

Dos aplicaciones de Mapas Auto-organizativos: *Clustering* y Generalización

Dr Diego RIVERA

Laboratorio de Hidrología Ambiental - UdeC

I CLEIA - VII CONEIA 2010
Universidad Nacional Agraria La Molina, Perú

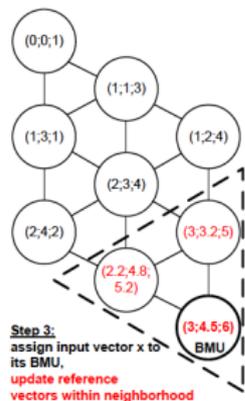
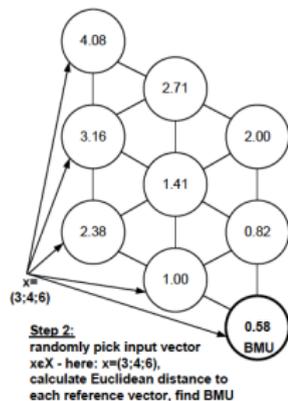
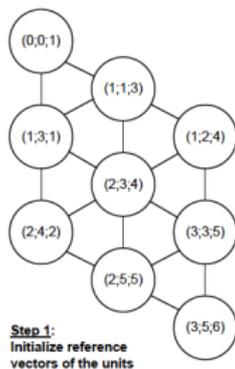
Outline

- 1 Self-organizing Maps
- 2 Agrupamiento
 - Datos de muestras de suelos
 - Precipitación y Temperatura Superficial del Mar (SST)
- 3 Pronóstico de Precipitación basado en Temperatura Superficial del Mar (SST)

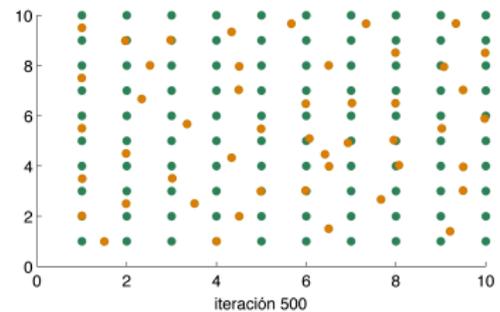
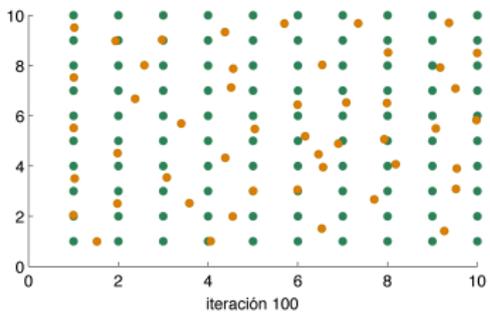
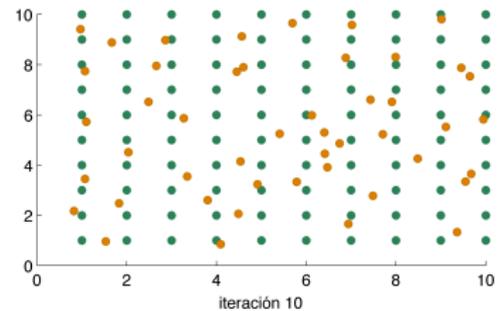
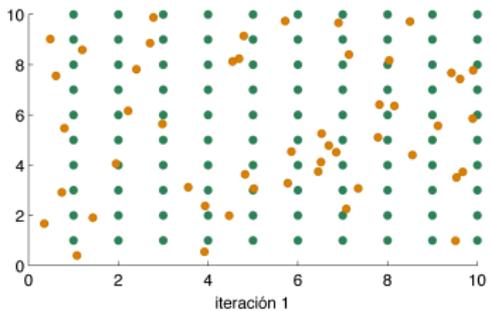
- Los Mapas Auto-organizativos (MAOs) son un tipo de Redes Neuronales Artificiales que aplican un algoritmo de aprendizaje no supervisado apropiado para el agrupamiento (*clustering*), visualización y abstracción de datos multidimensionales^{1,6}.
- MAOs mapea los vectores de entrada en posiciones similares de una grilla 2-dimensionalmanteniendo la topología¹.
- Patrones de entrada similares tendrán posiciones cercanas en el mapa.
- Algunas neuronas del mapa tendrán cierta *afinidad* con los correspondientes patrones de entrada, permitiendo *explorar* y *extraer* patrones.
- Capacidad de combinar variables en la construcción de las matrices de entrenamiento^{6,2}.
- Más allá de nombres sofisticados, el algoritmo se basa en calcular distancias Euclidianas ;

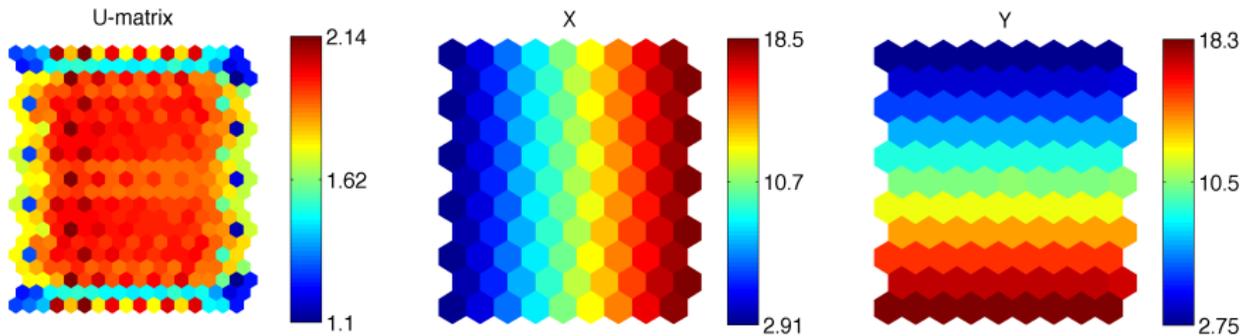


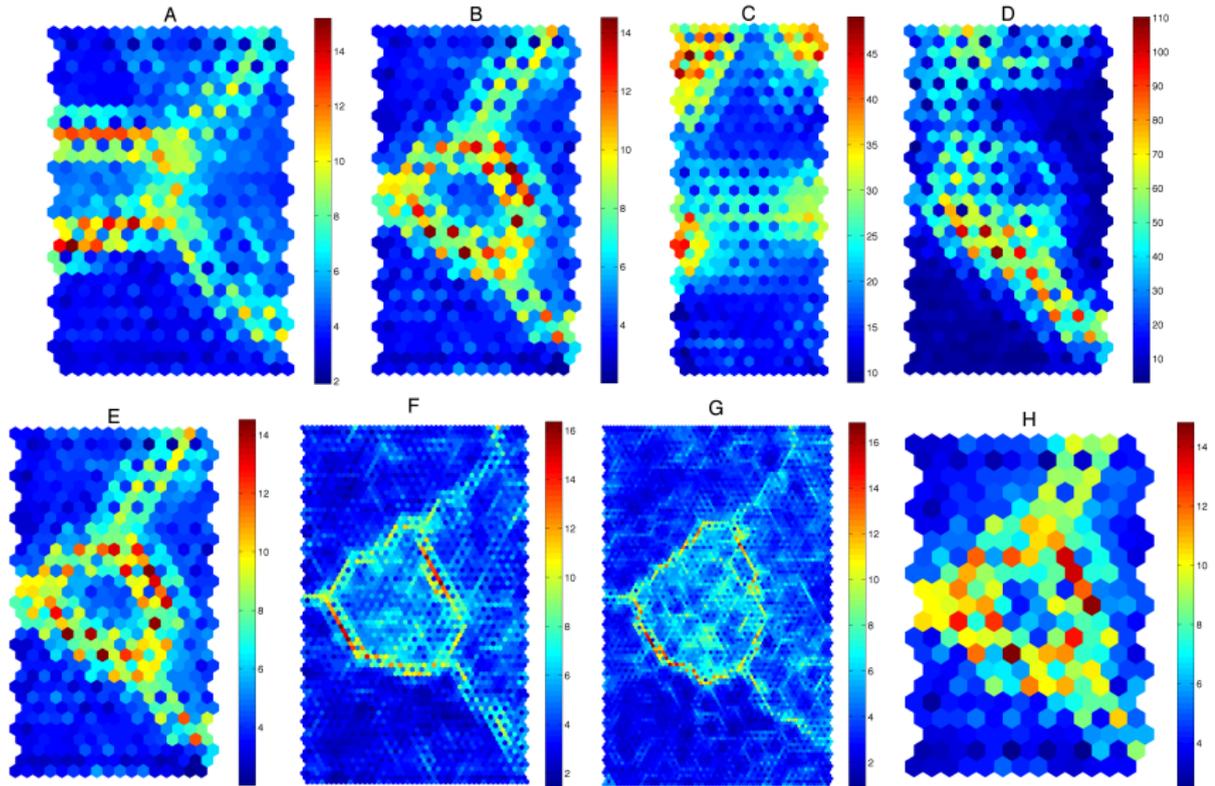
- 1 Cada i -ésima neurona está asociada a un vector prototipo \mathbf{p} .
- 2 Se elige aleatoriamente un vector del conjunto de entrenamiento \mathbf{m} .
- 3 Se calculan las distancias entre cada \mathbf{p} y \mathbf{m} .
- 4 Se encuentra la Best-matching Unit (BMU): $\|\mathbf{m} - \mathbf{p}_b\| = \min_i \{\|\mathbf{m} - \mathbf{p}_i\|\}$
- 5 Se actualizan los pesos: la BMU y sus vecinos se mueven a una posición más cercana a \mathbf{m} : $\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t) \cdot H_b(t) \cdot [\mathbf{m} - \mathbf{p}_i(t)]$



Step 4: repeat step 2 + 3 for each new $x \in X$







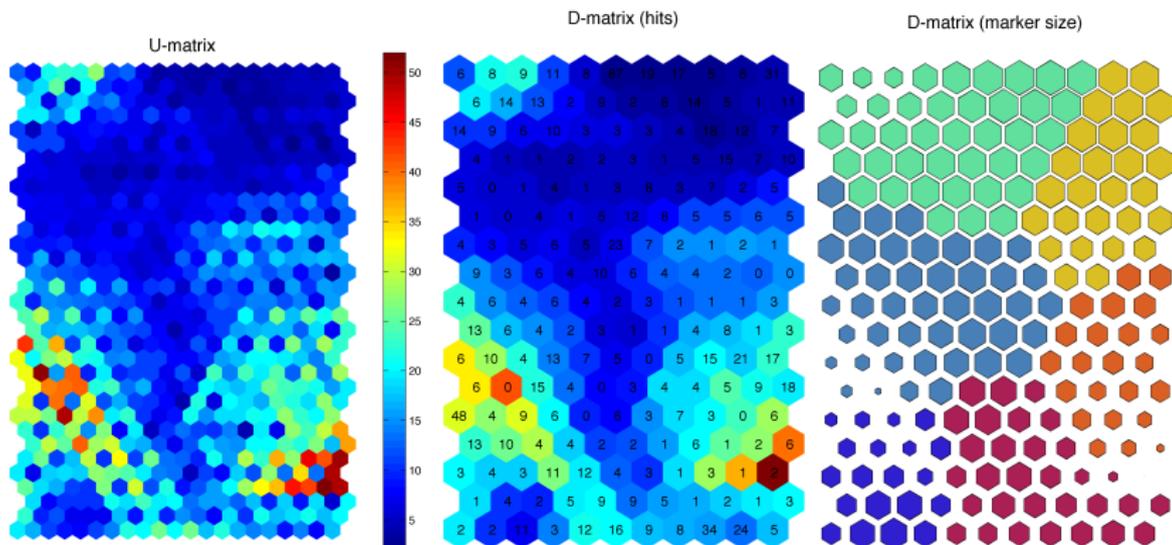


- *Data mining* busca responder una pregunta o resolver un problema basado en los datos (e información) disponible⁶:
 - 1 Exploración
 - 2 Pre-procesamiento
 - 3 Construcción y evaluación de modelos
 - 4 Consolidación del conocimiento
- El análisis de datos *confirmatorio* busca responder preguntas como ¿Los datos confirman la hipótesis?, mientras que el análisis de datos *exploratorio* busca responder preguntas como ¿Qué pueden decirme los datos acerca de esta asociación?³
- Las bases de datos de suelos son un buen caso de estudio: no etiquetado.



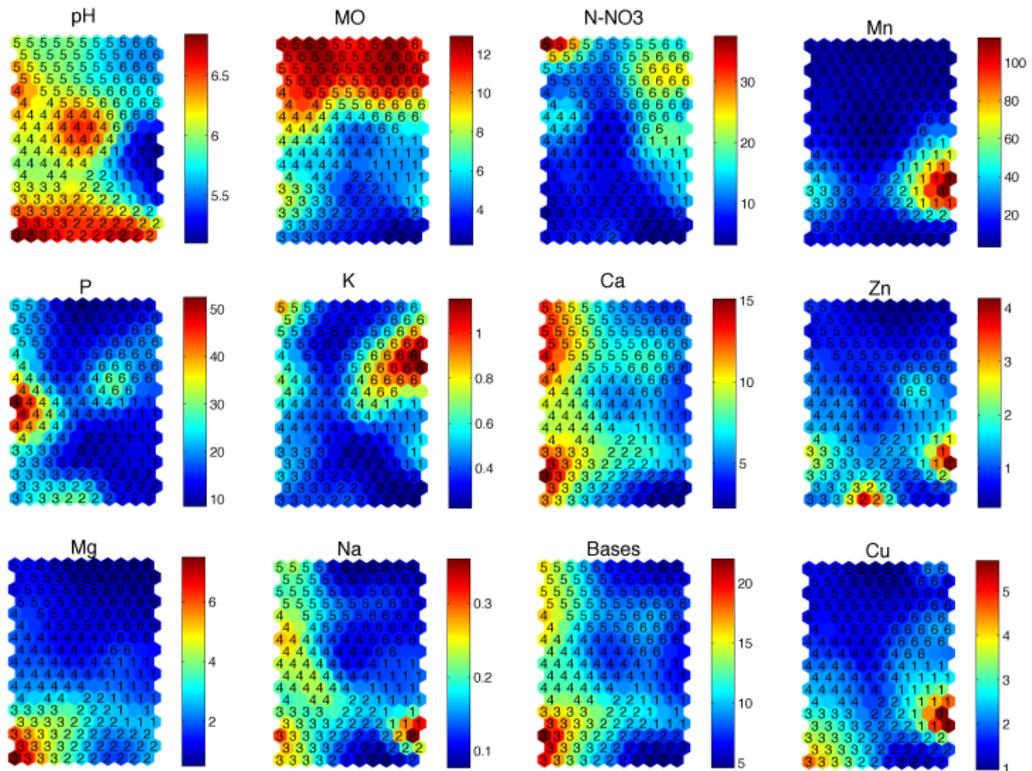
- Calidad de suelos es un concepto multidimensional⁷. No un parámetro, sino un conjunto de parámetros.
- Si se considera cada muestra de suelo como un vector, una base de datos puede definirse como una matriz $\mathbf{Q}_{s \times p}$ con p columnas (parámetros) y s filas (muestras), que se encuentran en un espacio p -dimensional⁸.

Resultados de Rivera, Sandoval, Arumí y Billib (2010) Exploring soil databases: A Self-organizing Maps approach. Enviado a *Geoderma*.





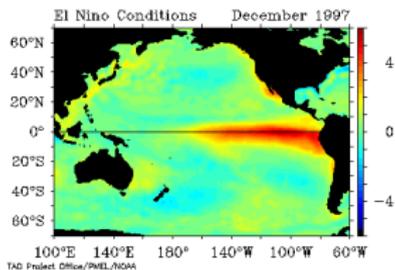
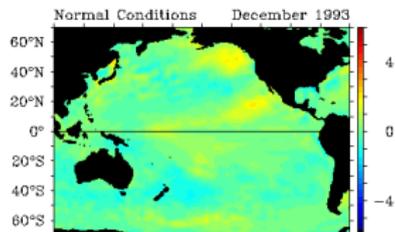
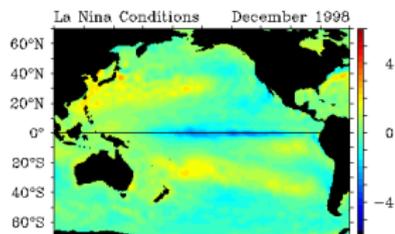
Datos de muestras de suelos



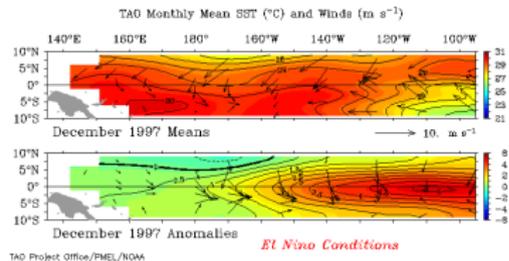
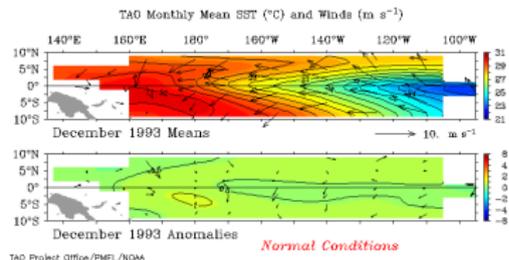
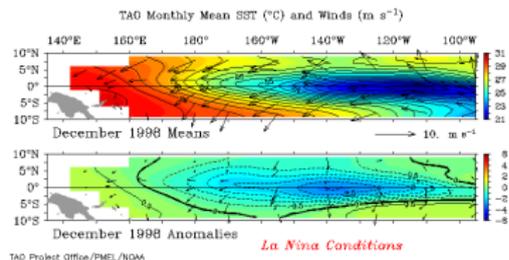


Precipitación y Temperatura Superficial del Mar (SST)

Reynolds Monthly SST Anomalies (°C)



TAD Project Office/PMEL/NOAA

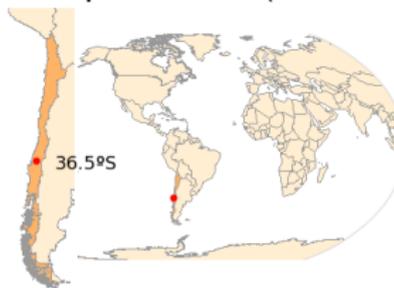


<http://www.pmel.noaa.gov/tao/el-nino/el-nino-story.html>





- **El Niño**: precipitación sobre el promedio 30° - 35° S en invierno (Jun-Jul-Ago) y 35° - 38° S en primavera (Oct-Nov). Anomalías

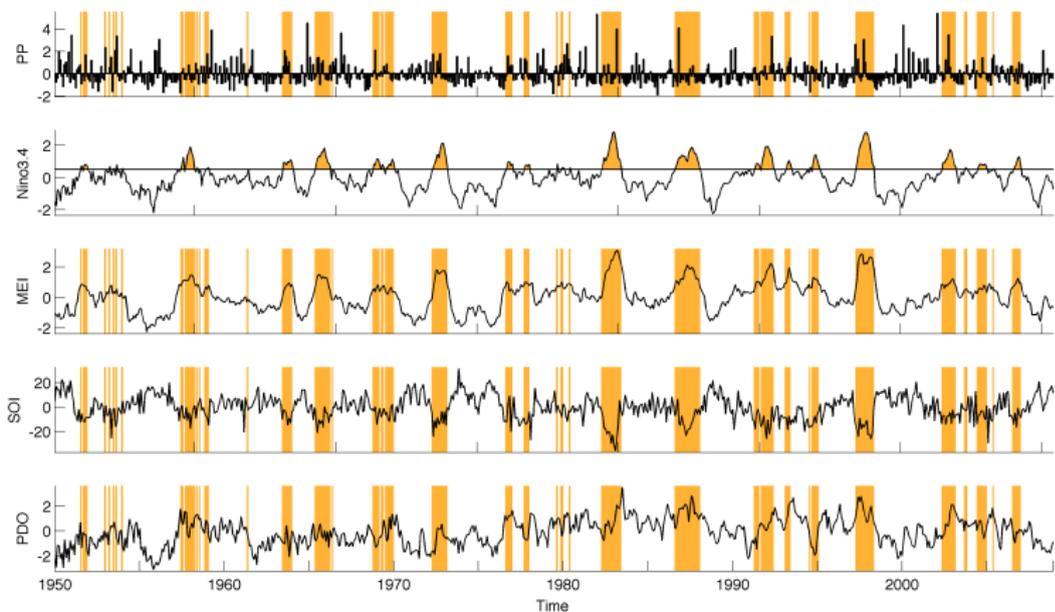


opuestas para **La Niña**⁴.

- Varianza de la precipitación explicada por SST en la región Niño 3.4 es menor en Apr-May, y mayor en Sep-Oct; mejor correlación para años cálidos que para los fríos⁵.
- Importantes meses para la agricultura.

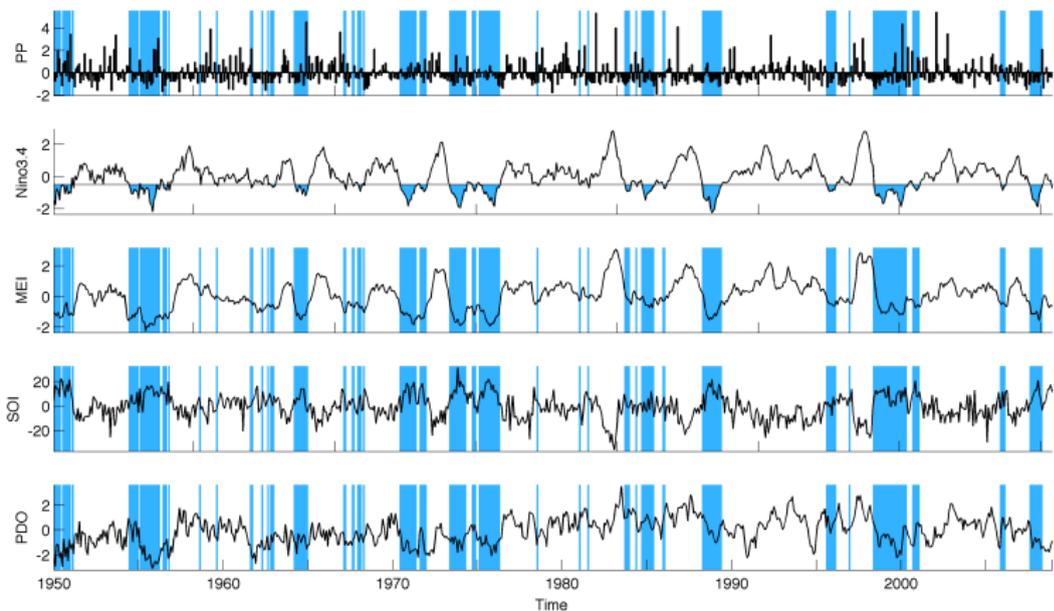


Precipitación y Temperatura Superficial del Mar (SST)





Precipitación y Temperatura Superficial del Mar (SST)





$$T = [\mathbf{PPA}_{\text{apr}} \mathbf{Nino3.4}_{\text{feb}} \mathbf{Nino3.4}_{\text{mar}} \mathbf{DIF}_{\text{Nino3.4}} \mathbf{MEI}_{\text{mar}}]$$

- Anomalía de precipitación estandarizada para Abril (**PPA_{apr}**) como "predictando" \Rightarrow patrones de SST asociados a precipitación de Abril ($\mu=72$ mm, $\sigma=59.5$ mm).
- Anomalía de SST en Nino3.4 (5° N - 5° S and 170° - 120° W) para Feb (**Nino3.4_{feb}**) y Mar (**Nino3.4_{mar}**).
- Diferencia entre **Nino3.4_{mar}** y **Nino3.4_{feb}** (**DIF_{Nino3.4}**) por persistencia.
- MEI index para Mar **MEI_{mar}** como indicador del estado del sistema atmósfera-oceáno.



Validación secuencial

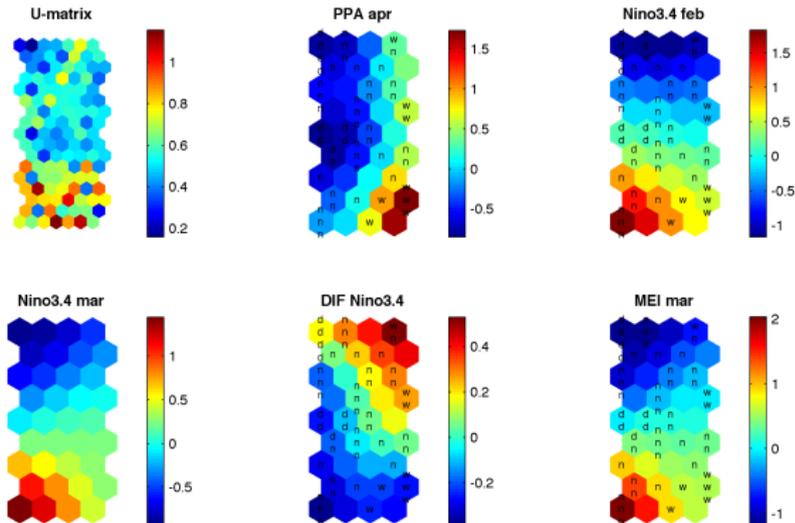
Simula una aplicación real

- 1 SOM entrenado con datos 1900 a X.
- 2 Buscar BMU para año $X+1$, enmascarando la precipitación (o sólo es un ejercicio académico).



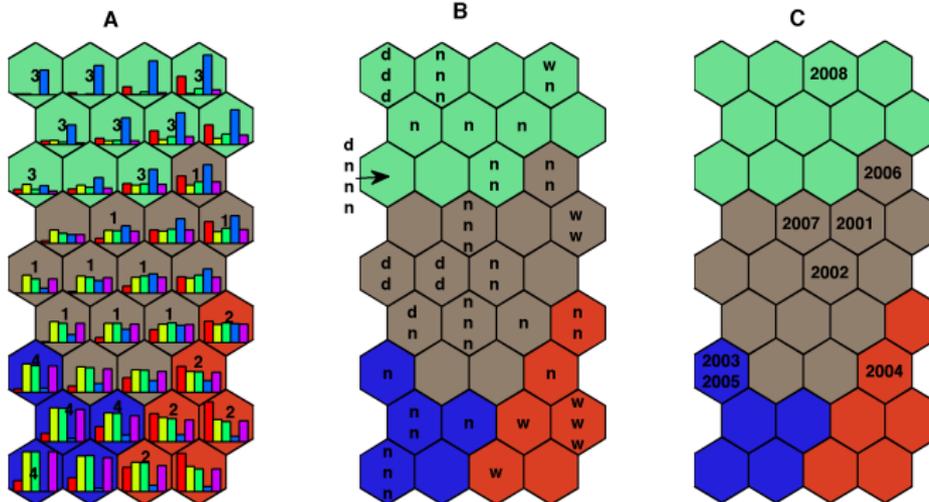
Precipitación y Temperatura Superficial del Mar (SST)

Resultados de Rivera, Uvo y Arumi y Billib (2010) Association between rainfall in Central Chile and Sea Surface Temperature patterns: A Self-organizing Map approach. Enviado a *International Journal of Climatology*.





Precipitación y Temperatura Superficial del Mar (SST)





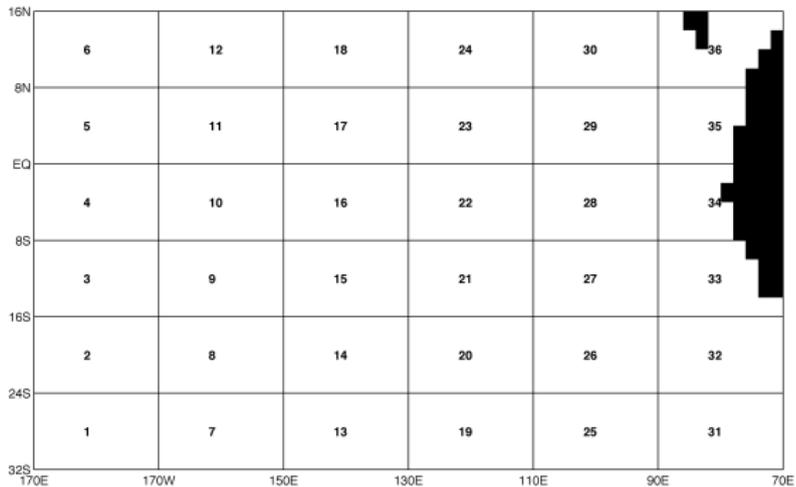
- 1 Campo de anomalías (ANOM), en cada i -ésima cedal en el campo de SST (promedio 1900-2008):

$$ANOM_i = \frac{SST_i - \overline{SST}_i}{\sigma_i}$$

- 2 La región 170°E-70°W x 16°N-32°S fue dividida en 36 regiones. Promedios para cada región.

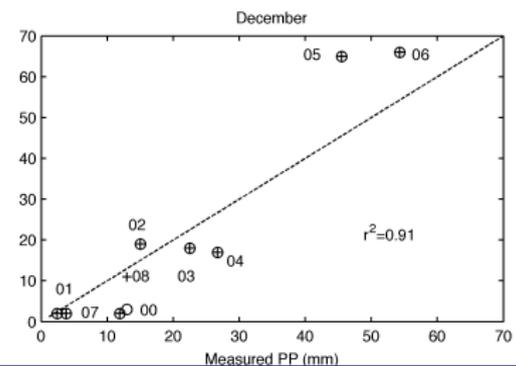
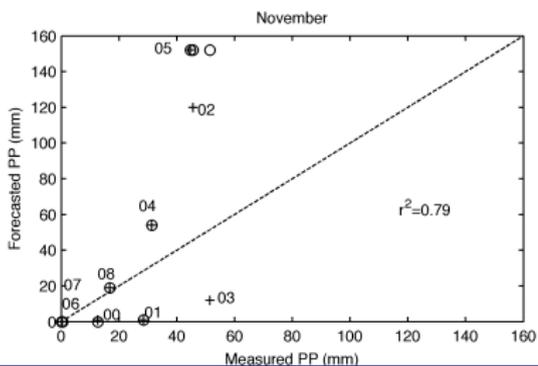
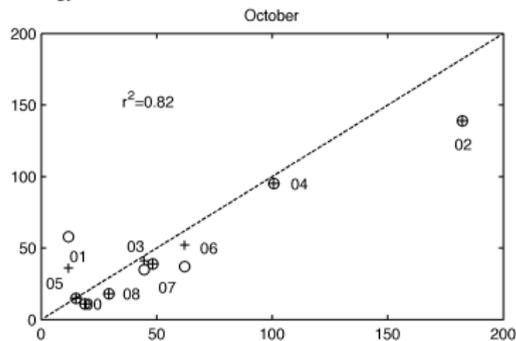
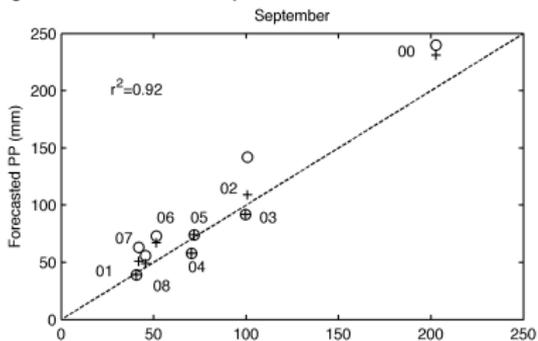
$$\mathbf{T} = [PPA_{apr} \ ANOM_{mar}^r \ ANOM_{feb}^r]$$

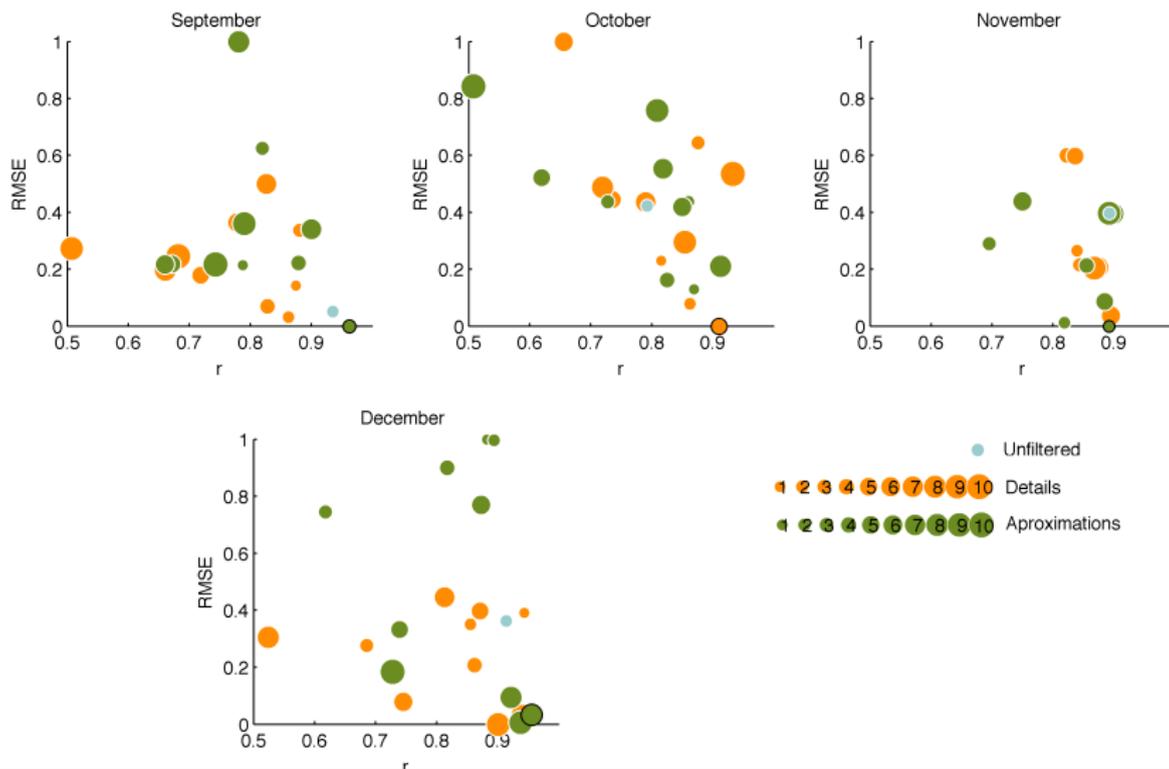
- 3 Campos promediados, pero el MAO es entrenado usando campos filtrados. Remoción de ruido.
- 4 Discrete wavelet transform usando el algoritmo *à trous*: redundante, pero los campos filtrados tiene la misma longitud de las series originales.





Resultados de Rivera, Uvo, Lillo, Billib y Arumí (2010) Forecasting monthly Precipitation in Central Chile: A Self-Organizing Map approach using filtered Sea Surface Temperature. Enviado a *Theoretical and Applied Climatology*.





Dr Diego Rivera, Ingeniero Civil
dirivera@udec.cl
Director de Programas de Magíster y Doctorado
Facultad de Ingeniería Agrícola
Profesor Asistente
Departamento de Recursos Hídricos
Universidad de Concepción

- [1] M Herbst and M Casper. Towards model evaluation and identification using Self-Organizing Maps. *Hydrology and Earth System Sciences*, 12:657–667, 2008.
- [2] E Manolakos, H Virani, and V Novotny. Extracting knowledge on the links between the water body stressors and biotic integrity. *Water Research*, 41:4041–4050, 2007.
- [3] W.L. Martinez and A.R. Martinez. *Exploratory data analysis with MATLAB*. CRC Press, 2005.
- [4] A. Montecinos and P. Aceituno. Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies. *Journal of Climate*, 16(2):281–296, 2003.
- [5] A. Montecinos, A. Díaz, and P. Aceituno. Seasonal diagnostic and predictability of rainfall in Subtropical South America based on Tropical Pacific SST. *Journal of Climate*, 13:746–758, 2000.
- [6] J Vesanto. *Data exploration process based on the Self-Organizing Maps*. PhD thesis, Helsinki University of Technology, Finland, 2002.
- [7] MB Villamil, FE Miguez, and GA Bollero. Multivariate Analysis and Visualization of Soil Quality Data for No-Till Systems. *Journal of Environmental Quality*, 37(6):2063, 2008.
- [8] R Webster. Statistics to support soil research and their presentation. *European Journal of Soil Science*, 52:331–340, 2001.