

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

**Clasificación geográfica de la ETo mediante la utilización del modelo de
predicción meteorológico WRF y el método de aprendizaje no supervisado
SOM**

Proyecto de Titulación

Carlos Alberto Flores Cadena

José David Vega, Ph.D.

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito para la obtención del título de Magíster
en Ciencia de Datos

Quito, 12 de mayo de 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

Clasificación geográfica de la ETo mediante la utilización del modelo de predicción meteorológico WRF y el método de aprendizaje no supervisado SOM

Carlos Alberto Flores Cadena

Nombre del Director del Programa:	Felipe Grijalva
Título académico:	Ph.D. en Ingeniería Eléctrica
Director del programa de:	Ciencia de Datos
Nombre del Decano del colegio Académico:	Eduardo Alba
Título académico:	Doctor en Ciencias Matemáticas
Decano del Colegio:	Ciencias e Ingenierías
Nombre del Decano del Colegio de Posgrados:	Dario Niebieskikwiat
Título académico:	Doctor en Física

Quito, mayo 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante: Carlos Alberto Flores Cadena

Código de estudiante: 00339148

C.I.: 1722656996

Lugar y fecha: Quito, 12 de mayo de 2025.

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>.

UNPUBLISHED DOCUMENT

Note: The following graduation project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>.

DEDICATORIA

Dedico este trabajo, en primer lugar, a Dios, fuente y razón de todo lo que soy. A mis padres, mi mamita Yolanda y mi papito Carlos, compañeros incondicionales, pilares de mi vida, cuyo amor y apoyo han sido mi mayor inspiración. A mi familia, en especial a mis abuelos y abuelas, cuyas raíces y enseñanzas han sembrado en mí los valores que me definen. A mis queridas mascotas, Beyli, Miji y Rufi, cuya compañía ha sido una fuente constante de alegría. A mis amigos y compañeros, los que me han acompañado desde mi juventud, y aquellos que aparecieron en los últimos años. Y a mi patria, el Ecuador, esta tierra que, con su cultura y tradición, nutre mi espíritu y me llena el corazón cada día.

AGRADECIMIENTOS

Agradezco profundamente a David Vega, Scott Williams, Luis Marcial, Israel Pineda por haber contribuido a que este estudio se hiciera realidad.

RESUMEN

La evapotranspiración de referencia (ET_o) es un parámetro fundamental para la planificación agrícola y la gestión eficiente del agua. Uno de los modelos más utilizados para estimar la ET_o es el Weather

Research and Forecasting (WRF), ampliamente empleado en estudios climáticos y simulaciones atmosféricas de alta resolución. Al integrar este modelo con mapas auto-organizados (Self-Organizing

Maps, SOM), una técnica de aprendizaje no supervisado en Machine Learning, se obtiene una herramienta poderosa para clasificar patrones espaciales de evapotranspiración. En esta investigación se incorpora, además, el uso de la biblioteca Somoclu, optimizada para el manejo eficiente de grandes volúmenes de datos, mejorando el rendimiento computacional y preservando la topología espacial.

Esta metodología se aplica para la clasificación de patrones de ET_o en las regiones de la Sierra y la Amazonía norte del Ecuador. Los resultados muestran agrupaciones consistentes a lo largo de varios años, contribuyendo al fortalecimiento de estrategias de riego y a una mejor gestión hídrica en zonas agrícolas vulnerables.

Palabras clave: ET_o, mapas auto-organizados SOM, clusterización, Modelo Weather Research and Forecasting (WRF), Somoclu.

ABSTRACT

Reference evapotranspiration (ET_o) is a fundamental parameter for agricultural planning and efficient water management. One of the most widely used models for estimating ET_o is the Weather Research and Forecasting (WRF) model, extensively employed in climate studies and high-resolution atmospheric simulations. By integrating this model with Self-Organizing Maps (SOM), an unsupervised machine learning technique, a powerful tool is obtained for classifying spatial patterns of evapotranspiration. In this research, the Somoclu library is also incorporated, optimized for the efficient handling of large data volumes, improving computational performance and preserving spatial topology. This methodology is applied to classify ET_o patterns in the Sierra and northern Amazon regions of Ecuador. The results show consistent groupings over several years, contributing to the strengthening of irrigation strategies and improved water management in vulnerable agricultural areas.

Key words: ET_o, Self-Organizing Maps, SOM, Clustering, Weather Research and Forecasting Model (WRF), Somoclu.

TABLA DE CONTENIDO

I.	Introducción	11
A.	Estado del Arte	11
II.	Marco Teórico	12
A.	Evapotranspiración de Referencia (ET _o)	12
B.	Weather Research and Forecasting (WRF)	13
C.	Self-Organizing Maps (SOM)	13
D.	Somoclu	14
III.	Metodología	14
A.	Área de estudio	14
B.	Datos meteorológicos	14
C.	Preprocesamiento de Datos	14
D.	Aplicación de SOM a través de Somoclu	14
	D1. Configuración del Modelo	14
	D2. Entrenamiento y Clasificación	15
	D3. Evaluación del Modelo	15
	D4. Plataforma Computacional	15
IV.	Resultados y Discusión	15
A.	Clasificación espacial de ET _o mediante SOM	15
B.	Evaluación de la cohesión de clústeres	16
V.	Conclusiones	17
	References	17

ÍNDICE DE FIGURAS

1.	Self-Organizing Map. La metodología SOM busca realizar una reducción de la dimensionalidad de los datos de entrada proyectándolos en un espacio de dos dimensiones, manteniendo sus características topográficas.	13
2.	Área de estudio: región Sierra y Amazonía norte de Ecuador.	14
3.	Clasificación espacial de la ETo mediante SOM: 1 de enero de 2019.	15
4.	Clasificación espacial de la ETo para los años 2017 y 2018.	15
5.	Error de cuantización (QE) de enero a diciembre en los años 2017, 2018 y 2019.	16

Clasificación geográfica de la ETo mediante la utilización del modelo de predicción meteorológico WRF y el método de aprendizaje no supervisado SOM

Carlos Flores and José David Vega

Abstract—La estimación precisa de la evapotranspiración de referencia (*evapotranspiration*, por sus siglas en inglés ETo) es crucial para la gestión eficiente del riego y la administración sostenible de los recursos hídricos. Uno de los modelos más reconocidos para su evaluación es el *Weather Research and Forecasting* (WRF), ampliamente utilizado tanto en investigaciones climáticas como en la predicción meteorológica numérica. Al integrarse con la red neuronal artificial *Self-Organizing Map* (SOM), una técnica de aprendizaje no supervisado dentro del campo de *Machine Learning*, se constituye una herramienta robusta para la clasificación de patrones espaciales de evapotranspiración. Para optimizar el procesamiento de grandes volúmenes de datos meteorológicos simulados, en este trabajo se propone el empleo de la biblioteca Somoclu, una implementación eficiente de SOM que permite acelerar el entrenamiento y preservar la topología de los datos a alta resolución. Esta combinación metodológica permite la identificación y categorización de distintos "climas" de ETo en un mapeo geográfico detallado, lo cual resulta particularmente valioso para optimizar prácticas agrícolas en zonas específicas. En este contexto, se propone la aplicación de esta metodología para la clasificación de patrones climáticos asociados a la ETo en las regiones del Amazonas y los Andes ecuatorianos. Los resultados obtenidos evidencian agrupaciones estables y dinámicas de ETo, aportando información estratégica que puede contribuir significativamente a mejorar las prácticas de riego y la planificación de la gestión hídrica.

Index Terms—evapotranspiración, Self-Organizing Maps, WRF, Machine Learning, Aprendizaje No Supervisado

I. INTRODUCCIÓN

LA gestión del agua en la agricultura es uno de los desafíos más antiguos y persistentes en el mundo [1]. La disponibilidad limitada de agua en relación con la demanda por evapotranspiración ha afectado la expansión de áreas agrícolas [2]. Asimismo, los efectos del cambio climático agravan esta problemática, incrementando los

riesgos de aridización de tierras cultivables. La evapotranspiración de referencia (*evapotranspiration*, por sus siglas en inglés ETo) representa la demanda atmosférica de agua por evaporación y transpiración desde una superficie de referencia. Aunque existen metodologías consolidadas para su cálculo, como formulaciones matemáticas [3], la necesidad de datos meteorológicos continuos limita su aplicación en regiones en desarrollo [4].

Específicamente, el cálculo de la ETo depende de la disponibilidad de datos meteorológicos de alta calidad, como temperatura, humedad relativa, radiación solar y velocidad del viento. Sin embargo, en muchas regiones del mundo, especialmente en países en desarrollo, el acceso a mediciones continuas y confiables de estas variables es limitado [5]. Para superar esta carencia, se han desarrollado enfoques basados en modelos de simulación atmosférica, como el *Weather Research and Forecasting* (WRF) [6], los cuales permiten generar datos meteorológicos sintéticos de alta resolución espacial y temporal. Paralelamente, el crecimiento de técnicas de inteligencia artificial, en particular el aprendizaje automático (*Machine Learning*) no supervisado, ha abierto nuevas posibilidades para analizar y clasificar patrones espaciales en conjuntos de datos meteorológicos complejos, facilitando la comprensión de fenómenos como la dinámica de la evapotranspiración en distintas regiones.

Investigación Reproducible: El código de simulación está disponible en: https://github.com/carlosflores597/SOM_WRF

A. Estado del Arte

La evapotranspiración de referencia (ETo) es un parámetro esencial para la gestión eficiente del riego y la planificación agrícola. Tradicionalmente, se estima utilizando la ecuación de Penman-Monteith, que requiere datos meteorológicos como temperatura, humedad relativa, radiación solar y velocidad del viento. Sin embargo, en regiones con escasa infraestructura meteorológica, la obtención de estos datos puede ser limitada, lo que dificulta la aplicación precisa de este método [7]. Para superar estas limitaciones, se han explorado métodos alternativos que utilizan modelos de simulación atmosférica y técnicas de inteligencia artificial.

Carlos Alberto Flores Cadena estudiante de Maestría en la Universidad San Francisco de Quito, Quito, Ecuador.

David Vega profesor en la Universidad San Francisco de Quito, Quito, Ecuador.

Manuscrito recibido en Mayo 10, 2024. Revisado en Mayo 11, 2025. Modificado Mayo 11, 2025.

Por ejemplo, el modelo WRF ha sido empleado para generar datos meteorológicos simulados que alimentan algoritmos de aprendizaje automático para estimar la ETo [8].

Por otro lado, los mapas auto-organizados (*Self-Organizing Map*, por sus siglas en inglés SOM) son una técnica de redes neuronales no supervisadas que permiten la proyección de datos de alta dimensión en un espacio de menor dimensión, preservando la topología de los datos originales. Esta característica los hace adecuados para la clasificación y visualización de patrones complejos en datos meteorológicos. En esencia, la idea central propuesta originalmente en [9], consiste en agrupar regiones geográficas según variables climáticas cuantificables. Esto ha servido de inspiración para estas nuevas aproximaciones metodológicas en la clasificación de fenómenos hidrometeorológicos. Siguiendo este principio, investigaciones recientes han adaptado modelos de predicción atmosférica como WRF y técnicas de inteligencia artificial no supervisada como SOM para construir clasificaciones espaciales detalladas de la evapotranspiración de referencia. A diferencia de la clasificación climática tradicional basada únicamente en temperatura y precipitación, el enfoque contemporáneo incorpora un conjunto más amplio de variables meteorológicas simuladas, considerando además su variabilidad horaria y espacial, lo cual permite capturar de manera más precisa la dinámica de los procesos atmosféricos que afectan el balance hídrico. Por ejemplo, el estudio realizado en [10] integró datos generados por el modelo WRF a 3 km de resolución con SOM, aplicando dos enfoques de aprendizaje: superficial, utilizando un único corte horario diario, y profundo, considerando las 24 horas del ciclo diario. La investigación exploró la clasificación de ETo a partir de variables WRF brutas y derivadas, en versiones normalizadas y no normalizadas, analizando distintos números de clases para evaluar la estabilidad espacial y temporal de los patrones detectados. Los resultados evidenciaron que las zonas montañosas del Ecuador presentaron una clasificación más estable, mientras que la Amazonía y la Costa exhibieron mayor migración de clases, siendo el enfoque de aprendizaje profundo el que proporcionó agrupamientos más coherentes a lo largo del tiempo. Continuando esta línea de investigación, los autores en [11] desarrollaron una metodología para clasificar el "clima" de la ETo a escala mesoespacial, utilizando datos generados por el modelo WRF y técnicas de aprendizaje no supervisado basadas en SOM. A partir del análisis de histogramas de ocurrencia de clases de ETo "weather" en cada píxel geográfico, construyendo agrupamientos espaciales que representan tendencias climáticas de largo plazo en las regiones andina y amazónica del Ecuador.

En base a los avances metodológicos propuestos por los estudios antes mencionados, así como en la idea fundacional de [9] de agrupar regiones climáticas mediante variables cuantificables, la propuesta desarrollada en este trabajo no solo representa una evolución del paradigma clásico, sino que también introduce un enfoque innovador basado en el uso de modelos de predicción atmosférica y aprendizaje automático no supervisado para clasificar dinámicamente

patrones de ETo. Este enfoque amplía significativamente el espectro de variables consideradas, capta la variabilidad espacio-temporal de los procesos hidrometeorológicos, y proporciona una herramienta robusta para la gestión hídrica y la planificación agrícola en territorios donde los datos observacionales son escasos o inexistentes. Además, como contribución específica de este estudio, se incorpora la utilización de la biblioteca Somoclu para el entrenamiento eficiente de mapas auto-organizados, lo que permite abordar grandes volúmenes de datos meteorológicos simulados con alta resolución espacial, mejorando el desempeño computacional respecto a implementaciones anteriores y sentando las bases para futuras aplicaciones en sistemas de monitoreo y toma de decisiones estratégicas en regiones vulnerables.

II. MARCO TEÓRICO

En esta sección se presentan los conceptos fundamentales que sustentan esta investigación, incluyendo la definición de la ETo, el modelo WRF para la generación de datos meteorológicos simulados, y los mapas auto-organizados (SOM) como técnica de clasificación espacial. También se distingue entre los enfoques de ETo 'weather' y ETo 'climate', fundamentales para interpretar la variabilidad hidrometeorológica.

A. Evapotranspiración de Referencia (ETo)

La evapotranspiración de referencia (ETo) es un parámetro importante del ciclo del agua, que caracteriza el transporte de humedad desde la superficie terrestre a la atmósfera [1]. Este proceso consiste en el proceso de evaporación simultánea del agua del suelo y la transpiración de las plantas [2]. Además, este proceso juega un rol importante en la regulación del clima local y regional [12], en la gestión de recursos hídricos locales y globales [13], gestión del riego [14], entre otros. A escala global, la ETo representa más de la mitad de la precipitación total global, lo que significa que la mayor parte del agua transportada desde la atmósfera a la superficie no será almacenada [12]. En particular, para calcular la ETo se utiliza la ecuación de Penman-Monteith, la cuál utiliza variables como temperatura, humedad, velocidad del viento y radiación solar, y esta dada por [3]

$$ETo = \frac{0.408\Delta(R_n - G) + \gamma \frac{37}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}, \quad (1)$$

dónde, R_n es la radiación neta, G el flujo de calor en el suelo, T la temperatura del aire, Δ la pendiente de la curva de presión de vapor. Asimismo, γ representa la constante psicrométrica, e_s la presión de vapor de saturación, e_a la presión de vapor real y u_2 la velocidad del viento.

Aunque la expresión de Penman-Monteith dada en (1) es un estándar para estimar ETo, su aplicación efectiva requiere datos meteorológicos continuos que, en muchos países en desarrollo, son limitados o inaccesibles. Esta carencia ha impulsado el uso de modelos como el WRF para generar

datos simulados, y de técnicas de aprendizaje automático, como SOM, para clasificar patrones de ETo sin depender de mediciones locales.

B. Weather Research and Forecasting (WRF)

Varios modelos se han utilizado para medir la ETo siendo el modelo WRF uno de los más populares debido a la accesibilidad universal de sus datos. Específicamente, el enfoque WRF es un modelo atmosférico diseñado, tanto para la investigación como para la predicción numérica del clima [15]. Este es un modelo que integra pronóstico numérico del tiempo, simulación atmosférica y asimilación de datos, además de adoptar un esquema de proceso físico mejorado con múltiples niveles de anidamiento y la capacidad de ubicarse fácilmente en diferentes ubicaciones geográficas, este modelo se usa ampliamente en simulación y predicción del tiempo a mesoescala [12]. Sin embargo, la dificultad que posee este modelo es su alto costo computacional. Cabe mencionar que el modelo WRF consiste en una interpolación temporal de otro modelo climatológico, conocido como el modelo GFS (Global Forecast System) el cual genera datos meteorológicos cada 6 horas. WRF fue desarrollado en entornos FORTRAN, C++, C y Shell, con diferentes versiones adaptadas a fenómenos atmosféricos específicos. Antes de su ejecución, el sistema interactúa con los siguientes componentes clave [15]

- Datos de entrada geográficos y topográficos
- Datos de entrada meteorológicos y atmosféricos
- Sistema de Preprocesamiento del Modelo WRF.

Una vez que el WRF termina su ejecución, la salida es un archivo *Network Common Data Form* (NetCDF), el cual requiere un posprocesamiento para su análisis y visualización:

- **Visualización:** Se emplean herramientas especializadas, como Ncview, para explorar y representar gráficamente los archivos NetCDF producidos por el modelo.
- **Python:** Facilita el posprocesamiento de los datos mediante diversas técnicas, como análisis estadístico, agrupamiento no supervisado, generación de predicciones y creación de visualizaciones gráficas, entre otras funcionalidades.

C. Self-Organizing Maps (SOM)

Los Mapas Auto-organizados SOM, introducidos por Kohonen en [16], son ampliamente utilizados para tareas de agrupamiento y reducción de dimensionalidad. En el contexto de la ETo, la metodología SOM permite identificar patrones espaciales y temporales complejos, clasificando regiones de estudio según comportamientos similares de ETo derivados de variables meteorológicas simuladas. Esta técnica es especialmente valorada frente a otros algoritmos de agrupamiento, como K-Means, debido a su capacidad para conservar la topología espacial de los datos de entrada y su proceso de aprendizaje adaptativo a

través de tasas de aprendizaje decrecientes. En particular, SOM fue inicialmente desarrollada para realizar operaciones de agrupación, clasificación, predicción y extracción de patrones, especialmente aplicada a grandes conjuntos de datos [10]. Posteriormente, estudios en [10] y [11] han aplicado este enfoque para la gestión de recursos hídricos, mientras que trabajos como en [17] extendieron su uso a la categorización de la calidad del agua, suelo y sedimentos. En esencia, SOM proyecta datos de alta dimensión en una estructura reticular bidimensional, preservando las características topológicas del espacio de entrada a través de una fase de entrenamiento competitivo [11]. La Figura 1 ilustra este proceso, donde la reducción dimensional facilita la interpretación visual de patrones ocultos en los datos.

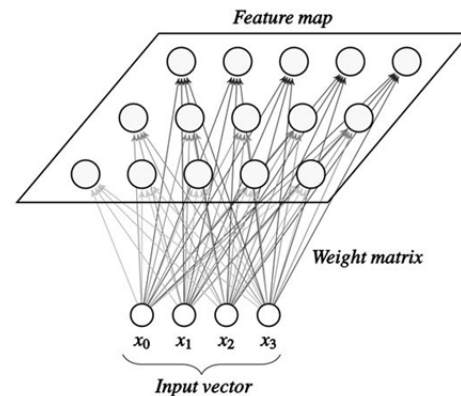


Figure 1. Self-Organizing Map. La metodología SOM busca realizar una reducción de la dimensionalidad de los datos de entrada proyectándolos en un espacio de dos dimensiones, manteniendo sus características topográficas.

El funcionamiento de SOM implica la disposición de neuronas en una cuadrícula bidimensional, donde cada neurona está asociada a un vector de pesos que representa las variables de los datos de entrada. Durante el entrenamiento, las neuronas compiten para ajustar sus pesos, adaptándose progresivamente a los patrones de los datos presentados. Este proceso, que se repite iterativamente, requiere de un conjunto de parámetros de entrada, entre los que destacan [11]:

- **Columnas:** Variables o características del conjunto de datos.
- **Filas:** Observaciones o registros del conjunto de datos.
- **m:** Número de nodos en el eje horizontal de la malla neuronal.
- **n:** Número de nodos en el eje vertical de la malla neuronal.
- **lr:** Valor inicial de la tasa de aprendizaje.
- **dim:** Número de características por observación (dimensionalidad de entrada).
- **max_iter:** Número máximo de iteraciones de entrenamiento.
- **sigma:** Parámetro inicial del decaimiento espacial (radio de vecindad).

Tras el entrenamiento, el modelo genera una salida en forma de etiquetas (labels), asignando a cada registro del conjunto de datos una neurona ganadora en función de su distancia al vector de pesos de cada nodo. Estas etiquetas permiten identificar agrupamientos naturales en el conjunto de datos. En un SOM estructurado con una malla de $m \times n$ nodos, el total de neuronas es $k = m \times n$. Cada neurona está conectada a las dim variables de entrada mediante un vector de pesos, resultando en una matriz de pesos de tamaño (dim, k) , lo que equivale a “dim” filas y “k” columnas

D. Somoclu

Somoclu es una implementación optimizada del algoritmo SOM, diseñada específicamente para manejar grandes volúmenes de datos y aprovechar la computación paralela. A diferencia de otras bibliotecas como MiniSom o Scikit-learn, Somoclu permite entrenar mapas autoorganizados de gran tamaño de manera más rápida y eficiente, utilizando tanto CPUs multicore como GPUs. Esta herramienta ha demostrado ser particularmente útil en escenarios donde la cantidad de observaciones y variables impone altos requerimientos computacionales, como ocurre en el análisis espacial de datos meteorológicos simulados [18]. En este trabajo, Somoclu se utiliza para la clasificación no supervisada de patrones de ETo a partir de datos generados por el modelo WRF, permitiendo una reducción efectiva de la dimensionalidad sin perder las características topológicas de los datos originales.

III. METODOLOGÍA

A. Área de estudio

El área de estudio de este trabajo abarca la región Sierra y la Amazonía norte del Ecuador, seleccionada por su ubicación estratégica y la disponibilidad de datos meteorológicos relevantes. Esta zona también fue elegida considerando la existencia de estudios previos [11], [10] que han explorado dinámicas hidrometeorológicas en la región. Específicamente, la Figura 2 corresponde a una extensión definida por una malla de 171 por 171 píxeles (recuadro de color rojo resaltado), cada uno de aproximadamente 3.3 km $\times$ 3.3 km. Las coordenadas iniciales del dominio son 0°36'00.0"N, 79°00'00.0"W, mientras que las coordenadas finales alcanzan 4°28'00.0"S, 73°55'00.0"W, los que aseguran una representación adecuada de las dimensiones espaciales y de la cobertura geográfica sobre la región de interés.

B. Datos meteorológicos

La fuente de datos meteorológicos utilizada fue el modelo Global Forecast System (GFS), operado por la Administración Nacional Oceánica y Atmosférica (NOAA) a través de sus Centros Nacionales de Predicción Ambiental (NCEP). El GFS genera datos meteorológicos globales cuatro veces al día (00:00, 06:00, 12:00 y 18:00 UTC), proporcionando variables como temperatura del aire a 2 metros (T2), presión superficial (PSFC), humedad relativa

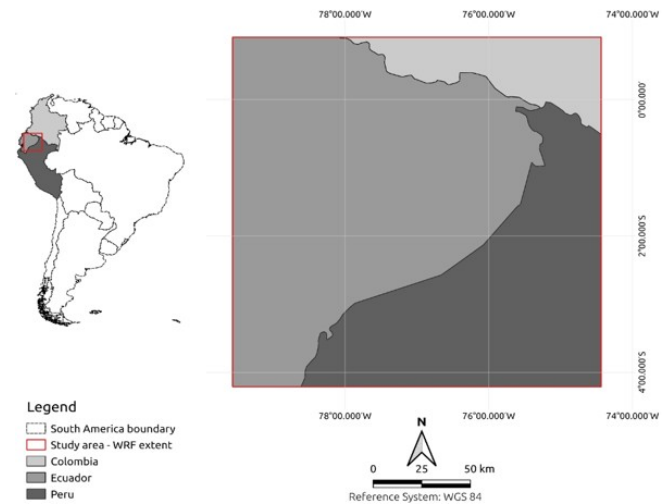


Figure 2. Área de estudio: región Sierra y Amazonía norte de Ecuador.

(Q2), radiación solar de onda corta (SWDOWN) y de onda larga (GLW), temperatura de la superficie (TSK), flujo de calor en el suelo (GRDFLX) y emisividad superficial (EMISS). Los datos se descargaron en formato NetCDF desde el repositorio público del NCEP (<https://ncep.noaa.gov/pub/data/nccf/com/gfs/prod>). Posteriormente, se realizaron simulaciones con el modelo WRF para refinar la resolución espacial a 3.3 km, adaptada al área de estudio.

C. Preprocesamiento de Datos

Una vez generados los archivos NetCDF del WRF, se procedió a su procesamiento utilizando Python. Se extrajeron las variables meteorológicas relevantes, las cuales fueron normalizadas en un rango [0, 1] mediante la aplicación de una función de escalamiento lineal:

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)}. \quad (2)$$

La normalización se aplicó a las variables: TSK, EMISS, SWDOWN, GLW, GRDFLX, T2, PSFC y Q2. Además, los coeficientes de normalización (pendiente y ordenada) fueron almacenados en archivos ‘coeffs’ para garantizar su reutilización durante las predicciones en los años posteriores. El conjunto de datos resultante fue estructurado como un arreglo de entradas por observación, optimizado para su procesamiento por redes neuronales. Posteriormente, este conjunto normalizado fue utilizado como entrada para el modelo SOM implementado mediante la biblioteca **Somoclu**, permitiendo la clasificación no supervisada de patrones espaciales de ETo.

D. Aplicación de SOM a través de Somoclu

D1. Configuración del Modelo: El modelo SOM fue implementado utilizando la biblioteca **Somoclu**, aprovechando su capacidad de entrenamiento en arquitecturas multicore y compatibilidad con unidades de procesamiento gráfico (GPU).

Se definieron los siguientes hiperparámetros para el entrenamiento:

- **Tamaño de la rejilla** ($m \times n$): 6×6 nodos (36 neuronas).
- **Inicialización de pesos**: Análisis de Componentes Principales (PCA).
- **Tipo de vecindario**: *bubble*.
- **Radio inicial** (radius0): 2.5.
- **Radio final** (radiusN): 0.5.
- **Enfriamiento del radio**: lineal.
- **Número de épocas de entrenamiento**: 10.
- **Función de distancia**: Euclidiana.

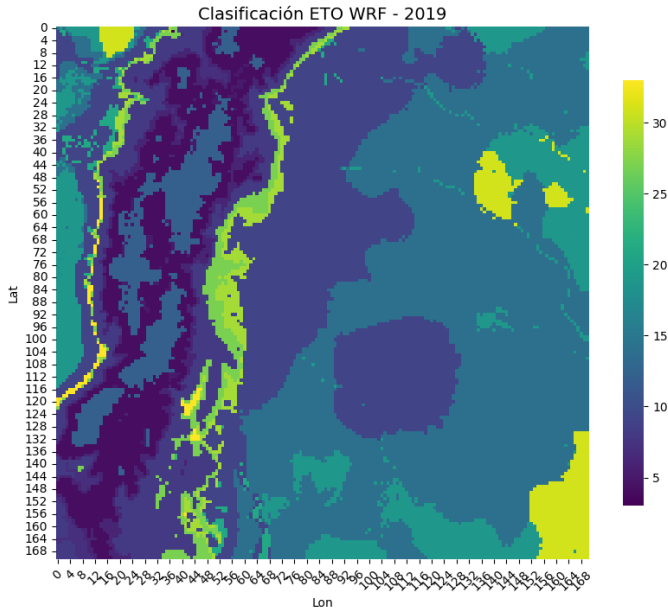


Figure 3. Clasificación espacial de la ETo mediante SOM: 1 de enero de 2019.

D2. Entrenamiento y Clasificación: El entrenamiento inicial se realizó utilizando exclusivamente los datos del año 2019. Posteriormente, se utilizó el modelo entrenado para clasificar los datos de los años 2017 y 2018, sin necesidad de reentrenamiento, preservando así la homogeneidad de los clústeres a lo largo de los años. El proceso de clasificación consistió en asignar a cada muestra del dataset la neurona (Best Matching Unit, BMU) más cercana en el mapa de características, basado en la distancia euclidiana mínima.

D3. Evaluación del Modelo: La calidad del modelo se evaluó a través del **Error de Cuantización (QE)**, definido como

$$QE = \frac{1}{N} \sum_{i=1}^N \|x_i - w_{c(i)}\|, \quad (3)$$

donde:

- N : número total de patrones de entrada (vectores de datos) considerados.
- x_i : el i -ésimo vector de datos de entrada.
- $w_{c(i)}$: el vector de peso (codebook vector) de la neurona ganadora correspondiente a x_i

- $\|x_i - w_{c(i)}\|$: la norma (distancia euclidiana) entre el vector de datos x_i y el vector de peso de su BMU.

Se utilizó la métrica de distorsión para comparar la cohesión de los clústeres en los años 2017, 2018 y 2019.

D4. Plataforma Computacional: Todo el procesamiento, entrenamiento y clasificación fue realizado en la plataforma **Google Colab**, utilizando TPUs (Tensor Processing Units) para acelerar el entrenamiento de Somoclu y facilitar el manejo de grandes volúmenes de datos.

IV. RESULTADOS Y DISCUSIÓN

A. Clasificación espacial de ETo mediante SOM

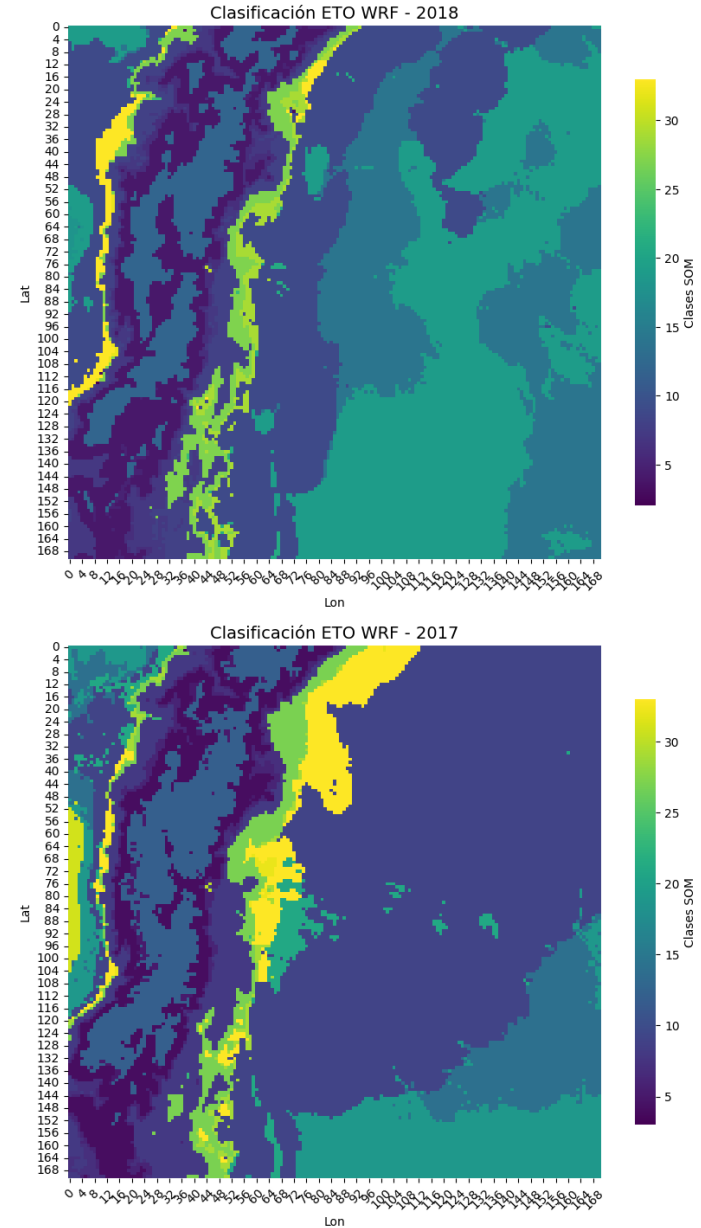


Figure 4. Clasificación espacial de la ETo para los años 2017 y 2018.

La clasificación espacial de la ETo a través del modelo SOM optimizado con **Somoclu** permitió identificar agrupamientos coherentes en las regiones Sierra y Amazonía del

Ecuador¹. En particular, la Figura 3 muestra la clasificación resultante para el 1 de enero de 2019, donde se observan agrupaciones espaciales consistentes en las zonas montañosas, evidenciando patrones climáticos relativamente estables. Específicamente, cada color en el mapa representa una clase distinta de comportamiento atmosférico respecto a la ETo. El modelo SOM fue configurado para generar un total de 36 clústeres posibles (6×6 nodos); sin embargo, como se observa en la barra de colores de la Figura 3, las clases efectivamente ocupadas alcanzan aproximadamente 30 clases distintas. Esta ligera reducción en el número de clases es una característica esperable en técnicas de clasificación no supervisada, donde la densidad de los datos y su distribución espacial determinan la ocupación real del mapa de características. Estos resultados son consistentes con lo reportado por [10], donde se señala que, bajo configuraciones de entrenamiento profundo (*deeper learning*), los patrones de clasificación de la ETo tienden a estabilizarse espacialmente, en especial en zonas montañosas. En contraste, las regiones de transición entre la Sierra y la Amazonía presentan una mayor variabilidad de clases, atribuida a la influencia combinada de factores topográficos y gradientes de humedad.

La Figura 4 muestra las clasificaciones espaciales obtenidas para el mismo día (1 de enero) en los años 2017 y 2018, utilizando el modelo previamente entrenado con los datos de 2019. Aquí, se observa que los patrones de clasificación espacial se mantienen notablemente consistentes entre los tres años analizados. En ambos mapas, la región andina preserva su agrupamiento característico, indicando estabilidad climática en las zonas de alta montaña, mientras que las zonas de transición hacia la Amazonía presentan nuevamente una mayor dispersión de clases, reflejando su inherente complejidad hidroclimática. Cabe destacar que, aunque las condiciones meteorológicas varían naturalmente año a año, los patrones fundamentales de evapotranspiración identificados a través del SOM demuestran resiliencia ante estas fluctuaciones, lo que evidencia la capacidad del modelo para generalizar de manera efectiva más allá del año de entrenamiento. Estos resultados no solo confirman la idoneidad del modelo para clasificar espacialmente la ETo en distintos periodos de tiempo, sino que también subrayan su utilidad potencial para aplicaciones prácticas en la planificación de estrategias de riego y la gestión de recursos hídricos a largo plazo.

B. Evaluación de la cohesión de clústeres

Para evaluar la calidad del agrupamiento generado por el SOM, se calculó el QE de manera mensual para los años 2017, 2018 y 2019, como se ilustra en la Figura 5. Aquí, se presenta la evolución de esta métrica a lo largo de los doce meses en cada uno de los años considerados.

En general, se observa que la variación del QE es relativamente estable a lo largo del año en los tres periodos

¹El modelo fue entrenado utilizando datos correspondientes al año 2019 y posteriormente aplicado a los años 2017 y 2018, permitiendo validar su capacidad de generalización.

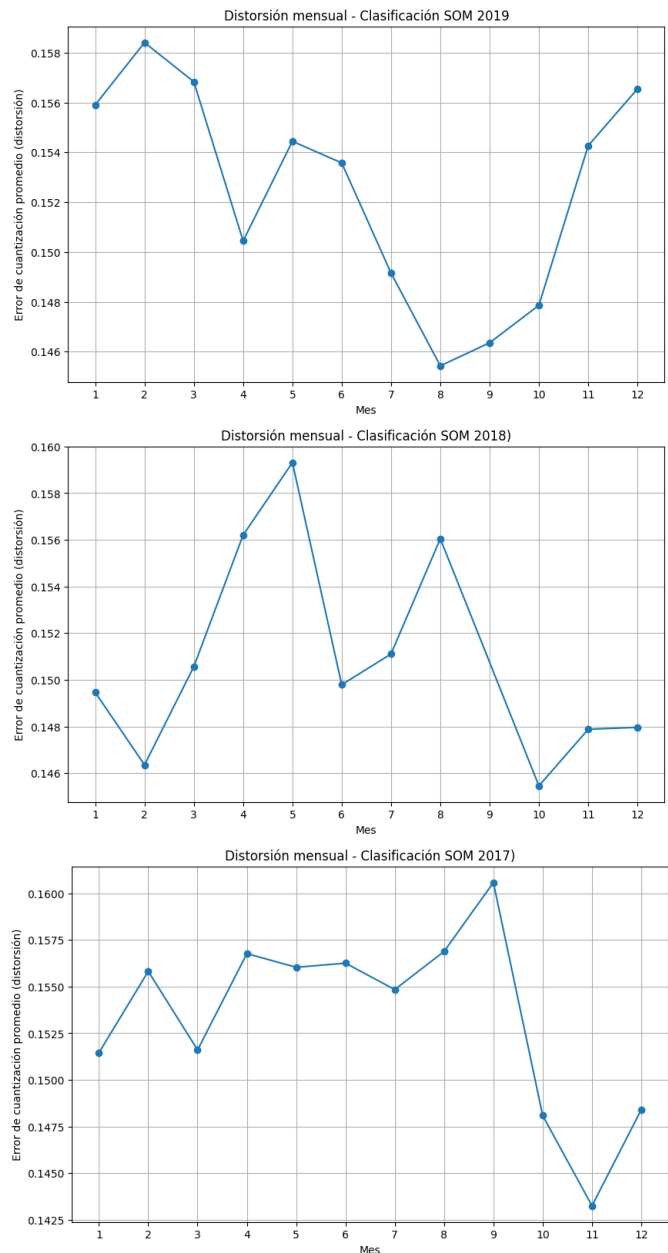


Figure 5. Error de cuantización (QE) de enero a diciembre en los años 2017, 2018 y 2019.

evaluados. Esta estabilidad indica que los patrones de ETo agrupados por el SOM mantienen una coherencia espacial robusta independientemente de las variaciones estacionales típicas de la atmósfera ecuatoriana. Para el año 2019, el QE oscila entre aproximadamente 0.146 y 0.158, presentando un descenso paulatino hacia los meses centrales del año (junio a septiembre), donde se alcanza el valor mínimo de error. Este comportamiento sugiere que en dichos meses los patrones de ETo son más homogéneos, posiblemente debido a condiciones atmosféricas más estables en la región durante la temporada seca.

Por otro lado, para el año 2018, el QE presenta mayor variabilidad intermensual, con picos destacados en los meses

de mayo y septiembre. Esta mayor dispersión puede estar asociada a la transición entre estaciones húmedas y secas, que afectan la estructura espacial de la evapotranspiración. No obstante, a pesar de esta variabilidad, los valores absolutos de QE se mantienen dentro de un rango aceptable, garantizando la robustez de la clasificación realizada. Por su parte, en el año 2017, el comportamiento del QE muestra picos de error relativamente moderados, con máximos en los meses de abril y octubre, aunque también presenta una baja marcada en diciembre. Este patrón sugiere que, si bien existen fluctuaciones estacionales, la estructura subyacente de los clústeres ETo no se altera significativamente, validando así la estabilidad del modelo SOM en múltiples condiciones climáticas.

Es importante mencionar que, los resultados previos son coherentes con lo reportado por [10], quienes encontraron que, aunque la ETo presenta variabilidad estacional natural, el uso de SOM entrenado sobre variables meteorológicas simuladas permite capturar patrones climáticos de fondo que se mantienen persistentes en el tiempo. El mantenimiento de niveles de QE relativamente bajos y consistentes refuerza la validez del enfoque WRF + SOM para estudios de clasificación hidrometeorológica en regiones de difícil acceso o de monitoreo limitado. Finalmente, en conjunto, la evaluación del QE confirma que el modelo SOM no solo clasifica adecuadamente los patrones espaciales de ETo en un año de entrenamiento (2019), sino que además logra generalizar su desempeño de forma efectiva sobre años adyacentes (2017 y 2018), preservando tanto la cohesión interna de los clústeres como la robustez temporal en la clasificación.

V. CONCLUSIONES

Los resultados obtenidos en este trabajo demuestran que la combinación de datos meteorológicos simulados por el modelo WRF y el uso de mapas auto-organizados (SOM) constituye una estrategia robusta para la clasificación espacial de patrones de ETo en regiones complejas como la Sierra y Amazonía norte del Ecuador. Particularmente, la implementación de la biblioteca Somoclu permitió un entrenamiento eficiente y escalable de mapas SOM de gran tamaño, superando las limitaciones de desempeño reportadas en estudios previos de la literatura que utilizaron versiones más básicas de SOM. Somoclu no solo facilitó el procesamiento de grandes volúmenes de datos meteorológicos simulados, sino que también garantizó una preservación adecuada de la topología espacial de los patrones detectados. La estabilidad de las clasificaciones a lo largo de los años 2017, 2018 y 2019, así como los bajos valores de QE obtenidos, confirman que el enfoque propuesto es viable para aplicaciones operativas en gestión hídrica y planificación agrícola en zonas vulnerables, incluso en ausencia de observaciones meteorológicas continuas. En síntesis, el uso de Somoclu representa un avance metodológico significativo que potencia la aplicabilidad de las técnicas de aprendizaje no supervisado en estudios hidrometeorológicos de alta resolución espacial y temporal.

ACKNOWLEDGMENT

El autor extiende su más sincera gratitud a Scott Williams por proporcionar la data empleada en este estudio.

REFERENCES

- [1] C. Carter and S. Liang, "Comprehensive evaluation of empirical algorithms for estimating land surface evapotranspiration," *Agricultural and Forest Meteorology*, vol. 256-257, pp. 334-345, 2018.
- [2] S. Ahmadi and Z. Javanbakht, "Assessing the physical and empirical reference evapotranspiration (eto) models and time series analyses of the influencing weather variables on eto in a semi-arid area," *Journal of Environmental Management*, vol. 276, p. 111278, 2020.
- [3] R. Allen, L. Pereira, D. Raes, and M. Smith, *Crop Evapotranspiration—Guidelines for Computing Crop Water Requirements—FAO Irrigation and Drainage Paper 56*, ser. FAO Irrigation and Drainage Paper. FAO, Rome, 1998, vol. 56, no. 300(9).
- [4] J. Sheffield, E. Wood, M. Pan, H. Beck, G. Coccia, and A. Serrat-Capdevila, "Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions," *Water Resources Research*, vol. 54, no. 12, pp. 9724-9758, 2018.
- [5] J. Sheffield, E. F. Wood, M. Pan, H. Beck, G. Coccia, A. Serrat-Capdevila, and K. Verbist, "Satellite remote sensing for water resources management: Potential for supporting sustainable development in data-poor regions," *Water Resources Research*, vol. 54, no. 12, pp. 9724-9758, 2018. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017WR022437>
- [6] V. Tallapragada, *Overview of the NOAA/NCEP Operational Hurricane Weather Research and Forecast (HWRF) Modelling System*. Dordrecht: Springer Netherlands, 2016, pp. 51-106. [Online]. Available: https://doi.org/10.5822/978-94-024-0896-6_3
- [7] F.-J. Chang, L.-C. Chang, H.-S. Kao, and G.-R. Wu, "Assessing the effort of meteorological variables for evaporation estimation by self-organizing map neural network," *Journal of Hydrology*, vol. 384, no. 1, pp. 118-129, 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022169410000399>
- [8] D. Wang, Y. Zhan, T. Yu, Y. Liu, X. Jin, X. Ren, X. Chen, and Q. Liu, "Improving meteorological input for surface energy balance system utilizing mesoscale weather research and forecasting model for estimating daily actual evapotranspiration," *Water*, vol. 12, no. 1, 2020. [Online]. Available: <https://www.mdpi.com/2073-4441/12/1/9>
- [9] W. P. Köppen, *Die Klimate der Erde: Grundriss der Klimakunde...* Walter de Gruyter & Company, 1923.
- [10] M. Solis-Aulestia, I. Pineda, E. J. Piispa, and S. L. Williams, "Evaluation of evapotranspiration classification using self organizing maps and weather research and forecasting variables," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2022, pp. 3195-3198.
- [11] I. Pineda, E. J. Piispa, S. L. Williams, and M. Solis-Aulestia, "Meso-scale standard evapotranspiration 'climate' classification derived from numerical weather prediction models and artificial intelligence," in *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023, pp. 3842-3845.
- [12] D. Wang, Y. Zhan, T. Yu, Y. Liu, X. Jin, X. Ren, X. Chen, and Q. Liu, "Improving meteorological input for surface energy balance system utilizing mesoscale weather research and forecasting model for estimating daily actual evapotranspiration," *Water*, vol. 12, no. 1, p. 9, 2019.

- [13] T. Biggs, M. Marshall, and A. Messina, "Mapping daily and seasonal evapotranspiration from irrigated crops using global climate grids and satellite imagery: Automation and methods comparison," *Water Resources Research*, vol. 52, no. 9, pp. 7311–7326, 2016.
- [14] H. Tabari, "Evaluation of reference crop evapotranspiration equations in various climates," *Water Resources Management*, vol. 24, pp. 2311–2337, 2010.
- [15] J. G. Powers, J. B. Klemp, W. C. Skamarock, C. A. Davis, J. Dudhia, D. O. Gill, J. L. Coen, D. J. Gochis, R. Ahmadov, S. E. Peckham, G. A. Grell, J. Michalakes, S. Trahan, S. G. Benjamin, C. R. Alexander, G. J. Dimego, W. Wang, C. S. Schwartz, G. S. Romine, and M. G. Duda, "The weather research and forecasting model: Overview, system efforts, and future directions," *Bulletin of the American Meteorological Society*, vol. 98, no. 8, pp. 1717–1737, 2017.
- [16] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [17] R. Olawoyin, A. Nieto, R. L. Grayson, F. Hardisty, and S. Oyewole, "Application of artificial neural network (ann)–self-organizing map (som) for the categorization of water, soil and sediment quality in petrochemical regions," *Expert Systems with Applications*, vol. 40, no. 9, pp. 3634–3648, 2013.
- [18] P. Wittek, S. Gao, I.-S. Lim, and X. Zhao, "Somoclu: An efficient parallel library for self-organizing maps," *Neurocomputing*, vol. 150, pp. 639–643, 2015.