EN EL DATUM DE REFERENCIA WGS 84.
2. SE UTILIZÓ UNA VISTA SATELITAL DE ESRI DE FONDO.

VISTA

ESCALA: 1 / 7 000 000



UNIVERSIDAD NACIONAL DE INGENIERÍA
Facultad de Ingeniería Civil

TESIS:

“REGIONALIZACIÓN Y ESTIMACIÓN DE LA PRECIPITACIÓN MÁXIMA PROBABLE (PMP) DE 24; 48 Y 72 HORAS EN EL PERÚ”

NOMBRE DEL MAPA:

PRECIPITACIÓN MÁXIMA PROBABLE DE 72 HORAS RASTERIZADA

© 2023, Universidad Nacional de Ingeniería. Todos los derechos reservados.
 “El autor autoriza a la UNI a reproducir la tesis en su totalidad o en parte, con fines estrictamente académicos.”
 Lucano Alarco, Dante Ramiro
 C: +51 956 151 109 | E: dlucanoa@uni.pe

TESISTA:

DANTE RAMIRO LUCANO ALARCO

SISTEMA DE COORDENADAS:

WGS 84

LIMA - PERÚ 2024

Nº MAPA:

F.7