ANÁLISIS DE CLUSTER DE SERIES DE TEMPERATURA MEDIA DIARIA EN LA REGIÓN PAMPEANA: COMPARACIÓN ENTRE DISTANCIA EUCLIDIANA Y DISTANCIA BASADA EN CORRELACIÓN

SUÁREZ, Josefina[1]; BARBONA, Ivana[2]

[1]Cátedra de Estadística. Facultad de Ciencias Agrarias. UNR

Palabras claves: Análisis de Cluster; distancia; series de tiempo.

INTRODUCCIÓN

El **Análisis de Clúster de series de tiempo** comprende métodos de clústers diseñados o adaptados para manejar datos con una estructura temporal. En este contexto, cabe preguntarse: ¿de qué manera la elección de la **medida de distancia** (la cual define cuándo dos series se consideran similares o diferentes) condiciona el agrupamiento de las series temporales?

OBJETIVO

Comparar la elección de una medida de distancia basada en magnitud y una basada en patrón sobre el agrupamiento de series de temperatura media diaria de la región pampeana.

DATOS

34 **series diarias de temperatura media** de la región pampeana (1962-2023) provenientes de estaciones sinópticas del Servicio Meteorológico Nacional, previamente procesadas.

METODOLOGÍA

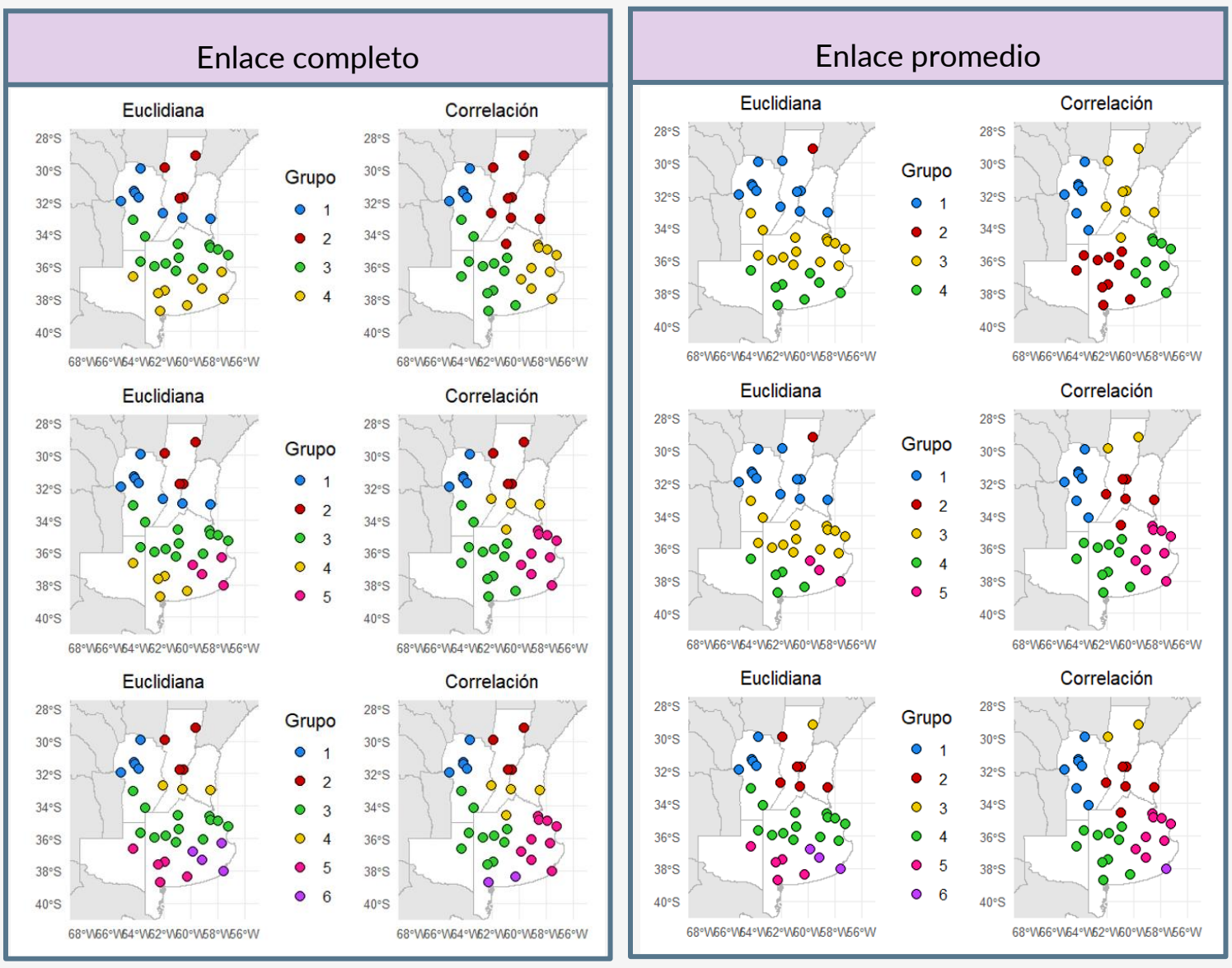
Distancia **euclidiana** y por **correlación** $\left(1 - r(\boldsymbol{X}, \boldsymbol{Y})\right)$

Análisis de Clúster **jerárquico** aglomerativo con enlaces **completo** y **promedio**

4, 5 y 6 grupos

RESULTADOS Y DISCUSIÓN

Figura 1. Agrupaciones obtenidas por clustering jerárquico aglomerativo con enlace completo (izquierda) y enlace promedio (derecha), para ambas medidas de distancia y distinta cantidad de grupos. Temperatura media diaria, período comprendido entre 01-01-1962 y 05-08-2023.



*Existen ciertos agrupamientos **estables***

*Las **diferencias** entre las distancias se manifiestan en las **agrupaciones**, en especial con 4 y 5 grupos.*

*Distancia **euclidiana** tiende a dividir en **franjas horizontales**, con mayores **valores** de temperatura hacia el norte y menores hacia el sur.*

*Distancia por **correlación** separa **verticalmente** en este y oeste, tanto en el norte como en el sur. Los patrones temporales de las localidades de la zona costera son similares, diferenciándose de las localidades más continentales del oeste.*

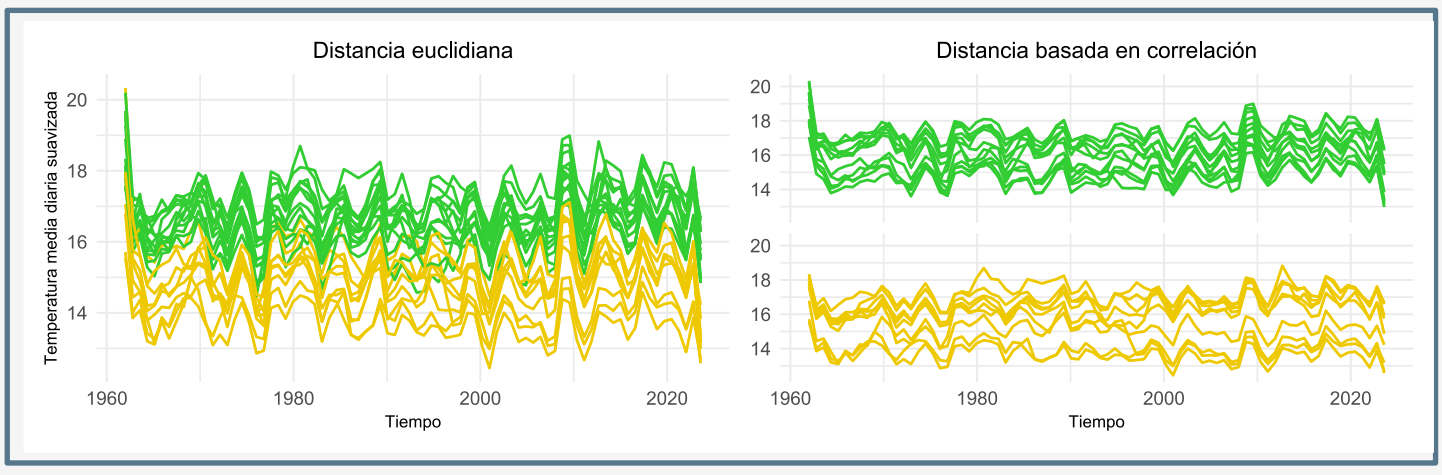


Figura 2. Series suavizadas de temperatura media diaria según pertenencia a clúster 3 (verde) o 4 (amarillo), obtenidos con enlace completo con la partición de cuatro grupos, para ambas distancias

CONCLUSIONES

Los clusters de las series diarias de temperatura media, para un mismo criterio de enlace y número de grupos, **cambian** de acuerdo con la **medida de distancia** adoptada, tanto en su **estructura** como en su **caracterización**. La correlación agrupa series con patrón similar aunque sus magnitudes difieran, mientras que la euclidiana reúne series de valores similares. Por eso, el enfoque basado en correlación resulta preferible cuando interesa la forma en que varía la temperatura.

Dado que ambas medidas captaron aspectos distintos de las localidades, la decisión debe ajustarse al **objetivo del análisis** y a la **naturaleza de las series** bajo estudio.

REFERENCIAS