

Análisis socio-económico del Área Metropolitana de Valencia por medio de un Mapa Auto-organizado de Kohonen.

Ignacio Martínez de Lejarza Esparducer
Universidad de Valencia

BIBLID [0213-7525 (2001); 59; 145-167]

PALABRAS CLAVE: Mapas auto-organizados de Kohonen (SOM), Redes Neuronales, Análisis Cluster, Areas metropolitanas, Area metropolitana de Valencia, Análisis de datos socio-económicos.

KEY WORDS: Kohonen's Self-Organized Maps, Neural Networks, Cluster Analysis, Metropolitan Areas, Valencian Metropolitan Area, Socio-economical Data Analysis.

RESUMEN

En este trabajo se plantea la posibilidad de utilizar los modelos de redes neuronales auto-organizadas en el análisis de la configuración interna de un territorio. Tras mostrar las características esenciales del funcionamiento y de los algoritmos de aprendizaje de los modelos de Kohonen con salida unidimensional (L.V.Q.) y bidimensional (S.O.M.), se argumenta que el primero de ellos, aplicado al análisis de un territorio, puede suponer una alternativa interesante a los métodos estadísticos de análisis cluster y que, por otra parte, los Mapas Auto-organizados (S.O.M.) pueden llegar a ofrecer clasificaciones potencialmente más ricas; puesto que dan como resultado una agrupación de las entidades territoriales sobre categorías dispuestas en un plano (tipología bidimensional), lo que aporta la posibilidad de incorporar criterios topológicos al análisis de la clasificación obtenida.

A partir de estas consideraciones, se lleva a cabo una aplicación empírica: partiendo de la información socioeconómica disponible sobre los municipios del Área Metropolitana de Valencia se procede a su agrupación utilizando estos modelos y algunos métodos tradicionales de análisis Cluster. Se compara la calidad de las distintas agrupaciones en función de la homogeneidad interna de los grupos obtenidos. Y, por último, se interpretan los resultados a través de distintos análisis estadísticos.

ABSTRACT

This paper is about the possibility of applying Kohonen's neural model to the study of a Region's internal structure. After showing the main characteristics of the working and learning of Kohonen models with uni-dimensional (L.V.Q.) and bi-dimensional (S.O.M.) exit, it is stated that the first -applied to the analysis of a territory- can be an interesting alternative to the statistical methods of cluster analysis. On the other hand, the Self Organized Maps (S.O.M.) can be of richer grading, as they result in a clustering of the territorial entities about categories placed in a map (bi-dimensional typology) - which gives the possibility of incorporating topological criteria into obtained clustering.

After these considerations an empirical application is carried out: starting from the socio-economic information available about Valencia's Metropolitan Area towns. We go on to group them using these models and some traditional methods of Cluster study. The quality of the different groupings is compared according to the internal homogeneity of those obtained. And, lastly, the results are interpreted by different statistical and topological analysis (Manova/Discriminant).

1. INTRODUCCIÓN

Desde la aparición de los primeros trabajos empíricos de la Ecología Factorial y los analistas de Áreas Sociales, no han dejado de aplicarse distintas técnicas estadísticas de agrupación, clasificación y discriminación con el objeto de analizar la estructura interna de los territorios considerados, pretendiendo, en última instancia obtener esas unidades territoriales distintivas y diferenciadas en su comportamiento social, económico y cultural, que a menudo han venido en llamarse áreas naturales (Zorbaugh, 1926; Burgess, 1964).

La naturaleza multifacética de la realidad socioeconómica regional y urbana, ha llevado casi siempre consigo la consecuencia de la utilización de técnicas estadísticas multivariantes de clasificación y agrupación con el fin de analizar la diferenciación interna de los espacios urbanos, metropolitanos o regionales. Las distintas técnicas de análisis cluster, se han mostrado siempre extraordinariamente útiles, complementadas o no con análisis factoriales previos o discriminaciones posteriores.

En el presente trabajo, se expone la posibilidad de utilizar un tipo particular de redes neuronales, conocidas como redes de aprendizaje no supervisado, que de alguna forma, suponen una alternativa a los métodos tradicionales (estadísticos) de agrupación y se lleva a cabo una aplicación de la más popular de ellas a un estudio concreto de ámbito metropolitano.

Las similitudes y disimilitudes entre los métodos estadísticos y neuronales de análisis de datos han sido puestas de relieve recientemente en varios trabajos (Ripley, 1993; Sarle, 1994; Shalkoff, 1992; Waller, Kaiser, Illian, y Manry, 1998) quedando de manifiesto que el conjunto de problemas asimilables al reconocimiento de patrones son igualmente abordables por ambas vías, llevando a planteamientos metodológicos, a menudo, semejantes y, en ocasiones, claramente diferenciados con resultados también distintos.

Así, en el ámbito que nos ocupa, el del reconocimiento no supervisado de patrones (Shalkoff, 1992) o clustering, algunos procedimientos neuronales son prácticamente equivalentes a algunos métodos estadísticos, pero otros, como los mapas auto-organizados de Kohonen, presentan rasgos diferenciados; en este caso, la tipologización en categorías o clusters bidimensionalmente distribuidas en un plano.

Sin descartar, desde luego, el necesario estudio teórico, la comparación empírica de unos métodos y otros, se muestra, por tanto, una vía interesante de trabajo. En este sentido, un estudio anterior aplicado sobre la comarca de l'Horta Sud (Mtz. de Lejarza, 1996) ha revelado cómo, en algunas ocasiones, incluso el modelo simplificado de Kohonen puede conducir a clusterizaciones preferibles a las que se obtienen por los métodos tradicionales, y ha servido de revulsivo para ensayar esta

investigación, sobre un ámbito de mayor escala, utilizando el modelo de mapas bidimensionales auto-organizados (S.O.M.), en la línea de otras aplicaciones al análisis territorial (Blayo y Demartines, 1991; Varfis y Versino, 1992; Kropp, 1998; Blunden, Pryce y Dreyer, 1998).

2. LOS MAPAS AUTO-ORGANIZADOS DE KOHONEN: ARQUITECTURA, FUNCIONAMIENTO Y APRENDIZAJE.

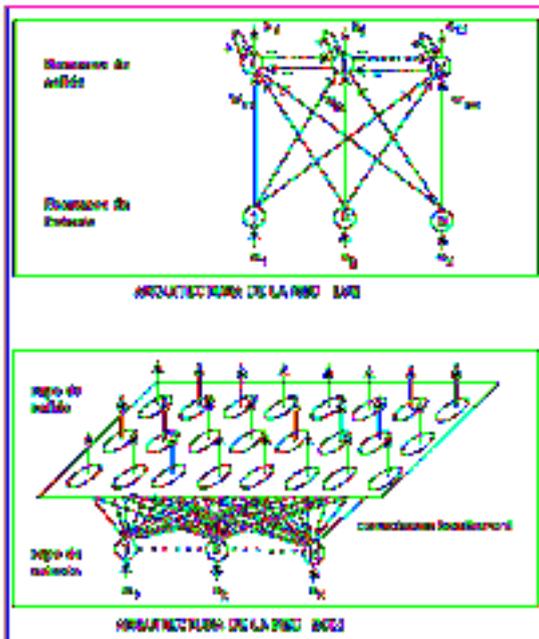
Los mapas auto-organizados de Kohonen (Kohonen, 1982; 1988; 1989) son redes neuronales con aprendizaje no supervisado; éste es, redes en las que la actualización de los pesos de las conexiones no requiere de información externa que le indique cuál debería ser la salida a obtener para cada conjunto de entradas. Los dos modelos más importantes de redes de Kohonen, los mapas auto-organizados (S.O.M.) y el Learning Vector Quantization (L.V.Q.) funcionan de forma similar y tienen un aprendizaje casi idéntico: Se trata en ambos casos de una red de dos capas feedforward: una capa de entrada por la que se introduce la información de los rasgos (variables) de los individuos y una de salida (bidimensional en el primer modelo y unidimensional en el segundo) que debe dar cuenta de las categorías de individuos obtenidas. La pretensión final, en ambos casos, es que, una vez entrenadas, las redes procedan de forma que, al serles suministrada la información de un cierto individuo, se active una y sólo una de las neuronas de la capa de salida; neurona que representará la categoría o cluster al que pertenece el individuo dado.

2.1. *Arquitectura de los modelos de Kohonen.*

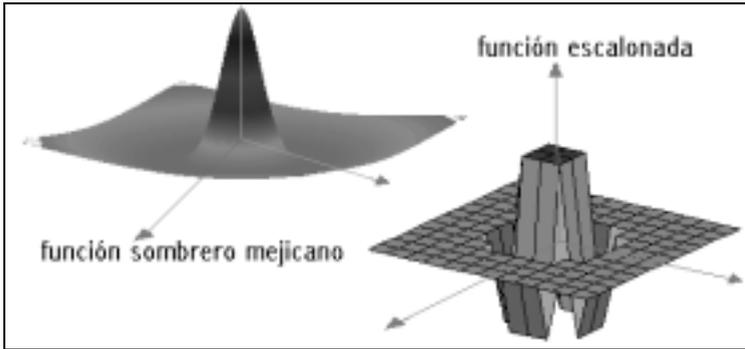
El modelo de Kohonen toma por objetivo fundamental el de crear una imagen de un espacio multidimensional de entrada en un espacio de salida de menor dimensionalidad. Las dos versiones del modelo (LVQ y SOM) constan de dos capas de neuronas una de entrada constituida por N neuronas de entrada que se limita a recoger y canalizar la información N -dimensional de entrada, y una capa de salida que procesa la información de entrada y da como resultado la representación reducida de la información introducida. Como ya se ha apuntado, la diferencia entre ambas versiones consiste en que mientras en el Learning Vector Quantization (LVQ) la capa de salida está constituida por un conjunto de M neuronas de salida dispuestas en una dimensión, el modelo SOM (mapa auto-organizado, propiamente dicho) dispone de una capa de salida constituida por $M = m_x \times m_y$ neuronas dispuestas en un plano (dos dimensiones). De esta forma, mientras el proceso llevado a cabo por el modelo LVQ puede asemejarse al análisis (estadístico) cluster tradicio-

nal, el resultado obtenido con un mapa auto-organizado de dos dimensiones resulta ser una clasificación de los vectores (información) de entrada de diferente naturaleza que no tiene un símil preciso en los métodos de análisis de datos tradicionales. Con todo, es obvio que el modelo LVQ puede considerarse como un caso unidimensional de mapa auto-organizado.

En este gráfico se presenta un esquema de la arquitectura de las dos redes auto-organizadas:



En ambos casos existirán conexiones laterales entre las neuronas de la capa de salida, dependiendo la influencia de cada neurona sobre las demás de una función de la distancia con ellas, habitualmente una función tipo sombrero mejicano; produciendo la excitación de las neuronas cercanas, la inhibición de las lejanas y siendo imperceptible en las muy lejanas. Precisamente estas conexiones laterales determinarán, en el proceso de aprendizaje, los pesos sinápticos que acabarán conectando las neuronas de la capa de entrada con las de salida.



Funciones de influencia lateral tipo sombrero mejicano, y escalonada

2.2. *Funcionamiento de la red.*

El funcionamiento de la red es relativamente simple. Cada vez que se presenta a la red una información de entrada; ésto es, cada vez que se introduce un vector N-dimensional de datos a través de la capa de entrada, la señal llega ponderada por los pesos w_{jk} (w_{jk} , en el caso LVQ) a las neuronas de la capa de salida por medio de las conexiones feedforward. Al mismo tiempo, estas neuronas reciben las correspondientes entradas del resto de las neuronas de salida a través de las conexiones laterales, y cuya influencia dependerá de la distancia a la que se encuentren:

Así, la salida generada en una neurona de salida u_{ij} ante el vector de entrada E_e dado por $E_e = (e_{1(e)}, \dots, e_{N(e)})$ vendrá dada por:

$$S_{ij}(t+1) = f (\sum w_{ijk} e_{k(e)} + \sum \text{Int}_{prij} S_{pr}(t))$$

(para $i, j \in \{(1,1), (1,2), \dots, (m_x, m_y)\}$ y $p, r \in \{(1,1), (1,2), \dots, (m_x, m_y)\}$)

En el modelo SOM, o bien:

$$S_i(t+1) = f (\sum w_{ijk} e_{k(e)} + \sum \text{Int}_{ri} S_r(t))$$

(para $i \in \{1, 2, \dots, m\}$ y $r \in \{1, 2, \dots, m\}$) en el modelo LVQ

Donde Int_{prij} es una función (tipo sombrero mejicano) que da cuenta de la influencia lateral de la neurona de salida u_{pr} sobre la u_{ij} . Y siendo la función de salida, f , una función continua de tipo sigmoideal o lineal adecuadamente ponderada para ir acercando a 1 la salida de las neuronas que van reforzándose y a cero las que van inhibiéndose.

Al tratarse de una red competitiva la red evolucionará hasta alcanzar una situación estable en la que, ante una determinada entrada, $E_e = (e_{1(e)}, \dots, e_{N(e)})$, sólo se active una neurona: la neurona ganadora. Así, la formulación de su funcionamiento puede simplificarse representando la activación final de las ($M = m_x \times m_y$) neuronas de salida como:

$$S_{i,j} = 1 \text{ si } \|E_e - W_{ij}\| = \min_{i,j} \{\|E_e - W_{ij}\|\}$$

$$S_{i,j} = 0 \text{ en caso contrario}$$

Siendo W_{ij} el vector de pesos ($w_{ij1}, w_{ij2}, \dots, w_{ijN}$) y siendo $\|E_e - W_{ij}\|$ una medida de la diferencia o discrepancia entre el vector de entrada y el vector de pesos; habitualmente la distancia euclídea.

Así pues, en la fase de entrenamiento, se van adaptando los pesos para que éstos acaben registrando los datos aprendidos y, de esta forma, en la fase posterior de funcionamiento, poder encontrar el dato "aprendido" al que más se parece la información de entrada suministrada. En definitiva, como vemos, la red lleva a cabo un proceso de clasificación, representando, la neurona activada, la clase a la que pertenece la información de entrada: Ante otra entrada parecida se activará la misma neurona, o quizá otra muy cercana (en el modelo bidimensional) garantizándose que las neuronas topológicamente próximas den cuenta de informaciones físicamente similares.

2.3. Aprendizaje de la red

El aprendizaje en los dos modelos de Kohonen es básicamente igual difiriendo obviamente en la dimensionalidad. Ambas aprenden según un procedimiento OFF LINE, secuencialmente separado del funcionamiento (y previo a él), utilizando, como ya se ha comentado, un aprendizaje no supervisado de tipo competitivo. En la fase de entrenamiento, se facilitan a la red un conjunto de vectores N- dimensionales de entrada (vectores de entrenamiento), para que ésta establezca, en función de su similitud, las categorías o clusters, (una por cada neurona de salida), que en el proceso de funcionamiento servirán para clasificar eventualmente nuevos datos. Se observa, por lo tanto, el enorme parecido entre el proceso de aprendizaje y el análisis cluster tradicional y se comprende, cómo el caso del aprendizaje en un mapa auto-organizado unidimensional (LVQ) pueda ser considerado, sin más, como un método de clusterización (tradicional) alternativo.

El algoritmo de aprendizaje empleado requiere de la iteración en la presentación de todos los patrones de aprendizaje, siendo necesario presentar varias veces (habitualmente, cientos o miles) el conjunto de datos de entrada para refinar el mapa topológico de salida e incluso conseguir la convergencia.

En el caso del modelo S.O.M., el algoritmo es, básicamente, el siguiente:

1. En primer lugar, se inicializan los pesos w_{ijk} con valores aleatorios pequeños, aunque también pueden establecerse a priori valores iniciales. Y se fija igualmente la zona inicial de vecindad entre las neuronas de salida.

2. A continuación, se presenta la información de entrada que la red debe aprender; el vector de entrada $E_e = (e_{1(e)}, \dots, e_{N(e)})$.

3. Se determina la neurona ganadora u_{ij} , cuyo vector de pesos, $W_{ij} = (w_{ij1}, w_{ij2}, \dots, w_{ijn})$ sea el más parecido al patrón de entrada introducido, E_e . Habitualmente la determinación se realiza a partir de la distancia euclídea al cuadrado:

$$d_{ij} = \sum (e_{k(e)} - w_{ijk})^2$$

(para $i, j \in \{(1,1), (1,2), \dots, (m_x, m_y)\}$).

Ello suele exigir, en las aplicaciones de clasificación, la normalización (tipificación) previa para evitar la sobrevaloración de algunas variables en detrimento de otras.

4. Una vez determinada la neurona ganadora (i^*j^*) se actualizan los pesos de las conexiones entre las neuronas de entrada y las neuronas de salida cercanas a la ganadora consiguiéndose con ello asociar la información de entrada con cierta zona de la capa de salida. La actualización se lleva a cabo de acuerdo con la regla:

$$w_{ijk}(t+1) = w_{ijk}(t) + a(t) (e_{k(e)} - w_{i^*j^*k}(t))$$

para las neuronas $u_{ij} \in G_{i^*j^*}(t)$.

Donde $G_{i^*j^*}(t)$ es una zona próxima alrededor de la neurona ganadora en la que se encuentran las neuronas cuyas pesos serán actualizados en el momento t del proceso de aprendizaje. Esta zona, en ocasiones, va reduciéndose con el proceso de aprendizaje y en otros casos se mantiene constante durante todo el proceso. Cuando se reduce esta zona suele hacerlo de forma lineal o exponencial de manera que, si $G_{i^*j^*}(t)$ viene dada por aquellas neuronas que distan de la ganadora una distancia menor a $R(t)$ (radio de aprendizaje), éste va variando según:

$$R(t) = R_o \left(1 - \frac{t}{a_2} \right) \text{ o bien: } R(t) = R_o \left(\frac{R_f}{R_o} \right)^{\frac{t}{a_2}}$$

siendo R_o el radio inicial de aprendizaje, R_f el radio final de aprendizaje y a_2 el número de iteraciones.

El término $a(t)$ es el llamado parámetro de ganancia, o factor de aprendizaje, toma valores comprendidos entre 0 y 1 y es habitual que decrezca con el número de iteraciones (con el tiempo de aprendizaje), de forma que al final del proceso sea prácticamente nulo. Habitualmente el factor de aprendizaje decrece lineal o exponencialmente, siendo lo más frecuente que dependa su decrecimiento del fac-

tor de aprendizaje inicial, a_1 , y del número de iteraciones, a_2 , que suelen convertirse así en los dos parámetros básicos del proceso. Las expresiones más habituales para el factor de aprendizaje son:

$$a(t) = 1/t; \quad a(t) = a_1 \left(1 - \frac{t}{a_2} \right); \quad \text{o bien } a(t) = a_1 \left(\frac{t}{a_2} \right)^{a_2}$$

5. Finalmente, el proceso, como se ha comentado se reitera un gran número de veces: Quinientas, mil o incluso más veces son presentadas a la red todos los patrones de entrenamiento: $\{E_e\}$ ($e= 1, 2, \dots, P$)

Por último, señalemos que la adecuada convergencia del proceso queda garantizada (Kohonen, 1992; Ripley, 1996) descartándose el inconveniente del sobre-entrenamiento que suele producirse en los modelos Backpropagation, si bien para evitar la convergencia a un mínimo local de la función de error en la clasificación se hace necesario elegir un adecuado conjunto de parámetros de entrenamiento, que supongan, de hecho, un relativamente grande radio inicial de aprendizaje en relación al radio total del mapa auto-organizado de salida y resulta conveniente varias inicializaciones aleatorias de los pesos iniciales y distintos esquemas de presentación de los datos de entrada (Ismail and Kamel, 1989). También debe señalarse que en evitación del llamado efecto borde, (inconveniente genuino de este tipo de redes), consistente en la tendencia a que se comprima excesivamente la representación de la zona del espacio de entrada que se asocia con los extremos del mapa, suele ser aconsejable utilizar un número de unidades de salida relativamente grande en relación al número de individuos a clasificar, sin menoscabo de que, lógicamente, el objetivo del análisis sea la clasificación de los individuos en un pequeño conjunto de clusters.

3. AGRUPACIÓN DE LAS ENTIDADES DE POBLACIÓN DEL ÁREA METROPOLITANA DE VALENCIA

A partir de lo expuesto hasta aquí, se ha procedido a la agrupación de los municipios integrantes de una zona geográfica restringida, aplicando para ello una red auto-organizada de Kohonen bidimensional (S.O.M.) y una red autoorganizada unidimensional (L.V.Q.) y se han comparado estos resultados con los que se obtienen aplicando un análisis cluster por el método de Ward.

3.1. *Ámbito de aplicación, información de partida y pre-proceso de los datos.*

La aplicación empírica ha consistido en la clasificación de los municipios del Área Metropolitana de Valencia, a partir de información socio-económica ya elabo-

rada. El ámbito territorial de aplicación ha quedado especificado, a partir de la última delimitación empírica disponible del área metropolitana, llevada a cabo mediante un modelo anisótropo de difusión con la distancia (Mtz. de Lejarza y Mtz. de Lejarza, 1995) incluyendo, además de la comarca metropolitana de L'Horta, algunos otros municipios de otras comarcas cercanas, abarcando un área geográfica compuesta por un total de 60 municipios:Alaquàs, Albal, Albalat de la Ribera, Albalat dels Sorells, Alboraià, Albuixech, Alcasser,Aldaia, Alfafar, Alfara del Patriarca, Algemesí, Alginet, Almassera, Almussafes, Benetússer, Benifaió, Beniparrell, Bétera, Bonrepós y Mirambell, Burjassot, Catarroja,Corbera, Quart de Poblet, Xirivella, l'Eliana, Emperador, Favara, Foios, Godella, Llaurí, Lloc Nou de la Corona, Manises, Massalfassar, Massamagrell, Massanassa, Meliana, Mislata, Moncada, Museros, Náquera, Paiporta, Paterna, Picanya, Picassent, Polinyà del Xúquer, Pobla de Farnals, El Puig, Puçol, Rafelbunyol, Riba-roja del Túria, Riola, Rocafort, Sedaví, Silla, Sollana, Sueca, Tavernes Blanques, Torrent, València y Vinalesa.

Con todo, de los iniciales 60 municipios englobados en Área Metropolitana, se ha prescindido del pequeño municipio de Lloc Nou de la Corona, ya que por razones de su escasa dimensión, en ocasiones, tenía un comportamiento anómalo para algunas de las variables y, en otras, simplemente se carecía de datos sobre el mismo.

La información utilizada en los análisis ha sido elaborada a partir de los datos socioeconómicos oficiales del Institut Valencià d'Estadística (I.V.E.). A partir de ellos se han seleccionado las más representativas y de comportamiento más diferenciado y, finalmente, para poder descontar los efectos debidos a la dimensión y a la escala, se han elaborado indicadores relativizados respecto de los comportamientos medios del Área Metropolitana. En la siguiente tabla se muestran los indicadores relativos utilizados.

TRPIND	TASA RELATIVA DE PARTICIPACIÓN EN EL SECTOR INDUSTRIAL
TRPSER	TASA RELATIVA DE PARTICIPACIÓN EN EL SECTOR SERVICIOS
IRNESTU	ÍNDICE RELATIVO DE NIVEL DE ESTUDIOS
TRPARO	TASA RELATIVIZADA DE PARO
NLTELH	COBERTURA TELEFÓNICA RELATIVA
IPFML	ÍNDICE DE PARTICIPACIÓN FEMENINA EN EL MERCADO LABORAL
IRDEN	ÍNDICE RELATIVO DE DENSIDAD DE POBLACIÓN
IRPREH	ÍNDICE RELATIVO DE PRESUPUESTO POR HABITANTE
IRESFIS	ÍNDICE RELATIVO DEL ESFUERZO FISCAL
IRTAC	TASA DE ACTIVIDAD RELATIVIZADA ¹
IRTACFE	TASA DE ACTIVIDAD FEMENINA RELATIVIZADA
IGEI	ÍNDICE RELATIVO DEL GASTO ENERGÉTICO INDUSTRIAL POR HAB.
IGSER	ÍNDICE RELATIVO DEL GASTO ENERGÉTICO EN SERVICIOS POR HAB.

Por último, la información sobre estos indicadores ha sido pre-procesada con el fin de descontar adecuadamente los efectos de la multicolinealidad de los indicadores utilizados. En la medida en que tanto el algoritmo de aprendizaje de la red SOM como el paquete estadístico (SPSS) empleado para el análisis cluster utilizan la distancia euclídea, es recomendable trabajar con los componentes principales de las variables a analizar, haciendo con ello equivalente el resultado final al que se obtendría utilizando la distancia de Mahalanobis. Así pues, las variables de entrada para los análisis han sido, finalmente, las cinco primeras componentes principales rotadas (rotación varimax- Kaiser-normalizada). Estas cinco componentes principales eran capaces de explicar un 81'7 % de la variabilidad total. Por otra parte, en función de la matriz de estructura factorial que reproducimos a continuación pueden, (tentativamente), interpretarse repectivamente como: *Infraestructura, Fuerza laboral, Modernidad socioeconómica, Potencia Industrial y Relevancia de la actividad industrial-concentración demográfica*. Las puntuaciones factoriales de los 59 municipios del Área Metropolitana de Valencia están recogidas en el anexo 1.

MATRIZ DE ESTRUCTURA FACTORIAL

	FACTOR 1 INFRAES- TRUCTURA	FACTOR 2 FUERZA LABORAL	FACTOR 3 MODERNIDAD	FACTOR 4 POTENCIA INDUSTRIAL	FACTOR 5 R.AC.IND. CONC. DEMOGRAF.
TRPIND	-0.19714	0.00699	-0.11071	0.18652	0.84540
TRPSER	0.08802	0.15895	0.89006	-0.20154	-0.00041
IRNESTU	0.11738	-0.03968	0.81299	0.02063	-0.08211
TRPARO	-0.09958	0.25523	-0.16233	-0.76166	0.12253
NLTELH	0.84167	-0.07576	0.18363	0.27900	0.09785
IPFML	-0.03186	0.51774	0.70417	0.14008	0.15644
IRDEN	-0.00354	0.06560	0.19892	-0.50605	0.64667
IRPREH	0.95057	0.01352	0.10712	0.04904	-0.09658
IRESFIS	0.89412	-0.07535	-0.05728	-0.04838	-0.18395
IRTAC	0.00269	0.96062	0.08957	-0.05596	0.07045
IRTACFE	0.06292	0.77769	0.46667	0.05225	0.23424
IGEI	0.13331	0.16302	-0.26393	0.75069	0.29010
IGSER	0.29538	-0.76556	0.08778	0.24139	0.34362

A partir de esta información, se ha realizado la agrupación de los municipios del Área metropolitana de Valencia para obtener una tipología bidimensional de 9 grupos, 3 filas por 3 columnas, utilizando la red autoorganizada de Kohonen.

Alternativamente se ha procedido a una agrupación en nueve grupos utilizando el modelo de mapa auto-organizado unidimensional, L.V.Q., tomando los mismos parámetros de entrenamiento que en el caso anterior.

Y, finalmente, se ha aplicado también un método jerárquico aglomerativo de análisis cluster, escogiéndose el método de Ward por ser el método jerárquico que habitualmente conduce a resultados más satisfactorios en términos de homogeneidad interna de los grupos, cortándose el dendograma aglomerativo por la solución de 9 grupos, para que todos los análisis fueran comparables.

3.2. Resultados de la aplicación del mapa auto-organizado .

Para obtener una tipología bidimensional de 3 filas por 3 columnas, se ha procedido a entrenar una red S.O.M. con una capa de entrada de 5 neuronas (una para cada rasgo de la información de entrada) y una capa de salida de 9 neuronas organizadas en 3 filas y 3 columnas. Para ello se ha utilizado el paquete de libre distribución para fines científicos "SOM_PAK v3.1" del "SOM Programming Team of the Helsinki University of Technology".

A partir de un proceso de auto-organización de 700 iteraciones, utilizando un factor de aprendizaje y un radio de alcance linealmente decrecientes con valores iniciales de $a_1 = 0.99$ y $R_0 = 2.25$, respectivamente, y partiendo de unos pesos iniciales aleatorios, se ha obtenido la cuantización final siguiente:

RASGOS-TIPO DE LA TIPOLOGIZACIÓN DE 3X3 GRUPOS

GRUPO	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
G(1,1)	4,947589	-0,581	-0,797	0,291	-0,794
G(1,2)	-0,0180345	0,308	-0,177	-0,084	-0,572
G(1,3)	-0,698	-0,0747	-0,015689	0,40495	0,1008926
G(2,1)	0,0301	0,337	-0,0456	-1,023588	1,355777
G(2,2)	-0,1495469	-0,825577	-1,674238	-1,027348	-1,453686
G(2,3)	1,358795	0,1998179	1,260159	0,2878099	-0,3504881
G(3,1)	0,733521	0,6120521	-1,134014	3,582967	1,664518
G(3,2)	-0,10925	-6,378389	0,1102	0,7146499	1,72009
G(3,3)	0,118	-0,212	2,831063	0,01469229	-0,6195372

Una vez obtenidas las ponderaciones, tras el proceso de auto-aprendizaje, el funcionamiento de la red procedería de forma que, al suministrar el patrón de entrada correspondiente a cada uno de los municipios del Área Metropolitana, se activaría la neurona de salida correspondiente al rasgo-tipo más cercano, adscribiéndose, por tanto, a tal grupo. Finalmente la tipología de 3x3 grupos quedaría según se muestra en el cuadro adjunto. (En el anexo 2 se muestra esta agrupación sobre el plano del Área Metropolitana). Con todo, debe tenerse presente que la disposición topológica de los grupos (celdas) obtenidos no se corresponde con una métrica bidimensional y, por lo tanto, no debemos pensar que la cercanía de dos celdas vecinas es métricamente equivalente a la de otras dos vecinas².

AGRUPACIÓN EN 9 (3X3) GRUPOS POR MAPA AUTO-ORGANIZADO (SOM)

1		2			3						
1	Náquera	Algemesí	Moncada	Puçol	Albal	Foios					
		Alginet	Museros	Rafelbunyol	Albalat Sor.	Alfara del Pat.	Massalfassar				
		Benifaíó	Paterna	Riba roja	Albuixech	Almassera	Massanassa				
		Bétera	Picanya	Sollana	Alcacer	Bonrepós i M. Meliana					
		Corbera	Picassent	Sueca	Aldaya	Catarroja	Silla				
		Massamagrell	Polínia X.	Torrent	Alfafar	Vinalesa					
2	Alaquás	Manises	Albalat de la R Ribera.		Llaurí	Alboraya	P. de Farnals				
	Benetússer	Mislata									
	Burjassot	Paiporta						Favara	Riola	Eliana, l'	Puig
	Quart	Sedavi						Emperador		Godella	Rocafort
Xirivella	Tavernes										
3	Almussafes	Beniparrell									

De la tipología obtenida, se pueden entresacar algunas consecuencias inmediatas interesantes como la singularidad de los municipios de Náquera y Emperador, o la caracterización como grupo compacto del conjunto de cuatro municipios más alejados del núcleo central del Área Metropolitana por la dirección Sur (grupo (2,2)), o la agrupación de los dos municipios de mayor carácter industrial: Almussafes, sede de la factoría Ford y Beniparrell, donde se ubica el polígono industrial de mayor consumo energético del Área metropolitana.

La disposición topológica de los grupos, también puede orientar, en gran medida, el análisis de los resultados, si bien, es necesario no extraer conclusiones precipitadas, puesto que la cercanía topológica entre los distintos grupos no es homogénea y así, por ejemplo, a pesar de que el grupo (3,3) se haya al mismo número de celdas de los grupos (2,3) y (3,2), la similitud entre ellos es muy diferen-

ción hecha de Valencia), sólo ocurre en los dos métodos neuronales (grupo 8 por el método L.V.Q. ; grupo (2,1) por el método S.O.M.).Y al igual ocurre con el grupo formado por los municipios de Alboraya, l'Eliana, El Puig y Pobla de Farnals, que al margen de su comportamiento en cuanto a las variables analizadas se caracterizan por ser entidades de población de reciente crecimiento residencial, que han ido decantándose desde la residencialidad estacional hacia el régimen de "commuters".

Una forma adecuada de comparar las agrupaciones obtenidas es llevar a cabo distintos análisis uni y multivariantes de la varianza. Cabe esperar, que en todos los casos se concluyan diferencias significativas en el comportamiento de los grupos obtenidos pero el análisis pormenorizado de los resultados puede permitir valorar la calidad de las agrupaciones en términos de una mayor homogeneidad interna-heterogeneidad externa. En el siguiente cuadro se muestran los resultados del análisis multi y univariante de la varianza tomando como variables dependientes los factores utilizados para la agrupación. En el MANOVA, se muestran en negrita, para cada uno de los criterios (Pillai, Wilks, Hotelling , Roy y traza de la matriz de suma de cuadrados interna), los valores que conllevan el resultado más satisfactorio; comprobándose un óptimo comportamiento de la agrupación obtenida por el método S.O.M.

MANOVA

ESTADÍSTICOS	S/ AGRUP. DE WARD	S/ AGRUPACIÓN LVQ	S/ AGRUPACIÓN SOM
V de Pillai	3.83054	3.65535	3.82250
F aprox.	20.47171	16.99031	20.28926
N. Significación	0.000	0.000	0.000
T de Hotelling	18.98994	16.96745	21.08293
F aprox.	21.07884	18.83387	23.40205
N. Significación	0.000	0.000	0.000
L de Wilks	0.00051	0.00094	0.00043
F aprox.	23.78807	20.04184	25.03762
N.Significación	0.000	0.000	0.000
R de Roy	0.85728	0.87002	0.88262
Traza de NW	67.822874	77.9895	68.29506

ANOVA

F ANOVA FACT 1 "Infraestructura"	9.77814	13.97407	14.64658
(n.s)	0.000	0.000	0.000
F ANOVA FACT 2 "Fuerza laboral"	28.84426	22.54540	28.97781
(n.s)	0.000	0.000	0.000
F ANOVA FACT 3 "Modernidad"	27.42437	17.34325	22.74814
(n.s)	0.000	0.000	0.000
F ANOVA FACT 4 "Potencia Ind."	27.37947	18.67358	19.78007
(n.s)	0.000	0.000	0.000
F ANOVA FACT 5	20.92628	14.38472	19.22337
(n.s)	0.000	0.000	0.000

VARIABLES MÁS DISCRIMINANTES EN CADA AGRUPACIÓN

AGRUPACIÓN	WARD	LVQ	SOM
VARIABLE	Índice de gasto energético industrial	Índice de gasto energético industrial	Índice de gasto energético industrial
ESTADÍSTICO F	104.97	96.90	100.88
DIFERENCIAS	G.5 > todos los demás	G.9 > todos los demás	G(3,1) > todos los demás
SIGNIFICATIVAS			
(Scheffé) pares de grupos			
VARIABLE	Índice relativo de actividad industrial	Índice relativo de actividad industrial	Índice relativo de actividad industrial
ESTADÍSTICO F	45.58	35.75	43.86
DIFERENCIAS	G.8 < todos los demás	G.1 < todos los demás	G(3,2) < todos los demás
SIGNIFICATIVAS	G.2 < G.3 // G.4 // G.1 // G.6	G.3 < G.8	G(2,2) < G(1,3) // G(1,2) // G(2,3) // G(2,1)
(Scheffé) pares de grupos			
VARIABLE	I.R. de densidad de pobl.	I.R. de cobertura telefónica	I.R. de cobertura telefónica
ESTADÍSTICO F	32.12	31.24	30.98
DIFERENCIAS	G.6 > todos los demás	G.2 > todos los demás	G(1,1) > todos los demás
SIGNIFICATIVAS		G.5 > G.7 // G.4 // G.8	G(2,3) > G(2,2) // G(1,3) // G(1,2) // G(2,1)
(Scheffé) pares de grupos		G.9 > G.7 // G.4 // G.8	G(3,1) > G(2,2) // G(1,3) // G(1,2) // G(2,1)
VARIABLE	I.R. de nivel de estudios	I.Gasto energético en servicios	I.Gasto energético en servicios
ESTADÍSTICO F	20.91	21.27	20.82
DIFERENCIAS	G.7 > todos los demás	G.1 > todos los demás	G(3,2) > todos los demás
SIGNIFICATIVAS			
(Scheffé) pares de grupos			

Si consideramos el comportamiento de los grupos obtenidos respecto de las variables originales, las conclusiones pueden ser más clarificadoras en lo que se refiere a la interpretación de estas agrupaciones. En el siguiente cuadro se detallan para cada agrupación, las cuatro variables que mejor discriminan cada una de ellas, detallándose, igualmente los pares de grupos que presentan diferencias significativas ($\alpha= 0.05$, según el criterio de Scheffé).

5. CONCLUSIONES

Al aplicar un mapa auto-organizado a la agrupación de los municipios de Área Metropolitana de Valencia, se ha obtenido una tipología bidimensional de los mismos, que mantiene una homogeneidad interna / heterogeneidad externa, comparable, o incluso preferible según algunos criterios a la alcanzada por la Clusterización de Ward. Al mismo tiempo, la disposición bidimensional de los grupos obtenidos permite sacar algunas conclusiones respecto de sus posiciones relativas, que añaden elementos de análisis nuevos en relación con los métodos tradicionales del análisis cluster. Con todo, a la hora de interpretar las relaciones topológicas entre los grupos obtenidos no se puede olvidar que las posiciones relativas de los grupos sobre la retícula obtenida no son homogéneas y que pares de grupos con el mismo grado de vecindad pueden tener discrepancias en su comportamiento de muy distinto orden.

En el caso que nos ocupa, de la tipología obtenida pueden concluirse algunas consideraciones importantes:

1. El comportamiento singular del municipio de Emperador, caracterizado por una escasa actividad, orientada fundamentalmente al sector servicios, y por valores inferiores a los medios en casi todas las variables relacionadas con la actividad económica.
2. En menor medida, la singularidad del municipio de Náquera, explicable en términos de un muy superior esfuerzo fiscal (4,7 veces el valor medio metropolitano), una elevada participación en el sector de la construcción (el doble de la media) y una cobertura telefónica tres veces superior a la media metropolitana; rasgos, que conjuntamente considerados parecen obedecer a zonas más bien espúreas.
3. La agrupación en un cluster homogéneo de los cuatro municipios más alejados , caracterizados fundamentalmente por una escasa participación femenina en el mercado laboral (con una tasa media de actividad femenina tres veces inferior a la tasa media metropolitana), y una actividad en el sec-

tor servicios inferior a la media, que, junto con una escasa actividad industrial y una no muy alta participación en la construcción, parece indicarnos una fuerte importancia relativa de la agricultura que no se da en el resto del Área Metropolitana.

4. La constitución de un cluster de marcado carácter industrial, integrado por los municipios de Almusafes y Beniparrell, que tienen un consumo energético industrial 10 veces superior a la media metropolitana.
5. La agrupación, junto con Valencia, de los municipios de Rocafort y Godella, municipios caracterizados, históricamente, por desarrollar una clara función residencial para clases acomodadas y que al constituir este grupo junto con Valencia, forman el cluster que alcanza un mayor nivel de formación académica con un indicador medio casi tres veces superior al del conjunto metropolitano y con una mayor participación femenina en el mercado laboral y una importancia relativa extraordinaria del sector servicios.
6. La proximidad tipológica (y topológica) del grupo de municipios de mayor crecimiento en su función residencial extrarradial (Commuters) al grupo anterior, que se explica principalmente por su similar comportamiento (aunque en inferior medida) en lo que se refiere a la participación femenina en el trabajo, en la formación académica y en la relevancia del sector servicios.
7. La agrupación de los municipios de mayor densidad y características típicamente urbanas (excepción hecha de Valencia) en un cluster homogéneo formado por: Alaquás, Benetusser, Burjassot, Quart, Xirivella, Manises, Mislata, Paiporta, Sedaví y Tavernes, caracterizado por una muy alta densidad de población, y una importante relevancia de la actividad industrial que no se ve acompañada, sin embargo, de un alto consumo energético industrial.
8. Y finalmente, la agrupación del resto de los municipios en dos cluster cercanos en cuanto a su comportamiento, y cercanos en la tipologización que apenas se distinguen en su distinta participación relativa en los sectores industrial y construcción, decantándose el grupo (1,2) hacia la industria en mayor medida que el grupo (1,3), y ocurriendo lo contrario por lo que respecta a la construcción.

El gráfico que mostramos a continuación señala finalmente la agrupación obtenida por el método SOM, reseñando con distintos colores, las características propias de cada grupo con respecto a las variables utilizadas. Igualmente, las flechas nos indican las distorsiones que hay que considerar en el mapa para interpretar adecuadamente la disposición topológica de los grupos: algunos de ellos están más alejados de lo que da a entender la disposición reticular y otros más cercanos de lo que aparenta el mapa.



Finalmente, debe remarcarse que la utilización de redes auto-organizadas en el análisis territorial parece mostrarse tan adecuada como los métodos tradicionales en cuanto a su validez estadística y su capacidad discriminativa, posibilitando además otras percepciones relacionadas con la disposición topológica de los grupos que pueden resultar, en ocasiones de gran interés. Cabe concluirse pues, que el de la clasificación es otro campo de aplicación prometedor de las redes neuronales en el ámbito de los estudios regionales.

BIBLIOGRAFÍA

- BLAYO,F y DEMARTINES,P.: "Data Analysis: How to compare Kohonen Neural Networks to other Techniques?".*Proceedings of IWANN 91,Ed. Prieto.Lectures Notes in Computer Sciences*.Springer-Verlag,pp.469-476.
- BLUNDEN, J. R.; PRYCE,W.T.R.; DREYER,P.(1998): "The Classification of Rural Areas in the European Context: An Exploration of a Typology Using Neural Network Applications" *RegionalStudies*; 32(2), April 1998, pages 149-60.
- BURGUESS, E.W.(1964) : "Natural Area" en Gould,J. y Kolb,W.(Eds.) "A dictionary of the Social Sciences".N.Y.; pp.458.
- ISMAIL, M.A., and KAMEL, M.S. (1989), "Multidimensional data clustering utilizing hybrid search strategies," *Pattern Recognition*, 22, 75-89.
- KOHONEN, T (1988b): "Learning Vector Quantization" *Abstracts of the First Annual INNS Meeting*, 303.
- KOHONEN, T. (1982): "Self-Organized formation of topologically correct feature maps" *Biological Cybernetics*, 43, pp.59-69. Reimpreso en: Anderson,J y Rosenfeld,E. (Eds.),(1988): *Neurocomputing*. MIT Press.
- KOHONEN,T. (1989): "Self-Organization and Associative Memory" Springer-Verlag.
- KOHONEN,T.(1988a):"An introduction to Neural Computing".*Neural Networks*, Vol.1,pp.3-16
- KROPP,J. (1998): "A Neural-Network Approach to the Analysis of City Systems" *Applied Geography*, Vol 18, Iss 1, pp 83-96,1998
- MTZ. DE LEJARZA, I. y MTZ. DE LEJARZA, J. (1995): "Un modelo de difusión anisótropa con la distancia para la delimitación de Áreas Metropolitanas". *XXI Reunión de Estudios Regionales*, pp. 65-80.
- MTZ. DE LEJARZA, Y.(1996): "Redes neuronales auto-organizadas y clustering: Una aplicación a la agrupación económico-funcional de entidades de población".*Quaderns de Treball*, nº 27 (nueva época).Facultad de CC. Económicas y Empresariales.Universidad de Valencia
- RIPLEY,B.D.,(1993): "Statistical Aspects of Neural Network". en *Barndorff-Nielsen,O.E.;Jensen,J.L. y Kendal,W.S.(eds.):"Networks & Chaos:Statistical and Probabilistics Aspects"* Chapman & Hall.
- RIPLEY, B.D. (1996) *Pattern Recognition and Neural Networks*, Cambridge: Cambridge University Press
- SARLE,W. (1994): "Neural Networks and Statistical Methods" *Proceedings of the Nineteenth Annual S.A.S. Users Group International Conference*.
- SCHALKOFF,R.(1992): "Pattern Recognition. Statistical, Structural and Neural Approaches" John Willey & Sons.

- VARFIS, A. y VERSINO, C. (1992): "Clustering of Socio-Economic data with Kohonen Maps", *International Journal on Neural and Mass-Parallel Computing and Informations Systems "Neural Network World"*, vol. 2, nº 6 pp. 813-833.
- WALLER, N.G.; KAISER, H.A.; ILLIAN, J.B.; MANRY, M. (1998): "A Comparison of the Classification Capabilities of the 1-Dimensional Kohonen Neural-Network with 2 Partitioning and 3 Hierarchical Cluster-Analysis Algorithms" *Psychometrika*, Vol 63, Iss 1, pp 5-22, 1998
- ZORBAUGH, H.W. (1926): "The natural areas of the city" *Publ. American Sociology Society*, 20. pp. 188-197.

Recibido, Abril de 1999; Aceptado, Abril de 2000.

ANEXO 1
PUNTUACIONES FACTORIALES DE LOS MUNICIPIOS DEL A.M.V.

MUNIC	FACT1	FACT2	FACT3	FACT4	FACT5
Alaquás	-0,26321	0,46799	-0,39823	-0,35906	1,19034
Albal	-0,0795	0,0983	-0,94009	0,35287	0,62363
Albalat de la Rib.	-0,28904	-0,77504	-1,983	-1,2715	-1,34183
Albalat dels Sor.	-0,57076	0,23958	0,02214	0,40883	-0,29338
Alboraya	0,28972	-0,09434	1,57895	0,75051	-0,37226
Albuixech	-0,9294	-1,54297	-0,03542	0,78209	-0,45689
Alcacer	-0,97871	0,1997	-0,09508	-0,09292	-0,3056
Aldaya	-0,95157	0,45067	-0,61244	0,36929	1,06071
Alfafar	-0,96559	-0,13424	-0,47978	-0,45101	0,66459
Alfara del Patria.	0,02868	-0,24308	0,36838	0,20429	0,2685
Algemesí	-0,08	0,62724	-0,0096	-0,69745	-0,60981
Alginet	-0,44481	0,27385	-0,62565	0,68038	-1,43873
Almassera	-0,41571	0,55209	-0,06488	0,5903	0,56041
Almussafes	-0,94569	0,97427	-1,02652	3,63736	1,10833
Benetússer	0,3059	-0,08828	-0,20178	-2,16782	2,53483
Benifaió	-0,18164	0,83772	-0,49747	0,87117	-0,57604
Beniparrell	2,3595	0,26132	-1,2381	3,5303	2,20307
Bétera	-0,02385	0,1236	0,36571	0,32864	-0,90651
Bonrepós i Mir.	-0,99459	0,72127	0,57923	0,44034	0,07844
Burjassot	-0,08841	-0,07534	0,85121	-1,10666	1,42492
Catarroja	-0,49074	-0,44083	0,31285	0,13384	0,16182
Corbera	-0,99606	0,43138	0,55511	-0,37657	-1,44652
Quart de Poblet	0,58324	0,58856	-0,67548	-0,57577	0,74914
Xirivella	-0,02162	0,50175	-0,05278	-1,26215	0,82181
Eliana, l'	1,80836	-0,03563	2,28704	0,26778	-0,39406
Emperador	-0,10925	-6,37839	0,1102	0,71465	1,72009
Favara	-1,14824	-0,68633	-1,13301	-0,51616	-0,56576
Foios	-1,20266	-0,26208	0,95206	0,71985	-0,48158
Godella	-0,21931	0,03164	2,70769	0,29472	-0,97316
Llaurí	1,41184	-0,60892	-2,1152	-1,55618	-2,05541
Manises	-0,74314	0,26463	-0,18258	-0,5005	1,0962
Massalfassar	-1,02618	-1,17921	-0,02075	1,1357	-0,62775
Massamagrell	-0,24146	0,56446	-0,2938	-0,66247	-0,25714
Massanassa	-0,19313	-0,30883	0,32586	-0,07151	0,00147

ANEXO 1
PUNTUACIONES FACTORIALES DE LOS MUNICIPIOS DEL A.M.V.

MUNIC	FACT1	FACT2	FACT3	FACT4	FACT5
Moncada	-0,15502	0,51563	0,19787	-0,19916	-0,27457
Museros	0,2463	0,85497	-0,48907	0,691	-0,22085
Náquera	4,94759	-0,58075	-0,79702	0,29064	-0,79377
Paiporta	0,32809	0,57714	-0,57003	-0,81469	0,73106
Paterna	0,24347	0,37595	0,5925	-0,52164	0,18186
Picanya	-0,02249	0,43075	0,07741	-0,46412	-0,33654
Picassent	0,24588	0,46842	-0,79876	0,01611	-0,23285
Polinyá de X.	-0,45997	-0,1949	-0,33172	-0,31796	-1,36222
Pobla de Farnals	1,60281	0,35983	0,37741	-0,58352	-0,26436
Puig	1,68944	0,54318	0,844	0,72636	-0,3727
Puçol	0,26009	0,55813	-0,19988	0,27772	-0,44066
Rafelbunyol	-0,42823	-0,22176	0,16975	0,08469	-0,91093
Riba roja de Túria	0,39892	0,3188	-0,7459	-0,31228	-0,00732
Riola	-0,59113	-1,2176	-1,47193	-0,77081	-1,81844
Rocafort	0,0003	-0,32766	2,94304	0,40676	-0,93396
Sedaví	0,11872	0,16355	-0,44881	-0,48717	1,10916
Silla	-0,93216	0,43355	-0,90238	0,49339	0,65313
Sollana	0,1453	-0,15688	-0,1774	0,41532	-0,98737
Sueca	0,81517	-0,39829	-0,55101	-0,36583	-0,80255
Tavernes B	-0,1142	0,73134	0,42442	-0,67183	1,79266
Torrent	0,1473	0,33445	-0,31988	-0,83736	0,19985
València	0,55377	-0,33136	2,83906	-0,63289	0,02225
Vinalesa	-0,33113	0,96451	0,18207	0,89815	-0,2919

ANEXO 2
TIPOLOGÍA DE 9 (3X3) GRUPOS SEGÚN S.O.M.

